2019

人工智能发展报告

Report of Artificial Intelligence Development

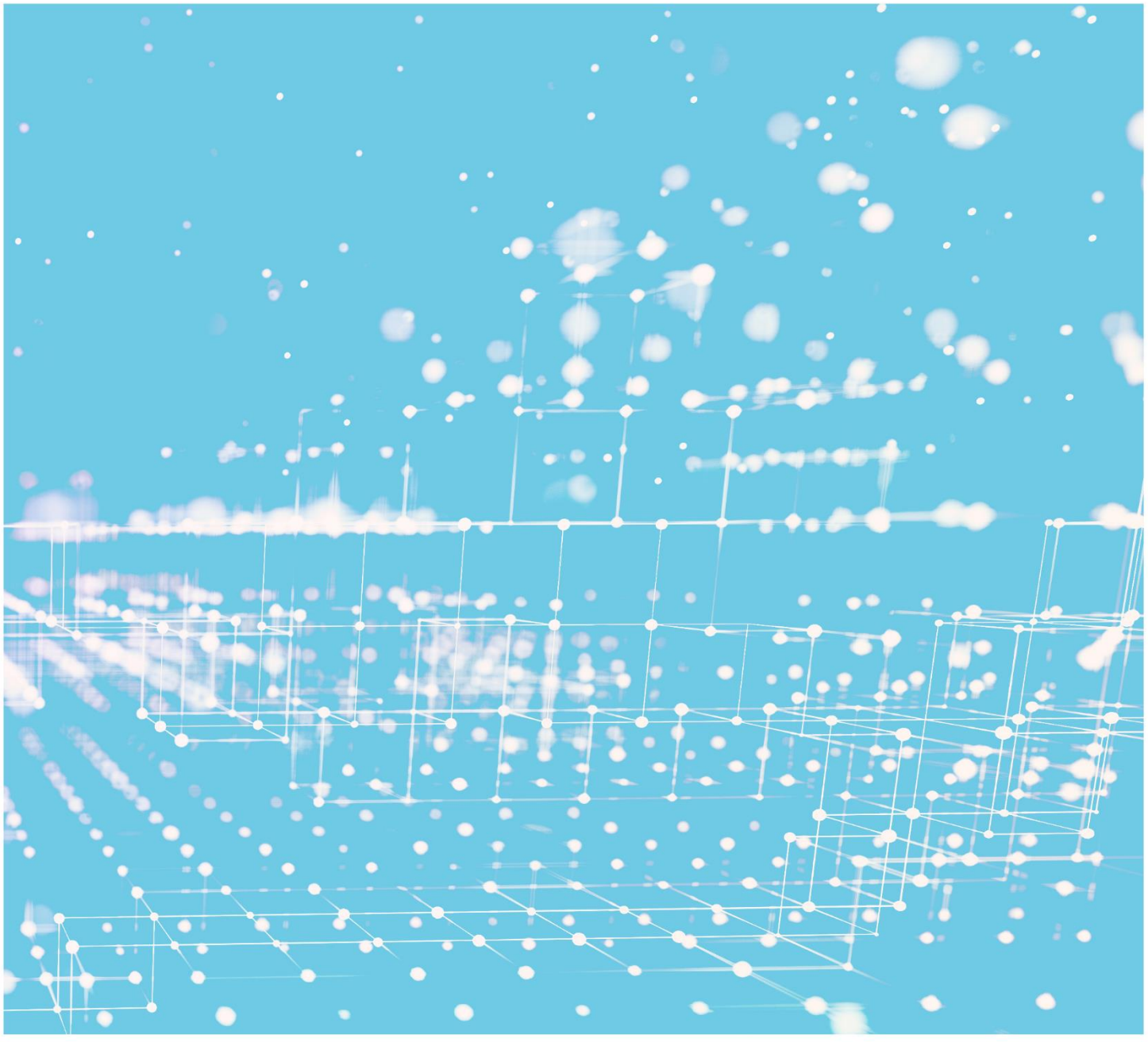


清华大学–中国工程院知识智能联合研究中心
中国人工智能学会吴文俊人工智能科学技术奖评选基地

2019年12月

人工智能发展报告

Report of Artificial Intelligence Development



清华大学–中国工程院知识智能联合研究中心
中国人工智能学会吴文俊人工智能科学技术奖评选基地

2019年12月

编写委员会（按姓氏拼音排序）

主　　编：李涓子　唐　杰

编　　委：曹　楠　程　健　贾　珈　李国良　刘华平

宋德雄　喻　纯　余有成　朱　军

责任编辑：景　晨　刘　佳

编　　辑：毕小俊　程时伟　韩　腾　侯　磊　刘德兵

刘　越　骆昱宇　麻晓娟　仇　瑜　王若琳

徐　菁

技术支持：北京智谱华章科技有限公司

CONTESTS

1. 中国人工智能发展报告编制概要

1.1 编制背景 ---- 1
1.2 编制目标与方法 ---- 1

2. 机器学习

2.1 机器学习概念 ---- 4
2.2 机器学习发展历史 ---- 6
2.3 机器学习经典算法 ---- 7
2.4 深度学习 ---- 21
2.5 人才概况 ---- 37
2.6 代表性学者简介 ---- 39
2.7 论文解读 ---- 60

3. 计算机视觉

3.1 计算机视觉概念 ---- 85
3.2 计算机视觉发展历史 ---- 87
3.3 人才概况 ---- 89
3.4 论文解读 ---- 91
3.5 计算机视觉进展 ---- 105

4. 知识工程

4.1 知识工程概念 ---- 107
4.2 知识工程发展历史 ---- 108
4.3 人才概况 ---- 111
4.4 论文解读 ---- 113
4.5 知识工程最新进展 ---- 129

5. 自然语言处理

5.1 自然语言处理概念 ---- 131
5.2 自然语言的理解发展历史 ---- 132
5.3 人才概况 ---- 133
5.4 论文解读 ---- 136
5.5 自然语言处理最新进展 ---- 153

目录

6. 语音识别

6.1 语音识别概念 155
6.2 语音识别发展历史 156
6.3 人才概况 158
6.4 论文解读 160
6.5 语音识别进展 173

7. 计算机图形学

7.1 计算机图形学概念 175
7.2 计算机图形学发展历史 175
7.3 人才概况 178
7.4 论文解读 181
7.5 计算机图形学进展 194

8. 多媒体技术

8.1 多媒体概念 197
8.2 多媒体技术发展历史 198
8.3 人才概况 200
8.4 论文解读 203
8.5 多媒体技术进展 215

9. 人机交互技术

9.1 人机交互概念 217
9.2 人机交互发展历史 218
9.3 人才概况 222
9.4 论文解读 225
9.5 人机交互进展 239

10. 机器人

10.1 机器人概念 241
10.2 机器人发展历史 242
10.3 人才概况 245
10.4 论文解读 247
10.5 机器人进展 260

11. 数据库技术

11.1 数据库概念 263
11.2 数据库技术历史 264
11.3 人才概况 266
11.4 论文解读 269
11.5 数据库技术重要进展 287

12. 可视化技术

12.1 可视化技术概念 289
12.2 可视化技术发展历史 290
12.3 人才概况 294
12.4 论文解读 296
12.5 可视化进展 312
12.6 可视化应用 314

13. 数据挖掘

13.1 数据挖掘概念 321
13.2 数据挖掘的发展历史 323
13.3 人才概况 324
13.4 论文解读 326
13.5 数据挖掘进展 337

14. 信息检索与推荐

14.1 信息检索与推荐概念 339
14.2 信息检索和推荐技术发展历史 341
14.3 人才概况 345
14.4 论文解读 347
14.5 信息检索与推荐进展 362

15. 结束语 365

参考文献 366

附录 372

1 编制概要

1.1 编制背景

21 世纪前两个十年，在大规模 GPU 服务器并行计算、大数据、深度学习算法和类脑芯片等技术的推动下，人类社会相继进入互联网时代、大数据时代和人工智能时代。当前，随着移动互联网发展红利逐步消失，后移动时代已经来临。当新一轮产业变革席卷全球，人工智能成为产业变革的核心方向：科技巨头纷纷把人工智能作为后移动时代的战略支点，努力在云端建立人工智能服务的生态系统；传统制造业在新旧动能转换，将人工智能作为发展新动力，不断创造出新的发展机遇。

现今，人工智能的发展对国民经济具有重要意义，人工智能通过综合各生产要素作用于国民经济活动，有利于提高生产力水平，助力实体经济发展，主要表现在以下四个方面：一是人工智能可以依托大数据，对庞大的信息资源进行处理，分析得到有效数据，避免了错误的经济决策，推进经济持续稳定的发展。二是人工智能可以通过智能化的精准控制来达到减少资源浪费、提高生产水平和生产效率的目的。三是人工智能可以赋能于商业生态，以电能为动力源的人工智能可以做到减少碳排放，达到节能环保的效果。四是在人工智能的驱动下，产业经济与信息经济相互整合，改变了传统的“需求-设计-制造-销售-服务”的生产模式。由于互联网等信息技术的应用，使得不同产业间的关联关系不断改变，新的产业不断涌现，跨界和融合发展成为产业生态的重要特征，提高了经济增长的质量，推动了经济整体结构的调整。

人工智能处于第四次科技革命的核心地位，在该领域的竞争意味着一个国家未来综合国力的较量。我国在人工智能领域的发展上有其独特优势，如稳定的发展环境、充足的人才储备、丰富的应用场景等；同时，需要注意的是，我国人工智能发展起步较晚，与以美国为主的发达国家相比还有一定差距。人工智能对于任何国家来说既是机遇又是挑战，世界格局极有可能因此而重新洗牌，对于错过前三次科技革命的我国来说，此次机遇尤为重要。近年来，我国政府高度重视人工智能的发展，相继出台多项战略规划，鼓励指引人工智能的发展。2015 年，

《国务院关于积极推进“互联网+”行动的指导意见》颁布，提出“人工智能作为重点布局的 11 个领域之一”；2016 年，在《国民经济和社会发展第十三个五年规划纲要（草案）》中提出“重点突破新兴领域人工智能技术”；2017 年，人工智能写入十九大报告，提出推动互联网、大数据、人工智能和实体经济深度融合；2018 年，李克强总理在政府工作报告中再次谈及人工智能，提出“加强新一代人工智能研发应用”；2019 年，习近平主席主持召开中央全面深化改革委员会第七次会议并发表重要讲话，会议审议通过了《关于促进人工智能和实体经济深度融合的指导意见》。目前，在多层次战略规划的指导下，无论是学术界还是产业界，我国在人工智能国际同行中均有不错的表现，在世界人工智能舞台上扮演了重要的角色，我国人工智能的发展已驶入快车道。

我国人工智能的发展也离不开人工智能团体组织与先进平台的参与和协助。中国人工智能学会（Chinese Association for Artificial Intelligence，CAAI）成立于 1981 年，是经国家民政部正式注册的我国智能科学技术领域唯一的国家级学会，目前拥有 48 个分支机构，包括 40 个专业委员会和 8 个工作委员会，覆盖了智能科学与技术领域，基本任务是团结全国智能科学技术工作者和积极分子通过学术研究、国内外学术交流、科学普及、学术教育、科技会展、学术出版、人才推荐、学术评价、学术咨询、技术评审与奖励等活动促进我国智能科学技术的发展，为国家的经济发展、社会进步、文明提升、安全保障提供智能化的科学技术服务。科技情报大数据挖掘与服务平台（AMiner）2006 年上线，经过十多年的建设发展，已收录 2.3 亿篇论文与 1.3 亿位学者，吸引了全球 220 个国家/地区、800 多万独立 IP 的访问，年度访问量 1100 万次。AMiner 平台曾获得 2017 年北京市科学技术奖一等奖，2013 年中国人工智能学会科学技术进步一等奖。AMiner 平台已经服务于科技部、中国科协、自然科学基金委、北京科委等政府机构，以及腾讯、华为、阿里巴巴、搜狗等企业机构。人工智能团体组织与先进平台的成立和发展已经成为团结优势资源共同促进人工智能发展的重要力量，见证并融入到了我国人工智能伟业的发展。

1.2 编制目标与方法

本报告由清华大学知识智能联合研究中心团队负责编写。依托于 AMiner 平台的数据资源及技术挖掘成果生成相关数据报告及图表，邀请清华大学、同济大学等高校专家解读核心技术及提出观点建议。报告遴选 13 个人工智能的重点领域进行重点介绍，包括：机器学习、知识工程、计算机视觉、自然语言处理、语音识别、计算机图形学、多媒体技术、人机交互、机器人、数据库技术、可视化、数据挖掘、信息检索与推荐等。在述说各领域概念及发展情况等内容的基础上，报告着重介绍了各领域人才情况以及对代表性文章的解读。

AMiner 平台推荐了各领域代表性的期刊/会议，并由专家进行补充，挖掘这些期刊/会议近 10 年论文，确定了 h-index 排名前 2000 的学者，构建各领域学者库。我们将这些学者供职机构的位置信息绘制于地图上得到了学者分布地图，研究各领域学者在世界及我国的分布规律；同时，我们进一步统计分析了各领域学者性别比例、h-index 分布等情况。对于中国在各领域的合作情况也进行了挖掘分析，通过统计中文合作论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的情况。

报告还选取这些期刊/会议上发表的高水平论文作为代表，对近年来的热点及前沿技术进行深度解读，既包括高引论文、最佳论文，又有专家推荐的代表性工作。解读前沿热点研究问题，深入探讨研究方法，展现最新研究成果。为读者了解近期人工智能相关领域的发展动向、基础及应用研究的代表性成果提供了信息窗口。

当前，人工智能正处在爆发期。我国在人工智能领域的科学技术研究和产业发展起步稍晚，但在最近十余年的时间里抓住了机遇，进入了快速发展阶段。在这个过程中，技术突破和创造性高端人才对人工智能的发展起着至关重要的作用。本报告对人工智能 13 个领域的人才情况及技术发展等内容进行了挖掘分析，希望能对我国人工智能的发展起到借鉴参考作用。以下各章将对各人工智能领域的基本概念、发展历史、人才情况、代表性论文解读以及近期重要进展进行详细介绍。

2 机器学习

2.1 机器学习概念

机器学习已经成为了当今的热门话题，但是从机器学习这个概念诞生到机器学习技术的普遍应用经过了漫长的过程。在机器学习发展的历史长河中，众多优秀的学者为推动机器学习的发展做出了巨大的贡献。

从 1642 年 Pascal 发明的手摇式计算机，到 1949 年 Donald Hebb 提出的赫布理论——解释学习过程中大脑神经元所发生的变化，都蕴含着机器学习思想的萌芽。

事实上，1950 年图灵在关于图灵测试的文章中就已提及机器学习的概念。到了 1952 年，IBM 的亚瑟·塞缪尔（Arthur Samuel，被誉为“机器学习之父”）设计了一款可以学习的西洋跳棋程序。它能够通过观察棋子的走位来构建新的模型，用来提高自己的下棋技巧。塞缪尔和这个程序进行多场对弈后发现，随着时间的推移，程序的棋艺变得越来越好[1]。塞缪尔用这个程序推翻了以往“机器无法超越人类，不能像人一样写代码和学习”这一传统认识。并在 1956 年正式提出了“机器学习”这一概念。他认为“机器学习是在不直接针对问题进行编程的情况下，赋予计算机学习能力的一个研究领域”。

对机器学习的认识可以从多个方面进行，有着“全球机器学习教父”之称的 Tom Mitchell 则将机器学习定义为：对于某类任务 T 和性能度量 P，如果计算机程序在 T 上以 P 衡量的性能随着经验 E 而自我完善，就称这个计算机程序从经验 E 学习。这些定义都比较简单抽象，但是随着对机器学习了解的深入，我们会发现随着时间的变迁，机器学习的内涵和外延在不断的变化。因为涉及到的领域和应用很广，发展和变化也相当迅速，简单明了地给出“机器学习”这一概念的定义并不是那么容易。

普遍认为，机器学习（Machine Learning，常简称为 ML）的处理系统和算法是主要通过找出数据里隐藏的模式进而做出预测的识别模式，它是人工智能（Artificial Intelligence，常简称为 AI）的一个重要子领域，而人工智能又与更广泛的数据挖掘（Data Mining，常简称为 DM）和知识发现（Knowledge Discovery

in Database，常简称为 KDD）领域相交叉。为了更好的理解和区分人工智能（Artificial Intelligence）、机器学习（Machine Leaning）、数据挖掘（Data Mining）、模式识别（Pattern Recognition）、统计（Statistics）、神经计算（Neuro Computing）、数据库（Databases）、知识发现（KDD）等概念，特绘制其交叉关系如下图所示：

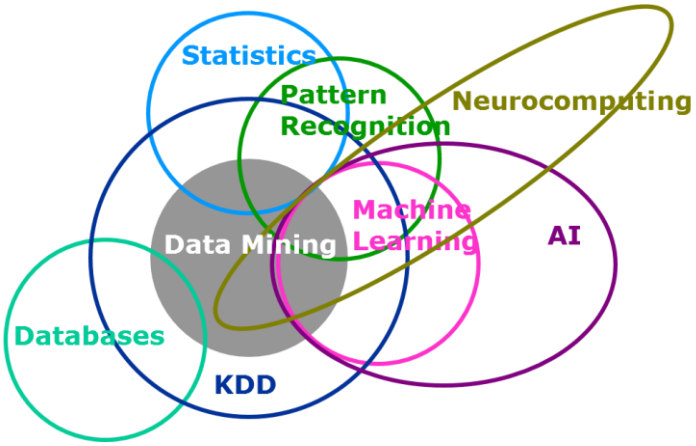


图 2-1 机器学习相关概念的辨识

机器学习是一门多领域交叉学科，涉及概率论、统计学、逼近论、凸分析、算法复杂度理论等多门学科。专门研究计算机怎样模拟或实现人类的学习行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能。其过程可以用下图简单表示：

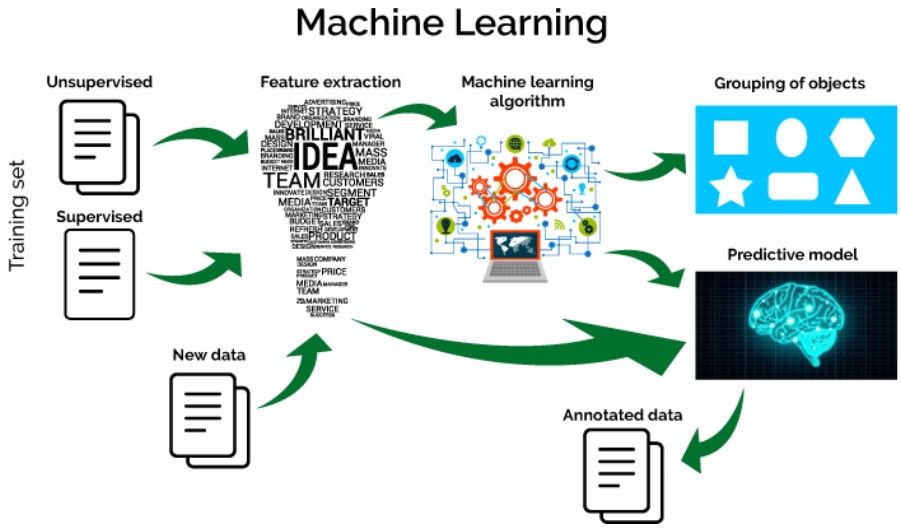


图 2-2 机器学习基本过程

2.2 机器学习发展历史

- 奠基时期

1950 年，阿兰·图灵创造了图灵测试来判定计算机是否智能。图灵测试认为，如果一台机器能够与人类展开对话（通过电传设备）而不能被辨别出其机器身份，那么称这台机器具有智能。这一简化使得图灵能够令人信服地说明“思考的机器”是可能的。

1952，IBM 科学家亚瑟·塞缪尔开发的跳棋程序。驳倒了普罗维登斯提出的机器无法超越人类的论断，像人类一样写代码和学习的模式，他创造了“机器学习”这一术语，并将它定义为：“可以提供计算机能力而无需显式编程的研究领域”。

- 瓶颈时期

从 60 年代中到 70 年代末，机器学习的发展步伐几乎处于停滞状态。无论是理论研究还是计算机硬件限制，使得整个人工智能领域的发展都遇到了很大的瓶颈。虽然这个时期温斯顿（Winston）的结构学习系统和海斯·罗思（Hayes Roth）等的基于逻辑的归纳学习系统取得较大的进展，但只能学习单一概念，而且未能投入实际应用。而神经网络学习机因理论缺陷也未能达到预期效果而转入低潮。

- 重振时期

伟博斯在 1981 年的神经网络反向传播（BP）算法中具体提出多层感知机模型。虽然 BP 算法早在 1970 年就已经以“自动微分的反向模型（reverse mode of automatic differentiation）”为名提出来了，但直到此时才真正发挥效用，并且直到今天 BP 算法仍然是神经网络架构的关键因素。有了这些新思想，神经网络的研究又加快了。在 1985-1986 年，神经网络研究人员相继提出了使用 BP 算法训练的多参数线性规划（MLP）的理念，成为后来深度学习的基石。在另一个谱系中，昆兰在 1986 年提出了一种非常出名的机器学习算法，我们称之为“决策树”，更具体的说是 ID3 算法。在 ID3 算法提出来以后，研究社区已经探索了许多改进（如 ID4、回归树、CART 算法等），这些算法至今仍然活跃在机器学习领域中。

- 成型时期

支持向量机（SVM）的出现是机器学习领域的另一大重要突破，算法具有非常强大的理论地位和实证结果。那一段时间机器学习研究也分为神经网络（Neural Network，NN）和 SVM 两派。然而，在 2000 年左右提出了带核函数的支持向量机后，SVM 在许多以前由 NN 占优的任务中获得了更好的效果。此外，SVM 相对于 NN 还能利用所有关于凸优化、泛化边际理论和核函数的深厚知识。因此 SVM 可以从不同的学科中大力推动理论和实践的改进。

- 爆发时期

神经网络研究领域领军者 Hinton 在 2006 年提出了神经网络 Deep Learning 算法，使神经网络的能力大大提高，向支持向量机发出挑战。2006 年，Hinton 和他的学生 Salakhutdinov 在顶尖学术刊物《Science》上发表了一篇文章，开启了深度学习在学术界和工业界的浪潮。2015 年，为纪念人工智能概念提出 60 周年，LeCun、Bengio 和 Hinton 推出了深度学习的联合综述。深度学习可以让那些拥有多个处理层的计算模型来学习具有多层次抽象的数据的表示，这些方法在许多方面都带来了显著的改善。深度学习的出现，让图像、语音等感知类问题取得了真正意义上的突破，离实际应用已如此之近[2]，将人工智能推进到一个新时代。

2.3 机器学习经典算法

机器学习算法可以按照不同的标准来进行分类。比如按函数 $f(x, \theta)$的不同，机器学习算法可以分为线性模型和非线性模型；按照学习准则的不同，机器学习算法也可以分为统计方法和非统计方法。

但一般来说，我们会按照训练样本提供的信息以及反馈方式的不同，将机器学习算法分为以下几类：

- 监督学习（Supervised Learning）

监督学习中的数据集是有标签的，就是说对于给出的样本我们是知道答案的。如果机器学习的目标是通过建模样本的特征 x 和标签 y 之间的关系：$f(x, \theta)$或 $p(y|x, \theta)$，并且训练集中每个样本都有标签，那么这类机器学习称为监督学习。根

据标签类型的不同，又可以将其分为分类问题和回归问题两类。前者是预测某一样东西所属的类别（离散的），比如给定一个人的身高、年龄、体重等信息，然后判断性别、是否健康等；后者则是预测某一样本所对应的实数输出（连续的），比如预测某一地区人的平均身高。我们大部分学到的模型都是属于监督学习，包括线性分类器、支持向量机等。常见的监督学习算法有：*k*-近邻算法（*k*-Nearest Neighbors，kNN）、决策树（Decision Trees）、朴素贝叶斯（Naive Bayesian）等。监督学习的基本流程如下图所示：

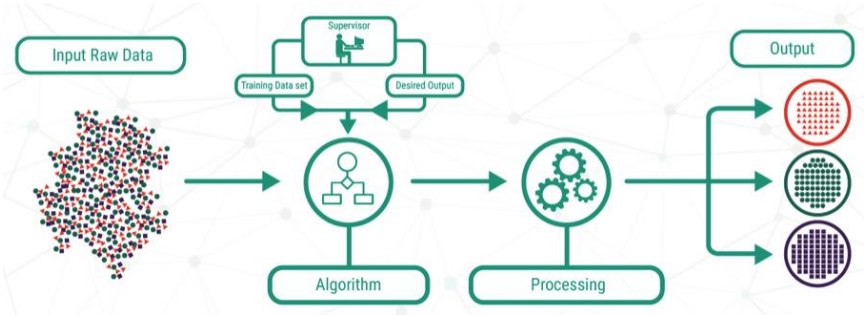


图 2-3 监督学习的基本流程

- 无监督学习（Unsupervised Learning, UL）

跟监督学习相反，无监督学习中数据集是完全没有标签的，依据相似样本在数据空间中一般距离较近这一假设，将样本分类。

常见的无监督学习算法包括：稀疏自编码（sparse auto-encoder）、主成分分析（Principal Component Analysis, PCA）、K-Means 算法（K 均值算法）、DBSCAN 算法（Density-Based Spatial Clustering of Applications with Noise）、最大期望算法（Expectation-Maximization algorithm, EM）等。

利用无监督学习可以解决的问题可以分为关联分析、聚类问题和维度约减。

关联分析是指发现不同事物之间同时出现的概率。在购物篮分析中被广泛地应用。如果发现买面包的客户有百分之八十的概率买鸡蛋，那么商家就会把鸡蛋和面包放在相邻的货架上。

聚类问题是指将相似的样本划分为一个簇（cluster）。与分类问题不同，聚类问题预先并不知道类别，自然训练数据也没有类别的标签。

维度约减：顾名思义，是指减少数据维度的同时保证不丢失有意义的信息。利用特征提取方法和特征选择方法，可以达到维度约减的效果。特征选择是指选择原始变量的子集。特征提取是将数据从高维度转换到低维度。广为熟知的主成分分析算法就是特征提取的方法。

非监督学习的基本处理流程如下图所示：

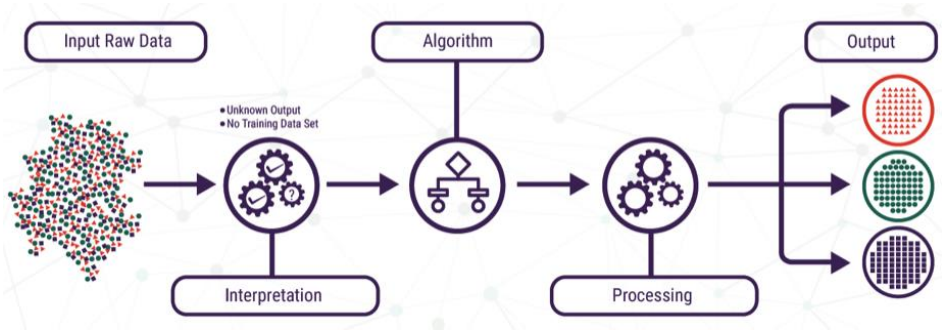


图 2-4 非监督学习的基本流程

可以很清楚的看到相对于监督学习，非监督学习的过程中没有监督者（Supervisor）的干预。下图是一个典型的监督学习和非监督学习的对比，左图是对一群有标签数据的分类，而右图是对一群无标签数据的聚类。

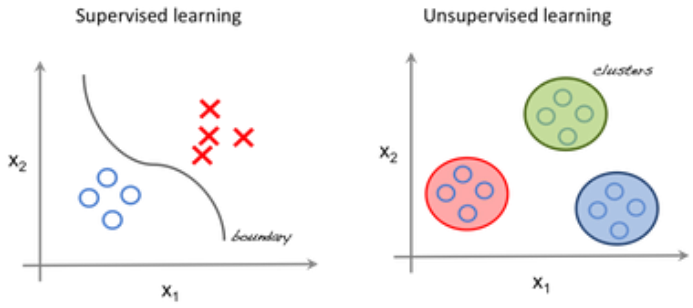


图 2-5 一个典型的监督学习和非监督学习对比

- 半监督学习（Semi-Supervised Learning）

半监督学习是监督学习与无监督学习相结合的一种学习方法。半监督学习一般针对的问题是数据量大，但是有标签数据少或者说标签数据的获取很难很贵的情况，训练的时候有一部分是有标签的，而有一部分是没有的。与使用所有标签数据的模型相比，使用训练集的训练模型在训练时可以更为准确，而且训练成本更低。常见的两种半监督的学习方式是直推学习（Transductive learning）和归纳学习（Inductive learning）。

直推学习（Transductive learning）：没有标记的数据是测试数据，这个时候可以用测试的数据进行训练。这里需要注意，这里只是用了测试数据中的特征（feature）而没有用标签（label），所以并不是一种欺骗的方法。

归纳学习（Inductive learning）：没有标签的数据不是测试集。

半监督学习的基本流程如图 2-6 所示：

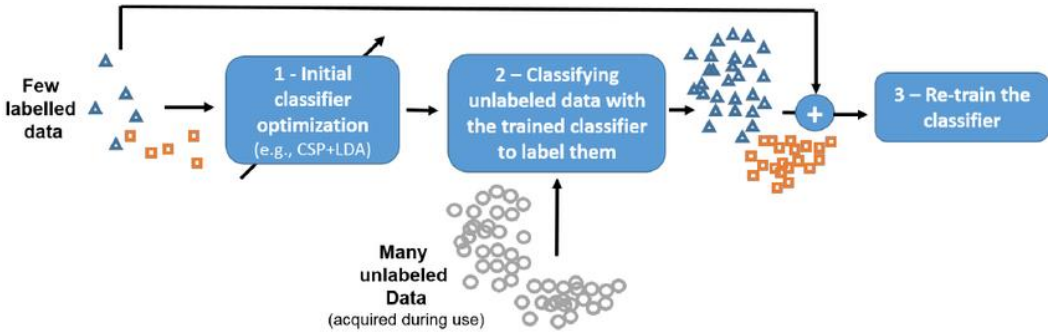


图 2-6 半监督学习的基本流程

监督学习、半监督学习和非监督学习之间的区别可以用图 2-7 表示：

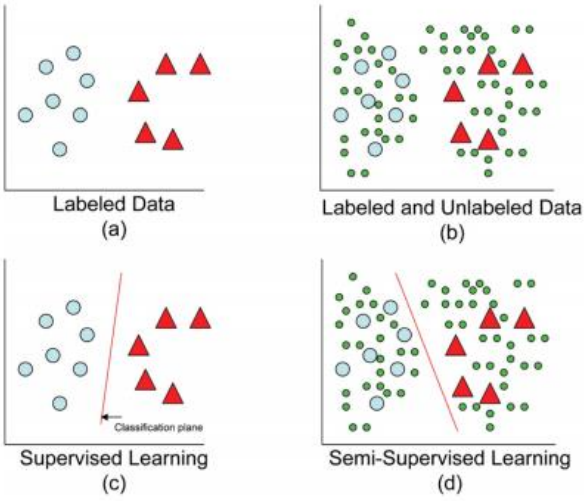


图 2-7 监督学习、半监督学习和非监督学习的简单对比

可以看到，图 2-7（a）中，红色三角形数据和蓝色圆点数据为标注数据；图 2-7（b）中，绿色的小圆点为非标注数据。图 2-7（c）显示监督学习将有标签的数据进行分类；而半监督学习如图 2-7（d）中部分是有标签的，部分是没有标签的，一般而言，半监督学习侧重于在有监督的分类算法中加入无标记样本来实现半监督分类。

- 强化学习（Reinforcement Learning, RL）

强化学习从动物学习、参数扰动自适应控制等理论发展而来，基本原理是：如果 Agent 的某个行为策略导致环境正的奖赏(强化信号)，那么 Agent 以后产生这个行为策略的趋势便会加强。Agent 的目标是在每个离散状态发现最优策略以使期望的折扣奖赏和最大。

强化学习在机器人学科中被广泛应用。在与障碍物碰撞后，机器人通过传感器收到负面的反馈从而学会去避免冲突。在视频游戏中，可以通过反复试验采用一定的动作，获得更高的分数。Agent 能利用回报去理解玩家最优的状态和当前应该采取的动作。

下图采用一只老鼠来模拟强化学习中的 Agent，其任务是走出迷宫，每走一步都有一个方法来衡量其走的好与坏，基本学习过程是当其走得好的时候就给其一定的奖励（如一块蛋糕）。通过这种方式，Agent 在行动评价的环境中获得知识，改进行动方案以适应环境。

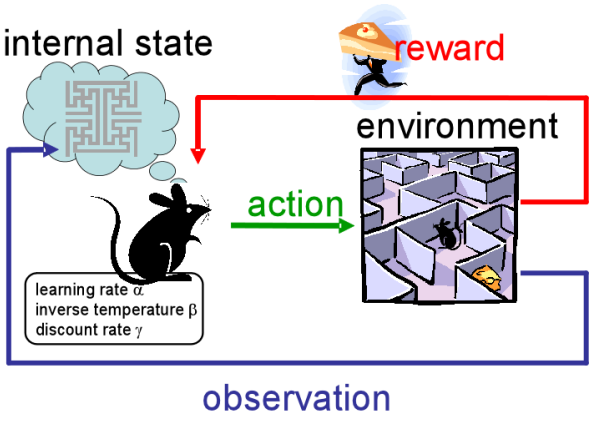


图 2-8 强化学习的基本学习流程

下面内容对部分机器学习代表算法进行了介绍。

- 线性回归

在机器学习中，我们有一组输入变量（x）用于确定输出变量（y）。输入变量和输出变量之间存在某种关系，机器学习的目标是量化这种关系。

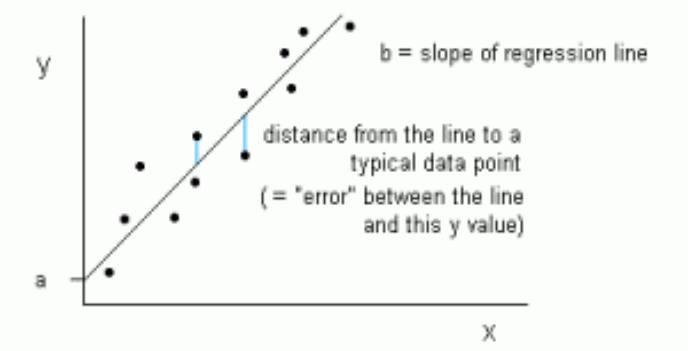


图 2-9 数据集的绘制 x 和 y 值

在线性回归中，输入变量（x）和输出变量（y）之间的关系表示为 $y = ax + b$ 的方程。因此，线性回归的目标是找出系数 a 和 b 的值。这里，a 是直线的斜率，b 是直线的截距。上图显示了数据集的 x 和 y 值，线性回归的目标是拟合最接近大部分点的线。

- 分类与回归树（CART）

CART 是决策树的一个实现方式，由 ID3，C4.5 演化而来，是许多基于树的 bagging、boosting 模型的基础。CART 可用于分类与回归。

CART 是在给定输入随机变量 x 条件下输出随机变量 y 的条件概率分布，与 ID3 和 C4.5 的决策树所不同的是，ID3 和 C4.5 生成的决策树可以是多叉的，每个节点下的叉数由该节点特征的取值种类而定，比如特征年龄分为（青年，中年，老年），那么该节点下可分为 3 叉。而 CART 为假设决策树为二叉树，内部结点特征取值为“是”和“否”。左分支取值为“是”，右分支取值为“否”。这样的决策树等价于递归地二分每一个特征，将输入空间划分为有限个单元，并在这些单元上预测概率分布，也就是在输入给定的条件下输出条件概率分布。

- 随机森林（Random Forest）

随机森林指的是利用多棵决策树对样本进行训练并预测的一种分类器。它包含多个决策树的分类器，并且其输出的类别是由个别树输出的类别的众数而定。随机森林是一种灵活且易于使用的机器学习算法，即便没有超参数调优，也可以在大多数情况下得到很好的结果。随机森林也是最常用的算法之一，因为它很简易，既可用于分类也能用于回归。

其基本的构建算法过程如下：

1. 用 N 来表示训练用例（样本）的个数，M 表示特征数目。

2. 输入特征数目 m，用于确定决策树上一个节点的决策结果；其中 m 应远小于 M。

3. 从 N 个训练用例（样本）中以有放回抽样的方式，取样 N 次，形成一个训练集（即 bootstrap 取样），并用未抽到的用例（样本）作预测，评估其误差。

4. 对于每一个节点，随机选择 m 个特征，决策树上每个节点的决定都是基于这些特征确定的。根据这 m 个特征，计算其最佳的分裂方式。

5. 每棵树都会完整成长而不会剪枝，这有可能在建完一棵正常树状分类器后被采用）。

一个简单的随机森林算法示意如下：

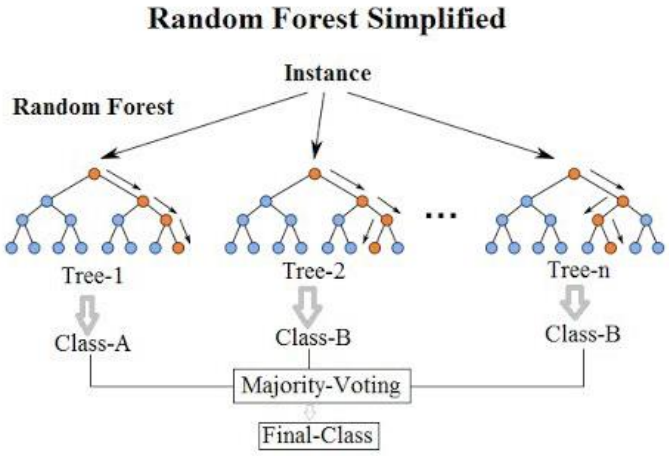


图 2-10 一个简单的随机森林算法示意

随机森林集成了所有的分类投票结果，将投票次数最多的类别指定为最终的输出，这就是一种最简单的 Bagging 思想。

- 逻辑回归

逻辑回归最适合二进制分类（$y = 0$ 或 1 的数据集，其中 1 表示默认类）例如：在预测事件是否发生时，发生的事件被分类为 1。在预测人会生病或不生病，生病的实例记为 1）。它是以其中使用的变换函数命名的，称为逻辑函数 $h(x) = 1 / (1+e^{-x})$，它是一个 S 形曲线。

在逻辑回归中，输出是以缺省类别的概率形式出现的。因为这是一个概率，所以输出在 0-1 的范围内。输出（*y* 值）通过对数转换 *x* 值，使用对数函数 $h(x) = 1 /（1+e^{-x}）$来生成。然后应用一个阈值来强制这个概率进入二元分类。

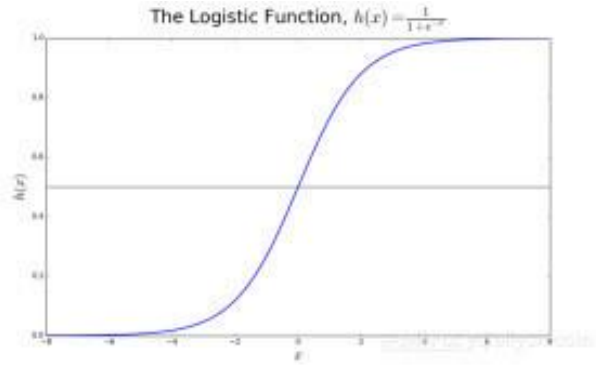


图 2-11 逻辑函数曲线图

上图判断了肿瘤是恶性还是良性。默认变量是 *y* = 1（肿瘤=恶性）；*x* 变量可以是肿瘤的信息，例如肿瘤的尺寸。如图所示，逻辑函数将数据集的各种实例的 *x* 值转换成 0 到 1 的范围。如果概率超过阈值 0.5（由水平线示出），则将肿瘤分类为恶性。

$$p(x) = \frac{e^{b_0 + b_1 x}}{1 + e^{b_0 + b_1 x}}$$

$$\log\left(\frac{p(x)}{1-p(x)}\right) = b_0 + b_1 x$$

逻辑回归的目标是使用训练数据来找到系数 b_0 和 b_1 的值，以使预测结果与实际结果之间的误差最小化。这些系数是使用最大似然估计来计算的。

- 朴素贝叶斯（Naive Bayesian）

朴素贝叶斯法是基于贝叶斯定理与特征条件独立假设的分类方法。朴素贝叶斯分类器基于一个简单的假定：给定目标值时属性之间相互条件独立。

通过以上定理和“朴素”的假定，我们知道：

P(*Category* | *Document*) = *P*(*Document* | *Category*) * *P*(*Category*) / *P*(*Document*)

朴素贝叶斯的基本方法：在统计数据的基础上，依据条件概率公式，计算当前特征的样本属于某个分类的概率，选择最大的概率分类。

对于给出的待分类项，求解在此项出现的条件下各个类别出现的概率，哪个最大，就认为此待分类项属于哪个类别。其计算流程表述如下：

（1）$x = \{a_1, a_2, ..., a_m\}$为待分类项，每个 a_i 为 x 的一个特征属性

（2）有类别集合 $C = \{y_1, y_2, ..., y_n\}$

（3）计算 $P(y_1|x), P(y_2|x), ..., P(y_n|x)$

（4）如果 $P(y_k|x) = max\{P(y_1|x)$

- k 最近邻（kNN）

kNN（*k*-Nearest Neighbor）的核心思想是如果一个样本在特征空间中的 *k* 个最相邻的样本中的大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性。该方法在确定分类决策上只依据最邻近的一个或者几个样本的类别来决定待分样本所属的类别。kNN 方法在做类别决策时，只与极少量的相邻样本有关。由于 kNN 方法主要靠周围有限的邻近的样本，而不是靠判别类域的方法来确定所属类别的，因此对于类域的交叉或重叠较多的待分样本集来说，kNN 方法较其他方法更为适合。

kNN 算法不仅可以用于分类，还可以用于回归。通过找出一个样本的 *k* 个最近邻居，将这些邻居的属性的平均值赋给该样本，就可以得到该样本的属性。如下图是 kNN 算法中，*k* 等于不同值时的算法分类结果：

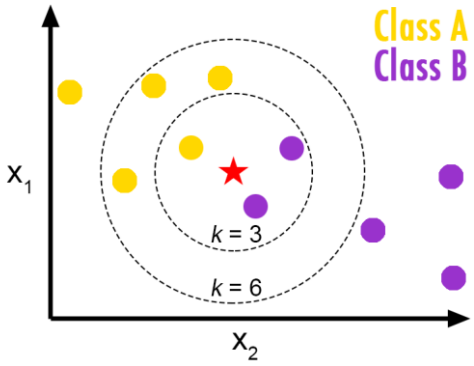


图 2-12 kNN 算法简单示例

简单来说，kNN 可以看成：有那么一堆你已经知道分类的数据，然后当一个新数据进入的时候，就开始跟训练数据里的每个点求距离，然后挑离这个训练数

据最近的 k 个点，看看这几个点属于什么类型，然后用少数服从多数的原则，给新数据归类。

- AdaBoost

Adaptive Boosting 或称为 AdaBoost，是多种学习算法的融合。它是一种迭代算法，其核心思想是针对同一个训练集训练不同的分类器(弱分类器)，然后把这些弱分类器集合起来，构成一个更强的最终分类器(强分类器)。其算法本身是通过改变数据分布来实现的，它根据每次训练集之中每个样本的分类是否正确，以及上次的总体分类的准确率，来确定每个样本的权值。将修改过权值的新数据集送给下层分类器进行训练，然后将每次训练得到的分类器融合起来，作为最终的决策分类器。

AdaBoost 是最常用的算法。它可用于回归或者分类算法。相比其他机器学习算法，它克服了过拟合的问题，通常对异常值和噪声数据敏感。为了创建一个强大的复合学习器，AdaBoost 使用了多次迭代。因此，它又被称为“Adaptive Boosting”。通过迭代添加弱学习器，AdaBoost 创建了一个强学习器。一个新的弱学习器加到实体上，并且调整加权向量，作为对前一轮中错误分类的样例的回应。得到的结果，是一个比弱分类器有更高准确性的分类器。

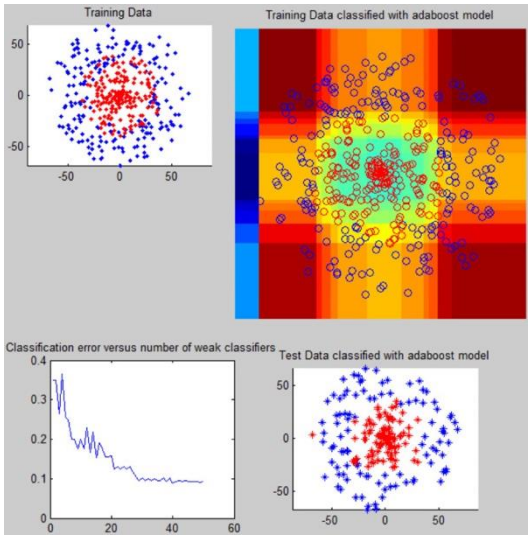


图 2-13 AdaBoost 执行

AdaBoost 有助于将弱阈值的分类器提升为强分类器。上面的图像描述了 AdaBoost 的执行，只用了简单易于理解的代码在一个文件中就实现了。这个函

数包含一个弱分类器和 boosting 组件。弱分类器在一维的数据中尝试去寻找最理想的阈值来将数据分离为两类。boosting 组件迭代调用分类器，经过每一步分类，它改变了错误分类示例的权重。因此，创建了一个级联的弱分类器，它的行为就像一个强分类器。

目前，对 Adaboost 算法的研究以及应用大多集中于分类问题，同时近年也出现了一些在回归问题上的应用。Adaboost 系列主要解决了：两类问题、多类单标签问题、多类多标签问题、大类单标签问题和回归问题。它用全部的训练样本进行学习。

- K-均值算法（K-Means）

K-均值是著名聚类算法，它找出代表聚类结构的 k 个质心。如果有一个点到某一质心的距离比到其他质心都近，这个点则指派到这个最近的质心所代表的簇。依次，利用当前已聚类的数据点找出一个新质心，再利用质心给新的数据指派一个簇。

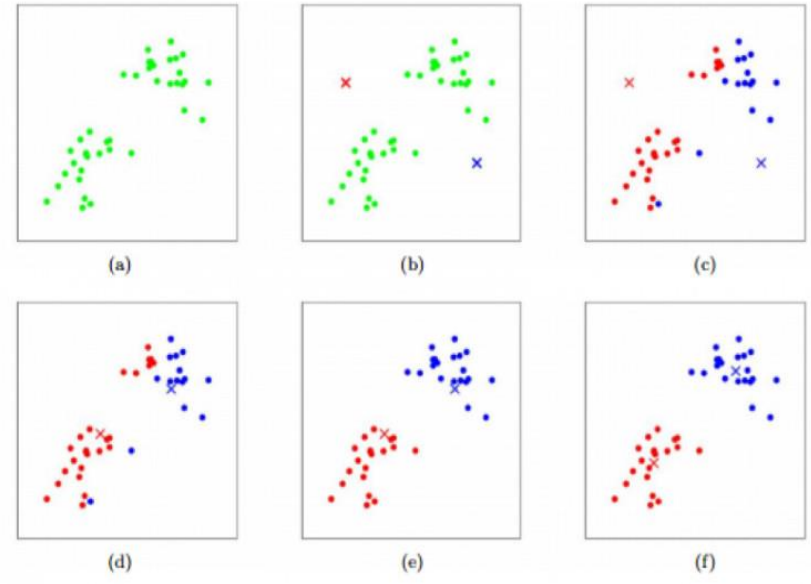


图 2-14 K-均值算法图示

K-均值算法——在上图中用“x”表示聚类质心，用点表示训练样本：

a) 原始数据集

b) 随机初始化聚类质心

c) (*c-f*)*k*-均值迭代 2 次的示意图

在每次迭代中每个训练样例都被指派到一个最近的聚类质心，每个聚类质心被移动到分配给它的点的平均值的位置。

- 支持向量机（SVM）

支持向量机（Support Vector Machine, SVM）是一类按监督学习（supervised learning）方式对数据进行二元分类（binary classification）的广义线性分类器（generalized linear classifier），其决策边界是对学习样本求解的最大边距超平面（maximum-margin hyperplane）。基本思想是：找到集合边缘上的若干数据（称为支持向量（Support Vector）），用这些点找出一个平面（称为决策面），使得支持向量到该平面的距离最大。由简至繁的 SVM 模型包括：

- 当训练样本线性可分时，通过硬间隔最大化，学习一个线性可分支持向量机；
- 当训练样本近似线性可分时，通过软间隔最大化，学习一个线性支持向量机；
- 当训练样本线性不可分时，通过核技巧和软间隔最大化，学习一个非线性支持向量机；

在分类问题中，很多时候有多个解，如下图左边所示，在理想的线性可分的情况下其决策平面会有多个。而 SVM 的基本模型是在特征空间上找到最佳的分离超平面使得训练集上正负样本间隔最大，SVM 算法计算出来的分界会保留对类别最大的间距，即有足够的余量，如下图右边所示。

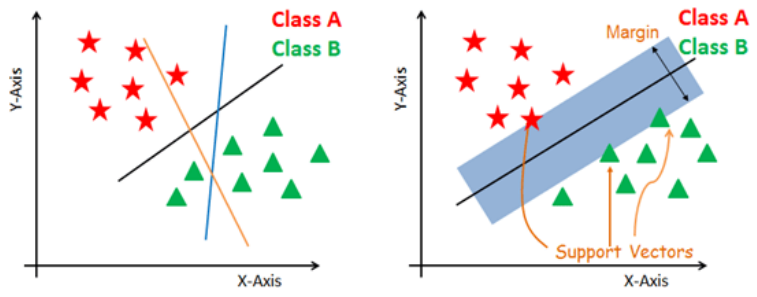


图 2-15 SVM 的决策平面

在解决线性不可分问题时，它可以通过引入核函数，巧妙地解决了在高维空间中的内积运算，从而很好地解决了非线性分类问题。如下图所示，通过核函数的引入，将线性不可分的数据映射到一个高纬的特征空间内，使得数据在特征空间内是可分的。如下图所示：

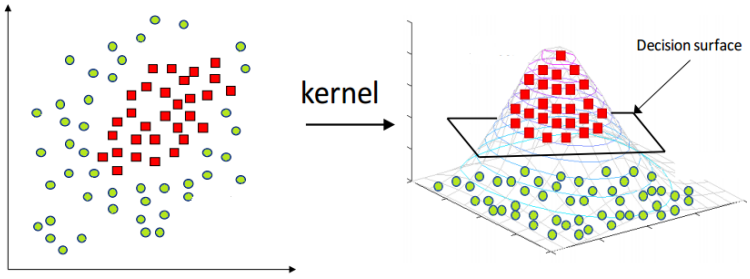


图 2-16 SVM 的核函数

- 人工神经网络 ANN（Artificial Neural Network）

人工神经网络 ANN（Artificial Neural Network）是由大量处理单元互联组成的非线性、自适应信息处理系统。它是一种模仿动物神经网络行为特征，进行分布式并行信息处理的算法数学模型。其基本过程可以概述如下：外部刺激通过神经末梢，转化为电信号，传导到神经细胞（又叫神经元）；无数神经元构成神经中枢；神经中枢综合各种信号，做出判断；人体根据神经中枢的指令，对外部刺激做出反应。其过程表述如下图所示：

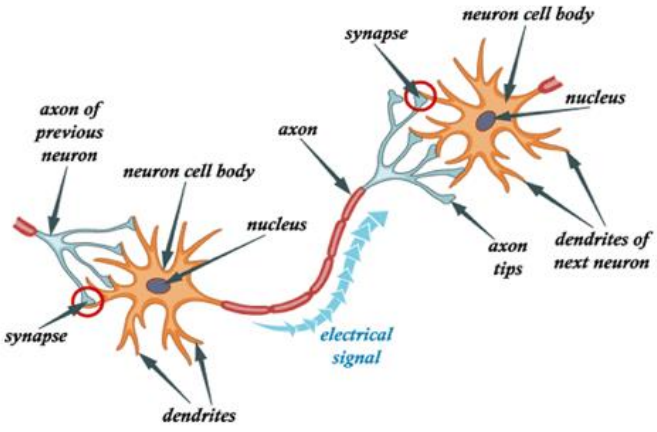


图 2-17 神经网络的传导过程

人工神经网络经历了漫长的发展阶段。最早是上个世纪六十年代提出的“人造神经元”模型，叫做“感知器”（perceptron）。感知机模型是机器学习二分类问题中的一个非常简单的模型。它的基本结构如下图所示：

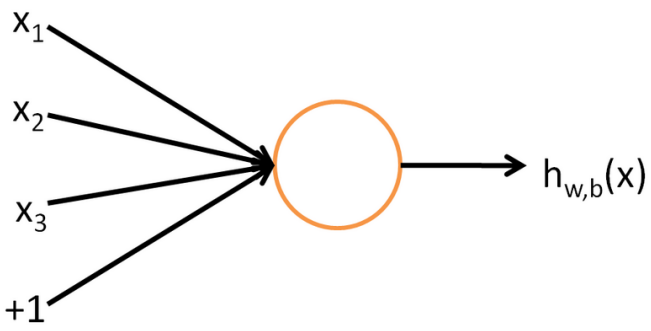


图 2-18 感知机

随着反向传播算法、最大池化（max-pooling）等技术的发明，神经网络进入了飞速发展的阶段。神经网络就是将许多个单一“神经元”联结在一起，这样，一个“神经元”的输出就可以是另一个“神经元”的输入。典型的人工神经网络具有以下三个部分：

结构（Architecture）结构指定了网络中的变量和它们的拓扑关系。

激励函数（Activity Rule）大部分神经网络模型具有一个短时间尺度的动力学规则，来定义神经元如何根据其他神经元的活动来改变自己的激励值。

学习规则（Learning Rule）指定了网络中的权重如何随着时间推进而调整。

一个典型的人工神经网络结构如下图所示：

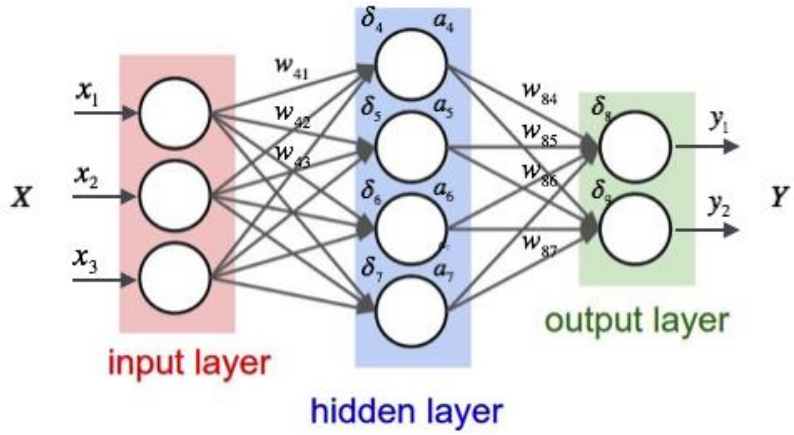


图 2-19 典型的人工神经网络结构

人工神经网络具有四个基本特征：非线性、非局限性、非常定性和非凸性。

人工神经网络的特点和优越性，主要表现在三个方面：具有自学习功能、具有联想存储功能和具有高速寻找最优解的能力。

2.4 深度学习

深度学习是近 10 年机器学习领域发展最快的一个分支，由于其重要性，三位教授（Geoffrey Hinton、Yann Lecun、Yoshua Bengio）因此同获图灵奖。深度学习模型的发展可以追溯到 1958 年的感知机（Perceptron）。1943 年神经网络就已经出现雏形（源自 NeuroScience），1958 年研究认知的心理学家 Frank 发明了感知机，当时掀起一股热潮。后来 Marvin Minsky（人工智能大师）和 Seymour Papert 发现感知机的缺陷：不能处理异或回路等非线性问题，以及当时存在计算能力不足以处理大型神经网络的问题。于是整个神经网络的研究进入停滞期。

最近 30 年来取得快速发展。总体来说，主要有 4 条发展脉络。

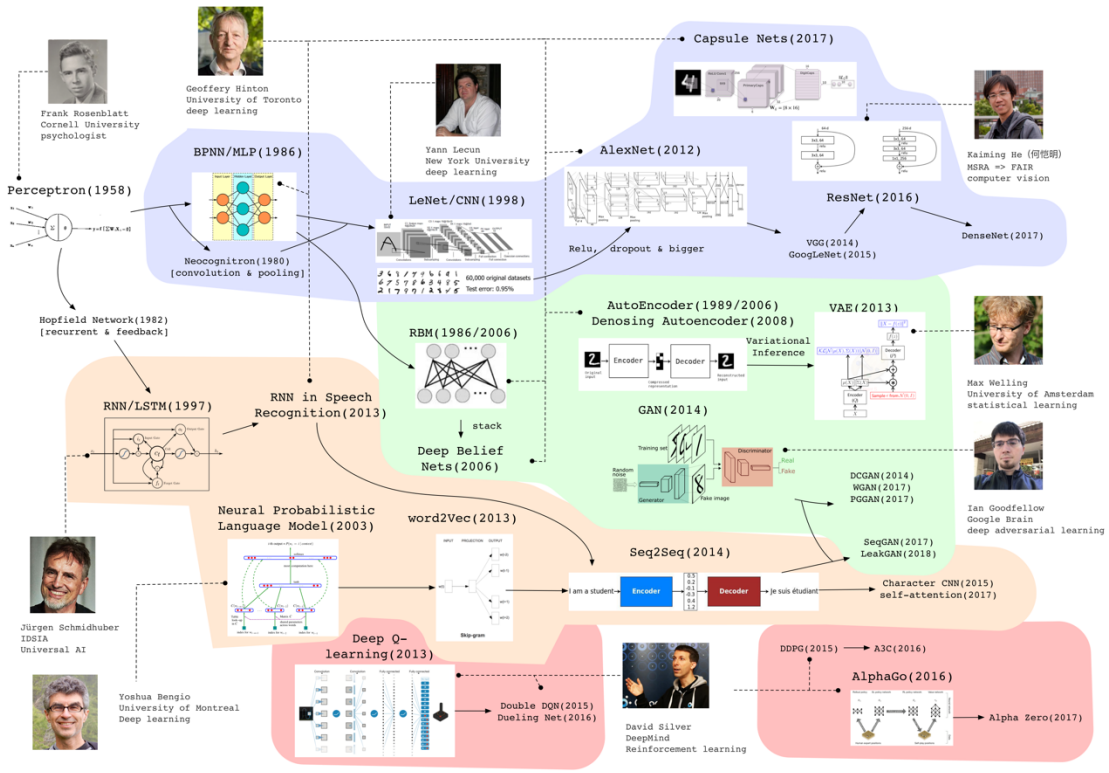


图 2-20 深度学习模型最近若干年的重要进展

第一个发展脉络（上图浅紫色区域）以计算机视觉和卷积网络为主。这个脉络的进展可以追溯到 1979 年，Fukushima 提出的 Neocognitron。该研究给出了卷积和池化的思想。1986 年 Hinton 提出的反向传播训练 MLP（之前也有几个类似的研究），该研究解决了感知机不能处理非线性学习的问题。1998 年，以 Yann LeCun 为首的研究人员实现了一个七层的卷积神经网络 LeNet-5 以识别手写数字。现在普遍把 Yann LeCun 的这个研究作为卷积网络的源头，但其实在当时由于 SVM 的迅速崛起，这些神经网络的方法还没有引起广泛关注。真正使得卷积神经网络荣耀登上大雅之堂的事件是，2012 年 Hinton 组的 AlexNet（一个设计精巧的 CNN）在 ImageNet 上以巨大优势夺冠，这引发了深度学习的热潮。AlexNet 在传统 CNN 的基础上加上了 ReLU、Dropout 等技巧，并且网络规模更大。这些技巧后来被证明非常有用，成为卷积神经网络的标配，被广泛发展，于是后来出现了 VGG、GoogLenet 等新模型。2016 年，青年计算机视觉科学家何恺明在层次之间加入跳跃连接，提出残差网络 ResNet。ResNet 极大增加了网络深度，效果有很大提升。一个将这个思路继续发展下去的是近年的 CVPR Best Paper 中黄高提出的 DenseNet。在计算机视觉领域的特定任务出现了各种各样的模型（Mask-RCNN 等），这里不一一介绍。2017 年，Hinton 认为反向传播和传统神经网络还存在一定缺陷，因此提出 Capsule Net，该模型增强了可解释性，但目前在 CIFAR 等数据集上效果一般，这个思路还需要继续验证和发展。

第二个发展脉络（上图浅绿色区域）以生成模型为主。传统的生成模型是要预测联合概率分布 $P(x, y)$。机器学习方法中生成模型一直占据着一个非常重要的地位，但基于神经网络的生成模型一直没有引起广泛关注。Hinton 在 2006 年的时候基于受限玻尔兹曼机（RBM，一个 19 世纪 80 年代左右提出的基于无向图模型的能量物理模型）设计了一个机器学习的生成模型，并且将其堆叠成为 Deep Belief Network，使用逐层贪婪或者 wake-sleep 的方法训练，当时模型的效果其实并没有那么好。但值得关注的是，正是式基于 RBM 模型 Hinton 等人开始设计深度框架，因此这也可以看做深度学习的一个开端。Auto-Encoder 也是上个世纪 80 年代 Hinton 就提出的模型，后来随着计算能力的进步也重新登上舞台。Bengio 等人又提出了 Denoise Auto-Encoder，主要针对数据中可能存在的噪音问题。Max Welling（这也是变分和概率图模型的高手）等人后来使用神经网络训练一个有一

层隐变量的图模型，由于使用了变分推断，并且最后长得跟 Auto-Encoder 有点像，被称为 Variational Auto-Encoder。此模型中可以通过隐变量的分布采样，经过后面的 Decoder 网络直接生成样本。生成对抗模型 GAN（Generative Adversarial Network）是 2014 年提出的非常火的模型，它是一个通过判别器和生成器进行对抗训练的生成模型，这个思路很有特色，模型直接使用神经网络 G 隐式建模样本整体的概率分布，每次运行相当于从分布中采样。后来引起大量跟随的研究，包括：DCGAN 是一个相当好的卷积神经网络实现，WGAN 是通过维尔斯特拉斯距离替换原来的 JS 散度来度量分布之间的相似性的工作，使得训练稳定。PGGAN 逐层增大网络，生成逼真的人脸。

第三个发展脉络（上图橙黄色区域）是序列模型。序列模型不是因为深度学习才有的，而是很早以前就有相关研究，例如有向图模型中的隐马尔科夫 HMM 以及无向图模型中的条件随机场模型 CRF 都是非常成功的序列模型。即使在神经网络模型中，1982 年就提出了 Hopfield Network，即在神经网络中加入了递归网络的思想。1997 年 Jürgen Schmidhuber 发明了长短期记忆模型 LSTM（Long-Short Term Memory），这是一个里程碑式的工作。当然，真正让序列神经网络模型得到广泛关注的还是 2013 年 Hinton 组使用 RNN 做语音识别的工作，比传统方法高出一大截。在文本分析方面，另一个图灵奖获得者 Yoshua Bengio 在 SVM 很火的时期提出了一种基于神经网络的语言模型（当然当时机器学习还是 SVM 和 CRF 的天下），后来 Google 提出的 word2vec（2013）也有一些反向传播的思想，最重要的是给出了一个非常高效的实现，从而引发这方面研究的热潮。后来，在机器翻译等任务上逐渐出现了以 RNN 为基础的 seq2seq 模型，通过一个 Encoder 把一句话的语义信息压成向量再通过 Decoder 转换输出得到这句话的翻译结果，后来该方法被扩展到和注意力机制（Attention）相结合，也大大扩展了模型的表示能力和实际效果。再后来，大家发现使用以字符为单位的 CNN 模型在很多语言任务也有不俗的表现，而且时空消耗更少。Self-attention 实际上就是采取一种结构去同时考虑同一序列局部和全局的信息，Google 有一篇很有名的文章“attention is all you need”把基于 Attention 的序列神经模型推向高潮。当然 2019 年 ACL 上同样有另一篇文章给这一研究也稍微降了降温。

第四个发展脉络（上图粉色区域）是增强学习。这个领域最出名的当属 Deep Mind，图中标出的 David Silver 博士是一直研究 RL 的高管。Q-learning 是很有名的传统 RL 算法，Deep Q-learning 将原来的 Q 值表用神经网络代替，做了一个打砖块的任务。后来又应用在许多游戏场景中，，并将其成果发表在 Nature 上。Double Dueling 对这个思路进行了一些扩展，主要是 Q-Learning 的权重更新时序上。DeepMind 的其他工作如 DDPG、A3C 也非常有名，它们是基于 Policy Gradient 和神经网络结合的变种。大家都熟知的 AlphaGo，里面其实既用了 RL 的方法也有传统的蒙特卡洛搜索技巧。Deep Mind 后来提出了的一个用 AlphaGo 框架，但通过主学习来玩不同（棋类）游戏的新算法 Alpha Zero。

下面对深度学习的不同方面进行分别解读。有些地方解读可能稍微会简单一些，不完整的地方还请见谅。

2.4.1 卷积神经网络

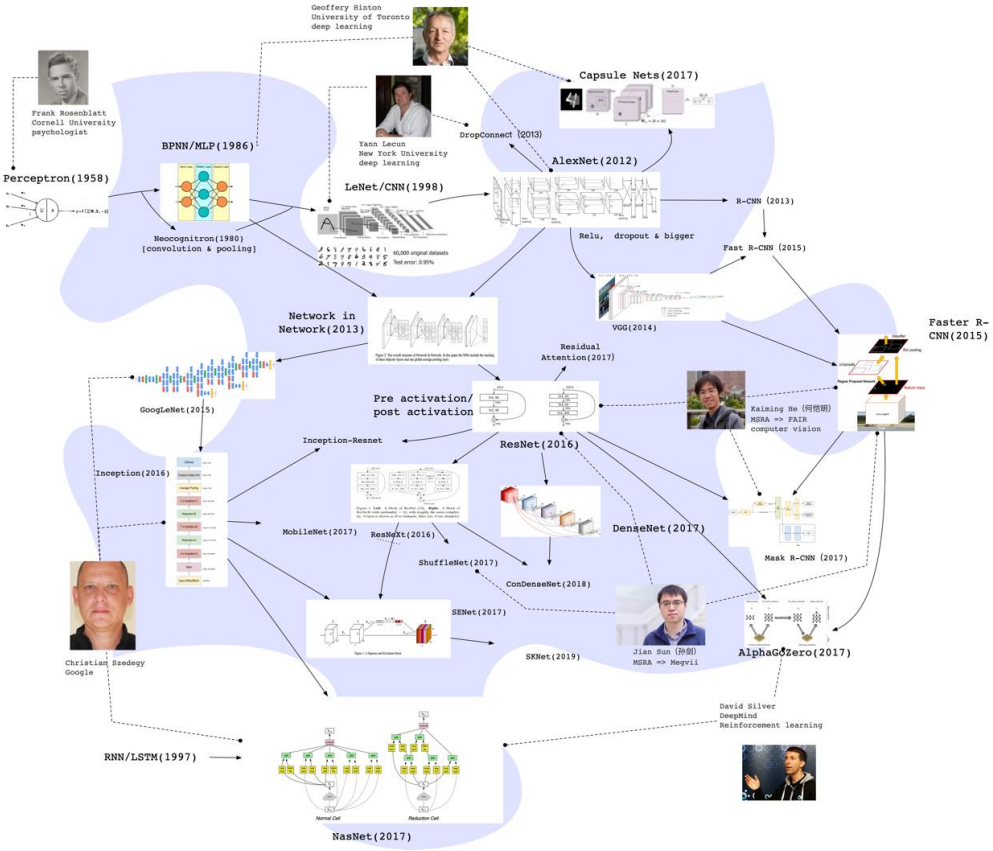


图 2-21 卷积神经网络的重要进展

卷积神经网络的发展，最早可以追溯到 1962 年，Hubel 和 Wiesel 对猫大脑中的视觉系统的研究。1980 年，一个日本科学家福岛邦彦（Kunihiko Fukushima）提出了一个包含卷积层、池化层的神经网络结构。在这个基础上，Yann Lecun 将 BP 算法应用到这个神经网络结构的训练上，就形成了当代卷积神经网络的雏形。

其实最初的 CNN 效果并不算好，而且训练也非常困难。虽然也在阅读支票、识别数字之类的任务上有一定的效果，但由于在一般的实际任务中表现不如 SVM、Boosting 等算法好，因此一直处于学术界的边缘地位。直到 2012 年，ImageNet 图像识别大赛中，Hinton 组的 AlexNet 引入了全新的深层结构和 Dropout 方法，一下子把 error rate 从 25%降低到了 15%，这颠覆了图像识别领域。AlexNet 有很多创新，尽管都不是很难的方法。其最主要的结果是让人们意识到原来那个福岛邦彦提出的、Yann LeCun 优化的 LeNet 结构原来是有很大改进空间的：只要通过一些方法能够加深这个网络到 8 层左右，让网络表达能力提升，就能得到出人意料的好结果。

顺着 AlexNet 的思想，LeCun 组 2013 年提出一个 DropConnect，把 error rate 降低到了 11%。而 NUS 的颜水成组则提出了一个重要的 Network in Network（NIN）方法，NIN 的思想是在原来的 CNN 结构中加入了一个 1*1 conv 层，NIN 的应用也得到了 2014 年 Imagine 另一个挑战——图像检测的冠军。Network in Network 更加引发了大家对 CNN 结构改变的大胆创新。因此，两个新的架构 Inception 和 VGG 在 2014 年把网络加深到了 20 层左右，图像识别的 error rate（越小越好）也大幅降低到 6.7%，接近人类错误率的 5.1%。2015 年，MSRA 的任少卿、何恺明、孙剑等人，尝试把 identity 加入到卷积神经网络中提出 ResNet。最简单的 Identity 却出人意料的有效，直接使 CNN 能够深化到 152 层、1202 层等，error rate 也降到了 3.6%。后来，ResNeXt, Residual-Attention，DenseNet，SENet 等也各有贡献，各自引入了 Group convolution，Attention，Dense connection，channelwise-attention 等，最终 ImageNet 上 error rate 降到了 2.2%，大大超过人类的错误率。现在，即使手机上的神经网络，也能达到超过人类的水平。而另一个挑战——图像检测中，也是任少卿、何恺明、孙剑等优化了原先的 R-CNN, fast R-CNN 等通过其他方法提出 region proposal，然后用 CNN 去判断是否是 object 的方法，提出了 faster R-CNN。Faster R-CNN 的主要贡献是使用和图像识别相同

的 CNN feature，发现 feature 不仅可以识别图片内容，还可以用来识别图片的位置。也就是说，CNN 的 feature 非常有用，包含了大量的信息，可以同时用来做不同的任务。这个创新一下子把图像检测的 MAP 也翻倍了。在短短的 4 年中，ImageNet 图像检测的 MAP（越大越好）从最初的 0.22 达到了最终的 0.73。何恺明后来还提出了 Mask R-CNN，即给 faster R-CNN 又加了一个 Mask Head，发现即使只在训练中使用 Mask Head，其信息可以传递回原先的 CNN feature 中，获得了更精细的信息。由此，Mask R-CNN 得到了更好的结果。何恺明在 2009 年时候就以一个简单有效的去雾算法得到了 CVPR Best Paper，在计算机视觉领域声名鹊起。后来更是提出了 ResNet 和 Faster R-CNN 两大创新，直接颠覆了整个计算机视觉/机器学习领域。

当然，CNN 结构变得越来越复杂，很多结构都很难直觉的来解释和设计。于是谷歌提出了自动架构学习方法 NasNet（Neural Architecture Search Network）来自动用 Reinforcement Learning 去搜索一个最优的神经网络结构。Nas 是目前 CV 界一个主流的方向，可以自动寻找出最好的结构，以及给定参数数量/运算量下最好的结构(这样就可以应用于手机)，是目前图像识别的一个重要发展方向。今年何恺明（2019 年 4 月）又发表了一篇论文，表示即使 Random 生成的网络连接结构（只要按某些比较好的 Random 方法），都会取得非常好的效果，甚至比标准的好很多。Random 和 Nas 哪个是真的正确的道路，这有待进一步的研究了。

卷积神经网络 CNN 的发展引发了其他领域的很多变革。比如：利用 CNN，AlphaGo 战胜了李世石，攻破了围棋（基础版本的 AlphaGo 其实和人类高手比起来是有胜有负的）。后来利用了 ResNet 和 Faster-RCNN 的思想，一年后的 Master 则完全战胜了所有人类围棋高手。后来又有很多复现的开源围棋 AI，每一个都能用不大的计算量超过所有的人类高手。以至于现在人们讲棋的时候，都是按着 AI 的胜率来讲了。

2.4.2 AutoEncoder

AutoEncoder 的基本思想是利用神经网络来做无监督学习，就是把样本的输入同时作为神经网络的输入和输出。本质上是希望学习到输入样本的表示（encoding）。早期 AutoEncoder 的研究主要是数据过于稀疏、数据高维导致计

算复杂度高。比较早用神经网络做 AutoEncoder 的可以追溯到 80 年代的 BPNN 和 MLP 以及当时 Hinton 推崇的 RBM。后来到了 2000 年以后还坚持在做的只剩下 Hinton 的 RBM 了。从 2000 年以后，随着神经网络的快速兴起，AutoEncoder 也得到快速发展，基本上有几条线：稀疏 AutoEncoder、噪音容忍 AutoEncoder、卷积 AutoEncoder、变分 AutoEncoder。最新的进展是结合对抗思想的对抗 AutoEncoder。

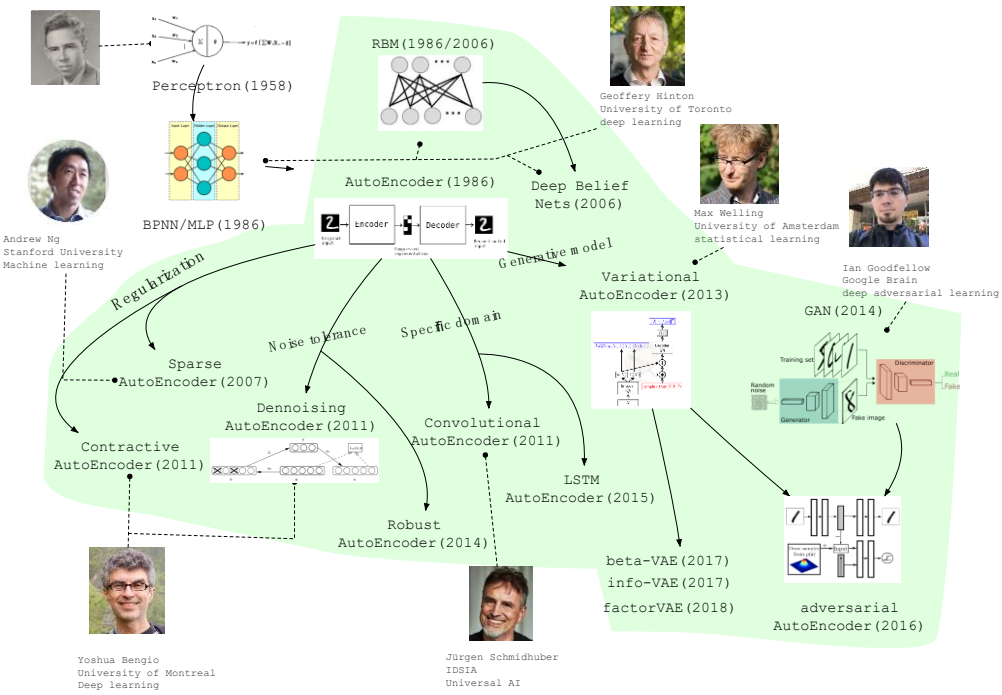


图 2-22 Auto-Encoder 的重要进展

稀疏 AutoEncoder 在学习输入样本表示的时候可以学习到相对比较稀疏的表示结果，这在 Overcomplete AutoEncoder（就是学习得到高维表示）方法中尤为重要。代表性人物包括斯坦福大学的 Andrew Ng 和蒙特利尔的 Yoshua Bengio 教授。具体方法就是在原来的损失函数中加一个控制稀疏化的正则化项，通过控制优化过程来实现。

Denoising AutoEncoder 的核心思想就是提高 Encoder 的鲁棒性，本质上就是避免可能的 overfitting。一个办法是在输入中加入随机噪音（比如随机置 0 一些输入，或者随机把部分输入变为 marked），这些思想后来在 BERT 等模型中也有广泛使用；另一个办法就是结合正则化的思想，比如在目标函数中加上 eEncoder 的 Jacobian 范数。Jacobian 范数可以让学习到的特征表示更具有差异性。

著名研究者 Jurgen Schmidhuber 提出了基于卷积网络的 AutoEncoder 以及后来的 LSTM AutoEncoder。Max Welling 基于变分思想提出变分 AutoEncoder 方法 VAE，这也是一个里程碑式的研究成果。后面很多研究者在这个工作上进行了扩展，包括 info-VAE、beta-VAE 和 factorVAE 等。最近还有人借鉴 Ian Goodfellow 等人提出的对抗建模思想提出 Adversarial AutoEncoder，也取得了很好的效果。这和之前的噪音容忍的 AE 学习也有一定呼应。除了上面的思想，就是可以把上面的各种方法 stacking 起来。

2.4.3 循环神经网络 RNN

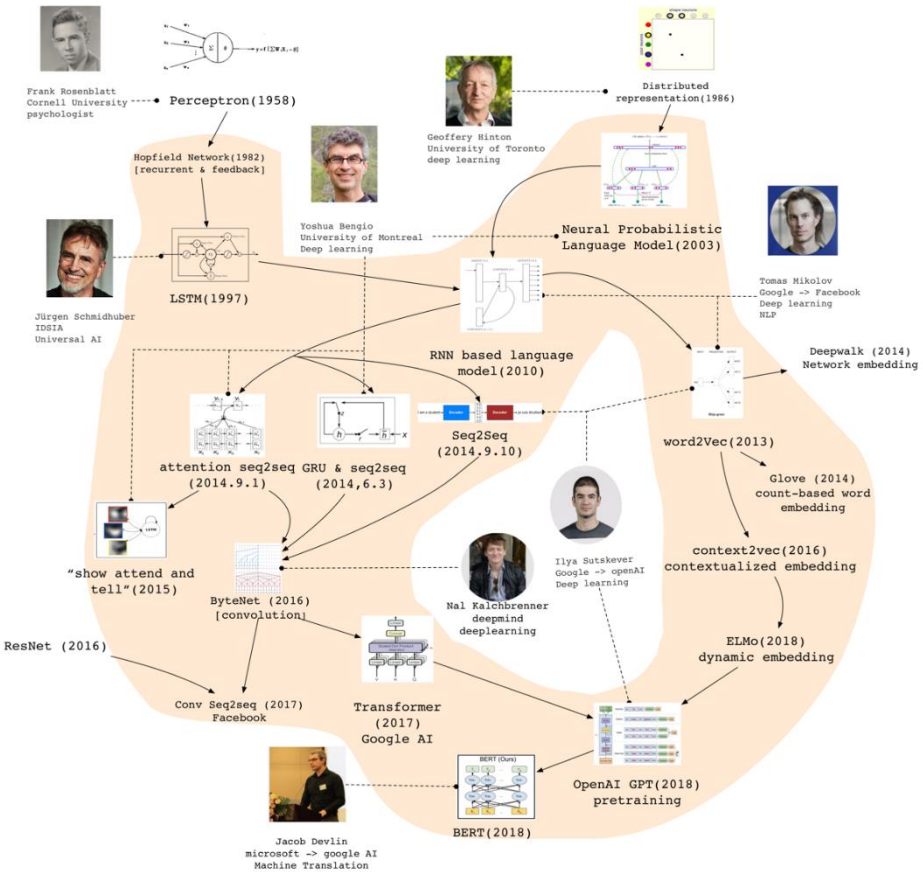


图 2-23 循环神经网络 RNN 的重要进展

1982 年，美国加州理工学院物理学家 John Hopfield 发明了一种单层反馈神经网络 Hopfield Network，用来解决组合优化问题。这是最早的 RNN 的雏形。86 年，另一位机器学习的泰斗 Michael I. Jordan 定义了 Recurrent 的概念，提出 Jordan Network。1990 年，美国认知科学家 Jeffrey L. Elman 对 Jordan Network 进行了简化,并采用 BP 算法进行训练，便有了如今最简单的包含单个自连接节点的 RNN

模型。但此时 RNN 由于梯度消失（Gradient Vanishing）及梯度爆炸（Gradient Exploding）的问题，训练非常困难，应用非常受限。直到 1997 年，瑞士人工智能研究所的主任 Jurgen Schmidhuber 提出长短期记忆（LSTM），LSTM 使用门控单元及记忆机制大大缓解了早期 RNN 训练的问题。同样在 1997 年，Mike Schuster 提出双向 RNN 模型（Bidirectional RNN）。这两种模型大大改进了早期 RNN 结构，拓宽了 RNN 的应用范围，为后续序列建模的发展奠定了基础。此时 RNN 虽然在一些序列建模任务上取得了不错的效果，但由于计算资源消耗大，后续几年一直没有太大的进展。

2010 年，Tomas Mikolov 对 Bengio 等人提出的 feedforward Neural network language model (NNLM)进行了改进，提出了基于 RNN 的语言模型（RNN LM），并将其用在语音识别任务中，大幅提升了识别精度。在此基础上 Tomas Mikolov 于 2013 年提出了大名鼎鼎的 word2vec，与 NNLM 及 RNNLM 不同，word2vec 的目标不再专注于建模语言模型，而是如何利用语言模型学习每个单词的语义化向量（distributed representation），当然 distributed representation 概念最早要来源于 Hinton 1986 年的工作。word2vec 引发了深度学习在自然语言处理领域的浪潮，除此之外还启发了 knowledge representation，network representation 等新的领域。

另一方面，2014 年 Bengio 团队与 Google 几乎同时提出了 seq2seq 架构，将 RNN 用于机器翻译。没过多久，Bengio 团队又提出注意力 Attention 机制，对 seq2seq 架构进行改进。自此机器翻译全面进入到神经机器翻译（NMT）的时代，NMT 不仅过程简单，而且效果要远超统计机器翻译的效果。目前主流的机器翻译系统几乎都采用了神经机器翻译的技术。除此之外，Attention 机制也被广泛用于基于深度学习的各种任务中。

近两年，相关领域仍有一些突破性进展，2017 年，Facebook 人工智能实验室提出基于卷积神经网络的 seq2seq 架构，将 RNN 替换为带有门控单元的 CNN，提升效果的同时大幅加快了模型训练速度。此后不久，Google 提出 Transformer 架构，使用 Self-Attention 代替原有的 RNN 及 CNN，更进一步降低了模型复杂度。在词表示学习方面，Allen 人工智能研究所 2018 年提出上下文相关的表示学习方法 ELMo，利用双向 LSTM 语言模型对不同语境下的单词，学习不同的向量表示，在 6 个 NLP 任务上取得了提升。OpenAI 团队在此基础上提出预训练模型

GPT，把 LSTM 替换为 Transformer 来训练语言模型，在应用到具体任务时，与之前学习词向量当作特征的方式不同，GPT 直接在预训练得到的语言模型最后一层接上 Softmax 作为任务输出层，然后再对模型进行微调，在多项任务上 GPT 取得了更好的效果。

不久之后，Google 提出 BERT 模型，将 GPT 中的单向语言模型拓展为双向语言模型（Masked Language Model），并在预训练中引入了 sentence prediction 任务。BERT 模型在 11 个任务中取得了最好的效果，是深度学习在 NLP 领域又一个里程碑式的工作。BERT 自从在 arXiv 上发表以来获得了研究界和工业界的极大关注，感觉像是打开了深度学习在 NLP 应用的潘多拉魔盒。随后涌现了一大批类似于“BERT”的预训练（pre-trained）模型，有引入 BERT 中双向上下文信息的广义自回归模型 XLNet，也有改进 BERT 训练方式和目标的 RoBERTa 和 SpanBERT，还有结合多任务以及知识蒸馏（Knowledge Distillation）强化 BERT 的 MT-DNN 等。这些种种，还被大家称为 BERTology。

2.4.4 网络表示学习与图神经网络（GNN）

这个方面的研究可以追溯到 Hinton 当年 1986 的 Distributed Representation，后来 Stanford 的 Andrew Ng 实验室做了个 Neural Tensor Network，本质就是把知识之间的关系和表示学习一起放到 tensor 里面来做，算是一个 smart 的扩展。后来 Facebook 的 Antonie Bordes 提出了 TransE 是一个 milestone 的工作，把知识网络的三元组融合到了表示学习中，这是 NLP 和知识图谱中的一个非常重要的研究，后面延续了一系列的工作，包括 TransH、TransR、TransA、TransG。

从表示学习本身来看，Neural Language Model 是对于单词和文本的表示，是对原来向量模型的一个自然扩展，其实本质上类似一个隐含语义分析，只是这里用的是神经网络来做学习。RNN based language model 是利用 RNN 进行表示学习，更好的保持了语言模型的连续性。但这个阶段的研究当时大部分都没有火起来，一是当时深度学习还没火起来，二是这些算法都还比较慢。2013 年 Tomas Mikolov 和 Jeff Dean 等人做 word2vec，可以说占据了“天时、地利、人和”：深度学习开始发热、算法简单有效、大神作品。现在 word2vec 已经轻松超过 1 万多引用了。后面的扩展也很多，如 pharagraph2vec、doc2vec，context2vec。以至

于后面有一段时间，“2vec”成了流行取名字的方法。最近的进展是 ELMo、OpenAI 的 GPT 和谷歌的 BERT。

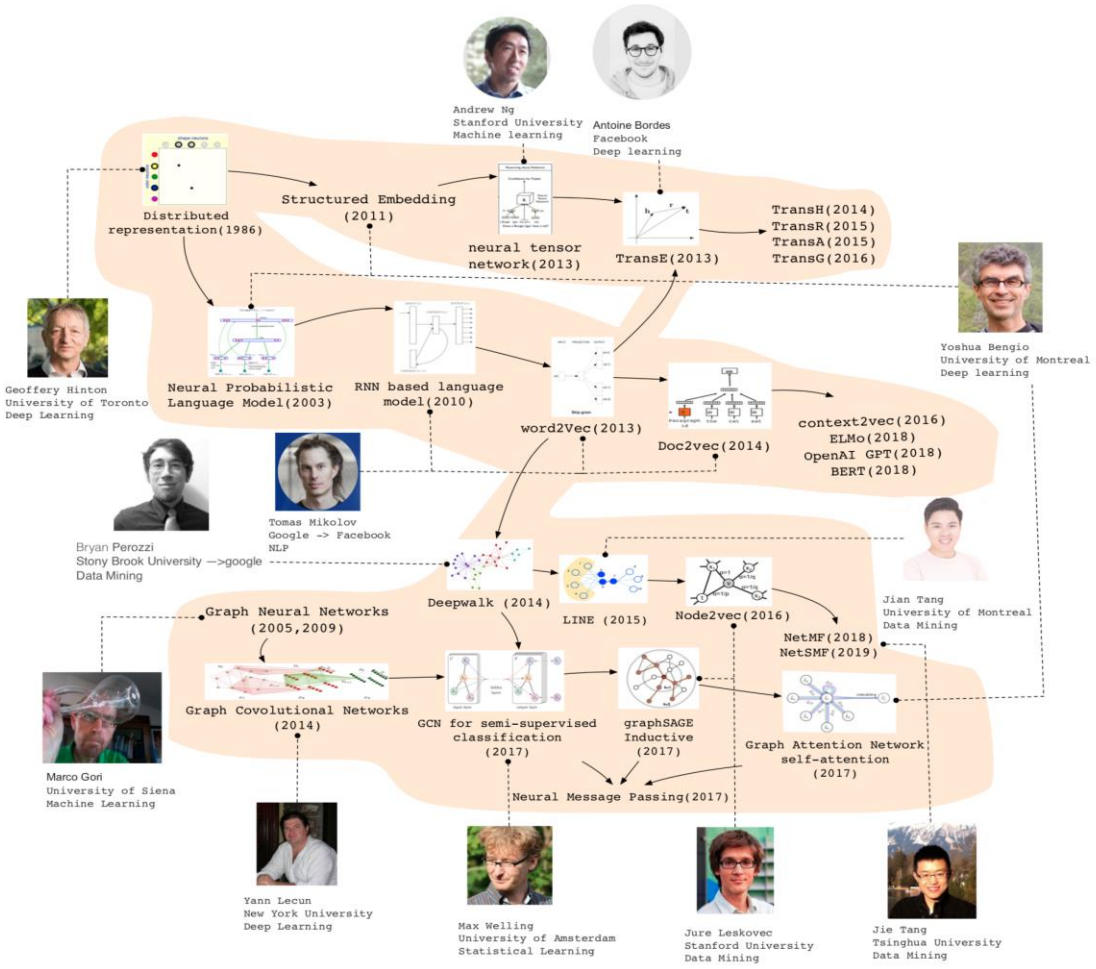


图 2-24 网络表示学习与图神经网络的重要进展

表示学习的另一个脉络就是扩展到网络数据上，在 NLP 领域的 Structured Embedding、TransE 等模型更多的是语言中的局部结构信息，而网络中还有更加复杂的拓扑结果。Bryan（原 Stony Brook 大学的，现在去了谷歌）提出 DeepWalk，这个算最早把 word2vec 稍微扩展了一下，应用于网络数据，这篇文章获得了当年 KDD 的最佳论文和后来 KDD 的最佳博士论文。很快这个工作吸引了大量关注，Jian Tang（原北大、微软，现在去了 Bengio 那边）等人做了两阶扩展，斯坦福的 Jure Leskovec 做了面向社交网络的“三阶”扩展 node2vec，后来清华也给出了一个理论证明，证明这些不同方法本质上都在做一个矩阵分解，并基于此提出了一个 NetMF 的算法以及其适用于大规模网络的实现 NetSMF。ProNE 是另

一个清华作品，其主要特点是高效和高精度。该方法非常简单，本质上是在原来的表示学习上引入了一个类似卷积但又不是卷积的操作，大大提高了精度。

最近的网络表示学习更多的是用卷积网络直接对图做，大方向是 Graph Neural Network，最早是 Siena 大学的 Marco 等人在 2005 和 2009 年提出的，但当时没引起太大关注。后来 Yann Lecun 提出的 Graph Convolutional Networks，还有 Kipf & Welling 等人提出的 semi-supervised 的 GCN。这一系列的研究本质上就是 Neural Message Passing，在最近引起大量关注。斯坦福的 Jure 也提出了 GraphSage，利用 NMP 简化了卷积，提高了速度，并且支持 inductive learning，再后来 Yoshua 他们团队又提出了 Graph Attention Network，进一步提高了图卷积精度。最近网络表示学习非常热，前前后后都能看到三大巨头 Hinton、Yoshua 和 Yann 的影子。在未来若干年还会继续是个研究热点。

2.4.5 增强学习

Deep Mind 的是一家英国人工智能公司，这是一个对增强学习影响最大的公司。创立于 2010 年，2014 年被 Google 收购。创始人哈萨比斯出身于伦敦，母亲为新加坡华裔，13 岁便已经获得国际象棋大师的头衔，19 岁开始学习围棋，当前是围棋业余初段。DeepMind 于 2014 年开始开发 AlphaGO。来看看 AlphaGO 的战绩吧。2015 年 10 月，AlphaGO 5:1 樊麾；2016 年 3 月，AlphaGo 4 : 1 李世石；2017 年 5 月，AlphaGO 3 : 0 柯洁；2017 年 10 月 19 日，AlphaGo Zero 发表在 Nature，其思路是从零开始，自我对弈，40 天超过所有版本。2018 年 12 月 7 日，AlphaZero 再次发表于 Science, AlphaZero 使用与 AlphaGo Zero 类似但更一般性的算法，在不做太多改变的前提下，并将算法从围棋延伸到将棋与国际象棋上。2018 年 12 月，Deep Mind 公司推出 AlphaFold, 可以根据基因序列预测蛋白质结构。2019 年 1 月 25 日，Deep Mind 公司 AlphaStar，在《星海争霸 II》以 10：1 战胜人类职业玩家。另一条在美国的战线，可能最著名的是 Open AI 公司，这是 Hinton 的高徒 Ilya Sutskever（AlexNet 发明人）创立的公司。2019 年 4 月，Open AI 推出 five dota2，2-0 战胜 Dota2 的 TI8 冠军战队 OG。

在研究方法上 Deep Q-Network（DQN）利用神经网络对 Q 值进行函数近似，并利用了 experience replay 和 fixed target network 的策略让 DQN 可以收敛，在

Atari 的不少游戏上都超过了人类水平。Double DQN 是深度学习版本的 double Q-learning，它通过微小的修改就成功减小了 DQN 中 max 操作带来的 bias。再后来，Dueling DQN 将 Q-network 分成了 action-dependent 和 action-independent 两个部分，从而提高了 DQN。DQN 是为 Value 的期望建模，greedy 的时候也是最大化期望的形式，Categorical DQN 的想法是直接为 Value 的分布进行建模。Noise DQN 在网络中添加了噪声，从而达到 exploration 的效果。DQN 还有非常多的提升版本，rainbow 整合了多种 DQN 版本。Ape-X 从 Rainbow 的工作中发现 Replay 的优先级对于性能影响是最大的，故扩大 Prioritised Replay Buffer，并使用 360 个 actor 做分布式的训练，比 rainbow 更快，也更好。

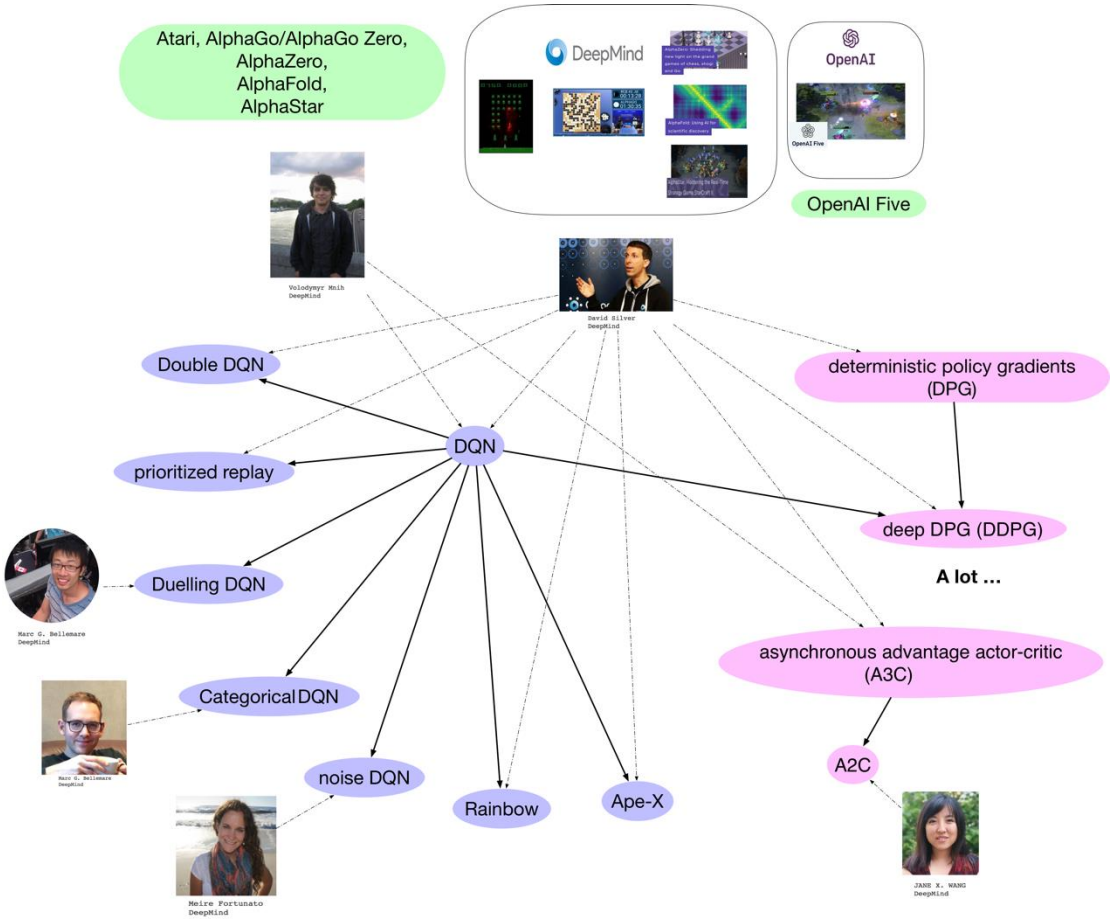


图 2-25 增强学习的重要进展

Deterministic policy gradients（DPG）将 policy gradients 方法中随机的 policy 推广为确定性 policy。Deep DPG 使用了神经网络表示高维 state，是结合了 DQN 和 DPG 的 actor critic 算法。A3C 是经典的 policy gradient 方法，可以并行 multiple

agent 的训练，并异步更新参数。A2C 是 A3C 的同步、确定性 policy 版本，同步的梯度更新，可以让并行训练更快收敛。

2.4.6 生成对抗网络

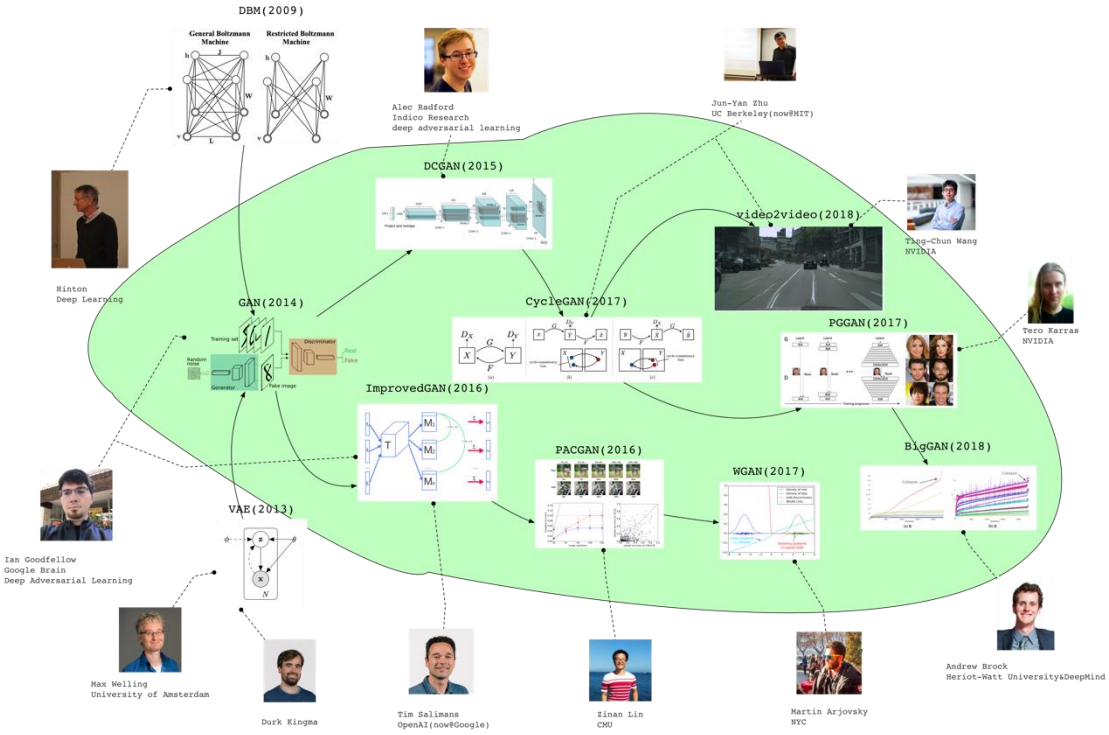


图 2-26 生成对抗网络的重要进展

GAN 最近几年发展非常快，这也是 Yoshua Bengio 获得图灵奖的贡献之一。传统的生成模型是要预测联合概率分布 $P(x, y)$。首先玻尔兹曼机（RBM）这个模型其实是一个基于能量的模型，1986 年的时候就有，Hinton 在 2006 年的时候重新拿出来作为一个生成模型，并且将其堆叠成为 Deep Belief Network，使用逐层贪婪或者 wake-sleep 的方法训练。

AutoEncoder 也是上个世纪 80 年代 Hinton 就提出的模型，此时由于计算能力的进步也重新登上舞台。Bengio 等人又搞了 Denoise AutoEncoder。Max welling 等人使用神经网络训练一个有一层隐变量的图模型，由于使用了变分推断，并且最后长得跟 AutoEncoder 有点像，被称为 Variational AutoEncoder。此模型中可以通过隐变量的分布采样，经过后面的 decoder 网络直接生成样本。

在生成模型方面，最近一个最重要的进展就是对抗生成网络（GAN），可以说是现在最火的生成模型。2014 年 Ian Goodfellow 在 NIPS 上发表了最初的 GAN

文章，到现在已经有近九千引用。为什么这个模型引起如此大的关注呢？一个原因是这个模型理论上非常优雅，大家理解起来简单方便；二就是效果确实好。看上面这一排，是基于 GAN 的一些应用文章，下面这些是改进 GAN 的训练的一些文章。这些文章都引起了广泛关注。可以看出，GAN 的发明人 Ian Goodfellow 是少年得志的典范。他本科在斯坦福，硕士在 Andrew Ng 手下，博士就跑到蒙特利尔 Yoshua Bengio 手下了。他另外还有一个导师 Aaron Courville。大家现在经常用的教科书《Deep Learning》，作者就是 Ian Goodfellow 和他两个博士导师。他是 85 年人，发表 GAN 在 2014 年，29，还差一年才 30。GAN 这个工作也给 Goodfellow 带来了很多荣誉，比如 17 年就被 MIT under 35 选中了。Goodfellow 博士毕业后去了 Google Brain，后来又跳到 Open AI，又跳回 Google，现在在苹果做特别项目机器学习项目负责人，实际上现在他也才 34 岁。另外，GAN 是 Ian Goodfellow 在蒙特利尔的时候的工作。大家知道今年图灵奖给了深度学习三巨头，其中的 Bengio，在图灵奖官网上给获奖理由，选的三个贡献之一就是 GAN。另外两个贡献分别是 90 年代的序列概率模型和 00 年代的语言模型。GAN 可以说是 Bengio 的代表作之一了，甚至可以说帮助他拿图灵奖。

另外还有几个有名的 GAN 的扩展，包括：cycleGAN 和 vid2vid。去年 NIPS 企业展示会场，英伟达把 vid2vid 配合方向盘，做了个实物 demo，非常引人关注。

2.4.7 老虎机

老虎机也是机器学习的一个重要分支，和深度学习有着或多或少的联系。老虎机实际上是个赌博机器。走进拉斯维加斯赌场，你就能看到一排排闪亮的机器。老虎机模型这个数学模型，现在追本溯源基本认为是一个病理学家 Thompson 在 1933 年提出的。他当时觉得验证新药的医学的随机双盲实验有些残酷的地方，对于被分到药效较差的新药的那一组病人并不公平。他想知道能否在实验中途就验证药物药效，从而避免给病人带来痛苦，因此他提出了一个序列决策模型。但是，实际使用还是有很多问题，比如中途效果不好评价。所以直到现在，美国 FDA 对在医学随机双盲实验中使用这种自适应调整的多臂老虎机方法，仍然只是建议使用。就现在而言，老虎机模型实际是在搜索和推荐方面的应用很多。

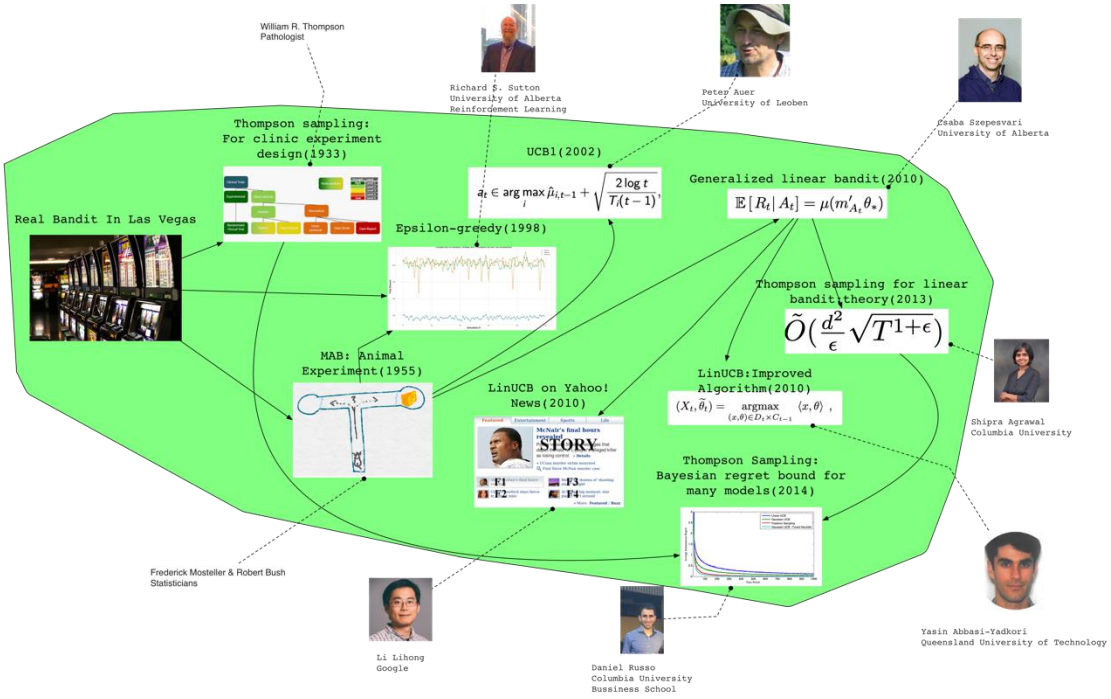


图 2-27 老虎机的重要进展

Epsilon-greedy 是种预留一点点机会去尝试的思想。这种想法很自然，学术界也不清楚最初的 credit 该给谁。现在就放在 sutton 名下。他是强化学习方面的大佬，写的那本教材 *Reinforcement Learning* 引用五位数，里边讲解了这个算法。Peter Auer 这个工作不仅分析了 UCB 算法的理论性质，还顺道分析了 Epsilon-greedy 的理论性质。这篇文章用到的技术，是此后很多更复杂技术的基础，很值得一看。这篇纯理论文章的引用量也达到了两千多。Frederick Mosteller 是哈佛统计系奠基人，20 世纪统计学界的超级牛人。他们当时做老虎机模型，主要是想给真实的动物或者人的序列决策建模，想抽象一个框架出来。所以他们作了一个老鼠找蛋糕的实验。当然，也做了关于人玩赌博用的老虎机的实验。Li Lihong 是清华 02 级校友。他 在 Yahoo! news 上的 LinUCB 的工作发表在 WWW 上，这篇应用文章获得了大量关注，引用上千。他后来又翻出来 Thompson sampling 这个很古旧的方法，作了一些系统性的实验，从实验结果的角度说明 Thompson sampling 效果很好。这篇文章发在 NIPS2011 上，也获得了大量关注。后来大批做理论的人就跟进，就把 Thompson sampling 在线性模型上的理论基础建立起来了。比如 Russo 这篇文章。那可以看到，从 Thompson 1933 年用 Thompson sampling，到 2010 年后这个方法的理论基础才建立起来，这个时间跨度是很大的。当然，因为线性情况下都还比较简单，所以 2011 年后收到广泛关注没几年，理论就建

立。这个现象和神经网络的理论建立基本是一个样子，都是线性的容易又基础，就先做着。研究老虎机模型确实比较偏理论，但老虎机应用也很广。上图里边除了有做医学的、做统计的、做计算机科学的，还有在商学院任教的，就是这个 Russo。

2.5 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为机器学习领域全球学者分布情况：



图 2-28 机器学习领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量遥遥领先且主要分布在其东西海岸；欧洲中西部也有较多的人才分布；亚洲的人才主要分布于我国东部及日韩地区；其他诸如非洲、南美洲等地区的学者非常稀少；机器学习领域的人才分布与各地区的科技、经济实力情况大体一致。此外，在性别比例方面，机器学习领域中男性学者占比 89.8%，女性学者占比 10.2%，男性学者占比远高于女性学者。

- h-index 分布

机器学习学者的 h-index 分布如下图所示，大部分学者的 h-index 都在 20 以上，其中 h-index 在 20-30 区间的人数最多，有 584 人，占比 28.8%，小于 20 区间的人数最少，共 7 人。

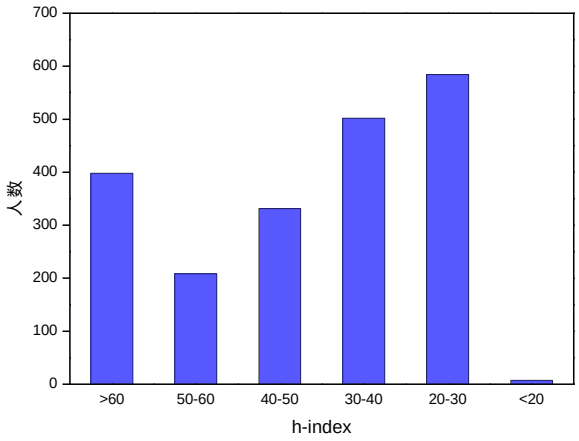


图 2-29 机器学习领域学者 h-index 分布

- 中国人才分布



图 2-30 机器学习领域中国学者分布

我国专家学者在机器学习领域的分布如上图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等亚洲国家相比，中国在机器学习领域学者数量较多。

中国与其他国家在机器学习的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 2-1 机器学习领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	总的学者数
中国-美国	511	26694	52	819
中国-英国	44	1398	32	73
中国-新加坡	36	1189	33	56
中国-澳大利亚	31	744	24	42
中国-印度	22	1123	51	19
中国-德国	17	419	25	39
中国-瑞士	11	233	21	22
中国-荷兰	6	93	16	10
中国-巴基斯坦	4	82	21	3
中国-以色列	3	23	8	6

从上表数据可以看出，中美合作的论文数、引用数、平均引用数、学者数遥遥领先，表明中美间在机器学习领域合作之密切；从地域角度看，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席；中国与印度合作的论文数虽然不是最多，但是拥有平均引用数依然位列第二，说明在合作质量上中印合作也达到了较高的水平。

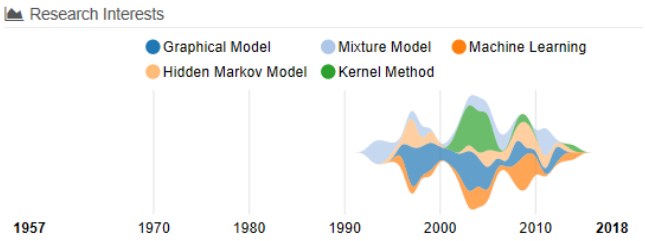
2.6 代表性学者简介

综合 h-index 以及领域知名度与活跃度，下面我们将对国内外机器学习领域代表性学者进行简要介绍，排名不分先后。此外，限于报告篇幅，我们对所有学者不能逐一罗列，如有疏漏，还请与 AMiner 编者联系，或者登录 https://www.aminer.cn/获取更多资料。

2.6.1 国际顶级学者

- Michael I. Jordan





Michael I. Jordan，美国三院（美国国家科学院、美国国家工程院、美国艺术与科学院）院士，机器学习泰斗，被誉为人工智能领域的“根目录”之一，伯克利大学机器学习实验室 AMP Lab 联合主任，IEEE Fellow，ACM Fellow。

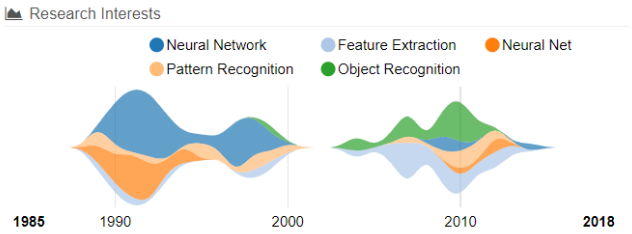
Michael I. Jordan 是美国科学家，加州大学伯克利分校电子工程系、计算机科学和统计系杰出教授，机器学习、统计学和人工智能研究员。他是机器学习领域的领军人物之一，并且在 2016 年被 *Semantic Scholar*（科学报）称为世界上最有影响力的计算机科学家。同年也被 AMiner 评为机器学习最有影响力学者。

他于 1985 年获得加利福尼亚大学圣地亚哥分校博士学位。自 1988 年至 1998 年，Michael I. Jordan 任麻省理工学院教授，他的研究方向包括了计算学、统计学、认知科学以及生物科学。近年来，他的研究兴趣集中在贝叶斯非参数分析、概率图模型、谱方法、核方法、分布式计算系统、自然语言处理、信号处理和统计遗传学等问题的应用上。深度学习领域的权威 Yoshua Bengio，贝叶斯学习领域权威 Zoubin Ghahramani 及前百度首席科学家吴恩达等人都是其门下学生。

他曾获得众多奖项，在 2016 年获得 IJCAI 研究卓越奖（IJCAI Research Excellence Award），2015 年获得了 David E. Rumelhart 奖，并在 2009 年获得了 ACM / AAAI Allen Newell 奖，2004 年获得 ICML 最佳学生论文奖。同时，他是 AAAI、ACM、ASA、CSS、IEEE、IMS、ISBA 和 SIAM 成员。

- Yann LeCun





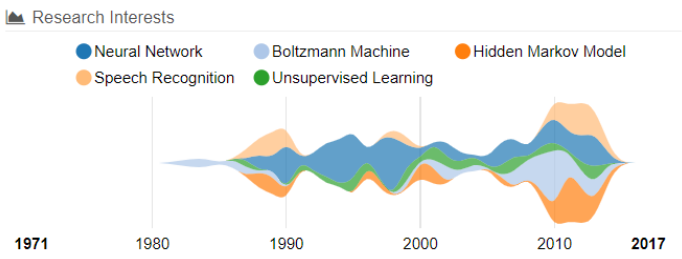
Yann LeCun，人工智能领域三大奠基人之一，被称为“卷积网络之父”。

Yann LeCun 是美国工程院院士，Facebook 人工智能研究院院长，纽约大学 Sliver 教授，同时还兼职于科学数据中心，数学科学交流学院，神经科学中心，以及电子工程计算机系。他以使用卷积神经网络（CNN）进行光学字符识别和计算机视觉方面的工作而闻名，并且是卷积网络的创始人。

他获得巴黎第六大学（Pierre et Marie Curie）大学的计算机科学博士学位，1987 年至 1988 年，是多伦多大学 Geoffrey Hinton 实验室的博士后研究员。他于 2003 年加入纽约大学，之后还在普林斯顿的 NEC 研究院短暂任职。在 2012 年，他创建了纽约大学数据科学中心，并担任主任。2013 年底，他被任命为 Facebook 人工智能研究总监，并继续在纽约大学做兼职教授。2015-2016 年，他在巴黎法兰西工学院做客座教授。

- Geoffrey Hinton





Geoffrey Hinton，人工智能领域三大奠基人之一，被称为“神经网络之父”，“深度学习鼻祖”。

Geoffrey Hinton 是英国计算机科学家，担任多伦多大学计算机科学系教授，多伦多大学向量学院（Vector Institute）首席科学顾问。人工智能三大奠基人之一 Yann LeCun 以及谷歌大脑研究科学家 Hugo LaRochelle 都是其博士后。

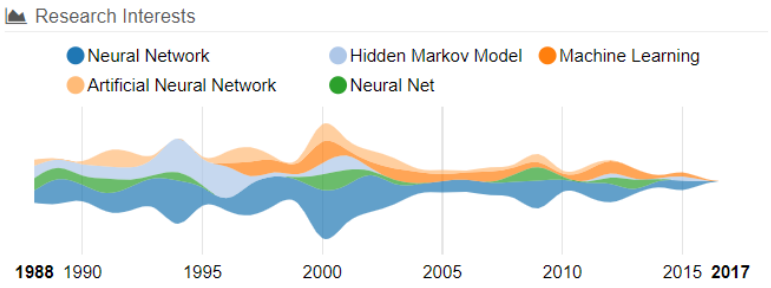
他于 1975 年获得爱丁堡大学人工智能方向博士学位，神经网络是他的研究重点。

2013 年，他加入谷歌并带领 AI 团队，将神经网络带入到研究与应用的热潮，将“深度学习”从边缘课题变成了谷歌等互联网公司的依赖的核心技术，并将 Backpropagation（反向传播）算法应用到神经网络与深度学习。

Geoffrey Hinton 获得诸多奖项。2016 年获得 NEC C&C Award，IEEE/RSE James Clerk Maxwell Medal；2014 年获得 IEEE Frank Rosenblatt Medal；2013 年获得 Doctorat honorifique, University of Sherbrooke；2012 年，获得了加拿大基廉奖（Killam Prizes，有“加拿大诺贝尔奖”之称的国家最高科学奖）。2011 年获得苏赛克斯大学理学博士荣誉学位；2005 年获得 JICAI 卓越研究奖项。

- Yoshua Bengio





Yoshua Bengio，加拿大计算机科学家，与 Geoffrey Hinton、Yann LeCun 一起，被称为人工智能三大奠基人。根据 MILA 的数据，在 *h* 指数至少为 100 的计算机科学家中，Yoshua Bengio 是每天都有被引用的一个。他在人工神经网络和深度学习方面做出了突出贡献。

Yoshua Bengio 于 1991 年获得加拿大麦吉尔大学计算机科学博学位，并是麻省理工学院和贝尔实验室的博士后。他自 1993 年以来担任蒙特利尔大学教授，担任计算机科学与运筹学系主任。他撰写了三本书，超过 500 种出版物（h-index 为 125，超过 135000 次引用），最常被引用在深度学习，复现神经网络，概率学习算法，自然语言处理和多元学习领域，其中，*Deep Learning* 是他于 GAN 之父 Ian Goodfellow 等人合著的入门深度学习必读经典教程。

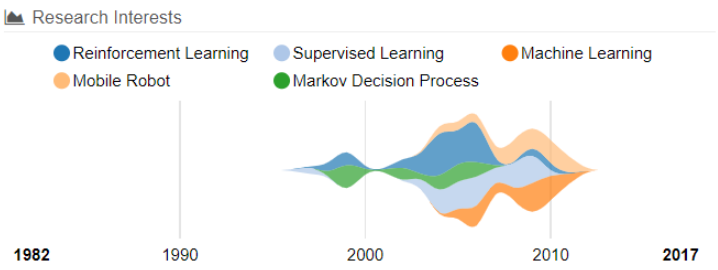
他是加拿大最受欢迎的计算机科学家之一，也是（或曾经是）机器学习和神经网络中顶尖期刊的副主编。自 2000 年起，他在统计学习算法中担任加拿大研究主席，自 2006 年成为 NSERC 工业主席，自 2005 年以来，他是加拿大高级研究所高级研究员，自 2014 年以来，他一直致力于深入学习。他是 NEURIPS 基金会的董事会成员，也是 NEURIPS 的课程主席和总裁。他共同组织了 14 年的学习研讨会，还共同组织了新的国际学习代表会议。他目前的兴趣集中于通过机器学习对 AI 的追求，并且包括关于深度学习和表征学习的基本问题，高维空间中的泛化几何，多元学习，生物学启发式学习算法以及统计机器学习的具有挑战性的应用。

2016 年 10 月，Yoshua Bengio 联合创立了 Element AI，这是一家位于蒙特利尔的企业孵化器，致力于将人工智能（AI）研究转化为实际的商业应用。2017 年 5 月，Bengio 宣布他将加入蒙特利尔的法律创业公司 Botler AI，担任战略顾问。他是 2017 年玛丽维克多魁北克奖获得者，加拿大皇家学会会员，他是 CIFAR 高级研究员并共同指导其在机器和大脑学习计划。此外，他还是 MILA（蒙特利尔大学学习算法学院）的创始人们兼科学主任。

Yoshua Bengio 的论文“*A neural probabilistic language model*”开创了神经网络 language model（语言模型）的先河。该论文的思路影响、启发了之后的很多基于神经网络做 NLP（自然语言处理）的文章。

- Andrew Y. Ng（吴恩达）





Andrew Y. Ng 最知名的事情是他所开发的人工神经网络通过观看一周 YouTube 视频，自主学会识别哪些是关于猫的视频。这个案例为人工智能领域翻开崭新一页。

Andrew Y. Ng 于 2002 年获得了加州大学伯克利分校的博士学位，并从这年开始在斯坦福大学工作。他是前文介绍的 Michael I Jordan 的弟子。他的主要兴趣领域在机器学习、深度学习、机器人、人工智能和计算机视觉等方面。2010 年，时任斯坦福大学教授的 Andrew Y. Ng 加入谷歌开发团队 XLab——这个团队已先后为谷歌开发无人驾驶汽车和谷歌眼镜两个知名项目，Andrew Y. Ng 加入后开始“谷歌大脑”项目。2014 年 5 月，吴恩达加入百度，担任百度公司首席科学家，负责百度研究院的领导工作，尤其是 Baidu Brain 计划。2017 年 10 月，吴恩达出任 Woebot 公司新任董事长，该公司拥有一款同名聊天机器人

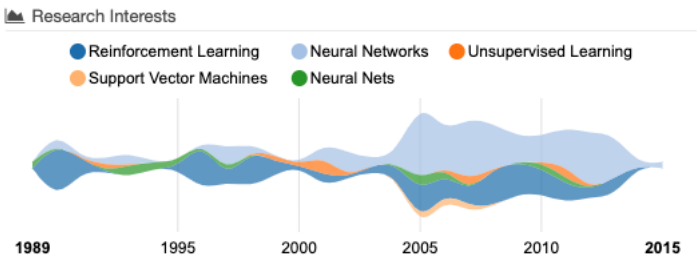
Andrew Y. Ng 最知名的事情是，他所开发的人工神经网络通过观看一周 YouTube 视频，自主学会识别哪些是关于猫的视频。这个案例为人工智能领域翻开崭新一页。

他 2007 年获得了斯隆奖（Sloan Fellowship），2008 年入选“the MIT Technology Review TR35”，即《麻省理工科技创业》杂志评选出的科技创新 35 俊杰，以及计算机思维奖（Computers and Thought Award），并在 2013 年入选《Time》杂志年度全球最有影响力的 100 人之一，其中共 16 位科技界人物。他也是“计算机和思想奖”的获得者。

他现在的兴趣主要是深度学习。他在 2013 年前共有 128 项学术著作，如 *Deep Learning with COTS HPC Systems*（Adam Coates,Brody Huval,Tao Wang,David J.Wu,Bryan Catanzaro and Andrew Y.Ng 等人在 ICML 2013 上发表）、*Parsing with Compositional Vector Grammars* 等，限于篇幅，本报告不一一列举。他所著 *Machine Learning Yearning* 于 2018 年出版，该书面向的用户群体为机器学习从业者，主要介绍机器学习实际使用时的一些策略和技巧，以便为开发指明方向，提升开发效率。

- Jürgen Schmidhuber



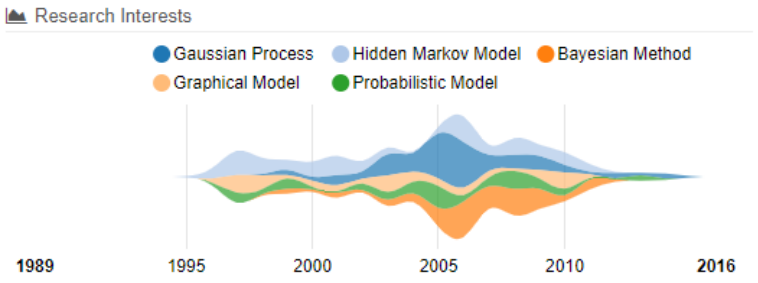


Jürgen Schmidhuber 出生于德国，是瑞士人工智能实验室（IDSIA）的研发主任，是 LSTM 的 发明人、深度学习元老，被称为递归神经网络之父。Schmidhuber 本人创立的公司 Nnaisense 正专注于人工智能技术研发。此前，他开发的算法让人类能够与计算机对话，还能让智能手机将普通话翻译成英语。

德国计算机科学家 Jürgen Schmidhuber 在接受英国《卫报》采访时表示，宇宙史上重大事件的发生间隔似乎在几何式地缩短——前后两个里程碑事件的间隔约为前一个间隔的四分之一。按照这一规律，人工智能可能在 2050 年超过人类智商。人工智能将造就一种新型的生命，像是生物大爆炸。

- Zoubin Ghahramani





Zoubin Ghahramani，是剑桥大学信息工程教授，他领导了由大约 30 名研究人员组成的机器学习小组，并领导了 Uber-AI 实验室的首席科学家。他曾担任英国国家数据科学研究所阿兰图灵研究所（Alan Turing Institute）的创始剑桥主任，勒沃胡姆未来情报中心（Leverhulme Centre for the Future of Intelligence）副学术主任，剑桥圣约翰学院（St John's College Cambridge）院士。

他在宾夕法尼亚大学学习计算机科学和认知科学，1995 年从麻省理工学院获得博士学位，并在多伦多大学做博士后。他的学术生涯包括同时被任命为伦敦

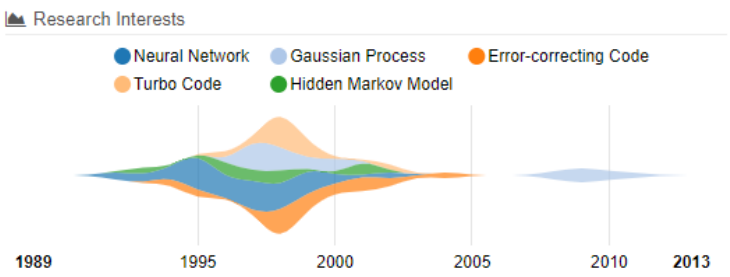
盖茨比计算神经科学部门的创始成员之一，以及 CMU 机器学习部门的教员超过 10 年。

他目前的研究兴趣包括统计机器学习、贝叶斯非参数、可伸缩推理、概率规划等。他发表了 250 多篇论文，获得 38000 多条引文（h 指数 84）。他的工作得到了 EPSRC、DARPA、微软、谷歌、Infosys、Facebook、亚马逊、FX Concepts、NTT 和其他一些工业合作伙伴的资助和捐赠。

2013 年，他获得了 75 万美元的谷歌奖，用于研究如何建立自动统计师。他曾担任微软剑桥研究院（Microsoft Research Cambridge）、VocalIQ（被苹果收购）、剑桥资本管理公司（Cambridge Capital Management）、Echobox、Informetis、Opera Solutions 和其他几家公司的顾问。他还担任过一些领导职务，担任机器学习领域主要国际会议的项目和总主席：AISTATS（2005 年）、ICML（2007 年、2011 年）和 NIPS（2013 年、2014 年）。2015 年，他被选为皇家学会会员，2016 年，他被评为机器学习领域十大最具影响力的学者之一。

- David J.C. MacKay





David J.C. MacKay，曾任剑桥大学卡文迪什实验室物理系自然哲学教授，现为剑桥大学工程系教授，能源和气候变化部首席科学顾问。

1967 年 4 月 22 日出生于英国特伦特的斯托克。在纽卡斯尔接受莱姆学校和剑桥三一学院的教育后，他于 1991 年在加州理工学院完成了计算和神经系统博士学位。

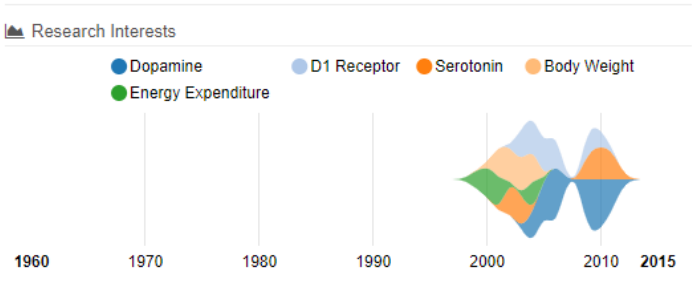
他的兴趣包括构建和实现发现数据模式的分层贝叶斯模型，开发神经网络的概率方法，以及纠错码的设计和解码。

他在机器学习、信息理论和通信系统方面的研究在国际上享有盛名，其中包括 Dasher 的发明，Dasher 是一种软件接口，可以用任何肌肉在任何语言中进行有效的通信。他从 1995 年开始在剑桥教物理。自 2005 年以来，他将越来越多的时间用于能源方面的公共教学。他是世界经济论坛全球气候变化议程理事会成员。

1985 年南斯拉夫国际物理奥林匹克运动会：银牌；一等奖，1999 年通信学会 Leonard G.Abraham 奖论文奖（与 R.J.McEliece 一起以及 J.－F.Cheng），2001 年、1999 年 IBM 合作伙伴奖，2009 年当选物理研究所院士、皇家学会会员，2010 年当选土木工程师学会会员，2013 年获梅尔切特奖。

- Christopher Bishop





Christopher Bishop，微软剑桥研究院技术研究员兼实验室主任，爱丁堡大学计算机科学教授，剑桥达尔文学院院士。

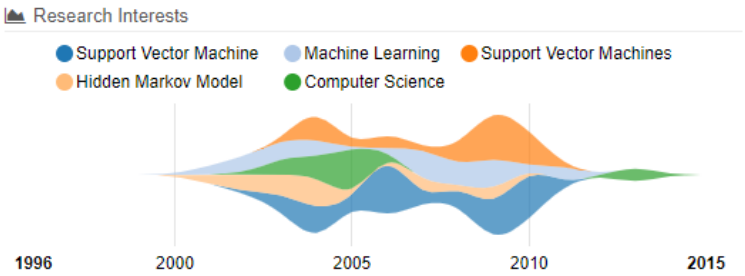
Chris 在牛津大学获得物理学学士学位，在爱丁堡大学获得理论物理学博士学位，并发表了一篇关于量子场论的论文。从那时起，他对模式识别产生了兴趣，并成为 AEA 技术应用神经计算中心的负责人。随后，他被选为阿斯顿大学计算机科学和应用数学系的主席，并在那里成立和领导了神经计算研究小组。

克里斯是两本被广泛引用的机器学习教科书的作者：《神经网络模式识别》（1995）和《模式识别与机器学习》（2006）。他还致力于机器学习在从计算机视觉到医疗保健等领域的广泛应用。克里斯是公众参与科学的积极倡导者，2008 年，他发表了著名的皇家学会圣诞讲座，1825 年由迈克尔法拉第创立，并在国家电视台播出。

他于 2004 年当选皇家工程院院士，2007 年当选爱丁堡皇家学会院士，2017 年当选皇家学会院士。

- Tony Jebara





哥伦比亚大学计算机科学系副教授，哥伦比亚大学机器学习实验室负责人。

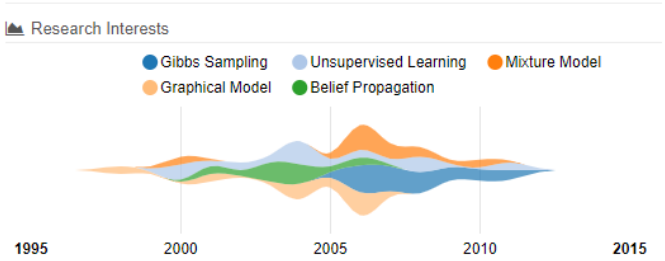
主要研究方向为计算机科学和统计学的交叉融合，在视觉、学习和时空建模等方面成就很高。

他于 2002 获得麻省理工学院博士学位。他指导哥伦比亚机器学习实验室（Columbia Machine Learning Laboratory），该实验室的研究与计算机科学和统计学交叉，开发新的数据学习框架，并将其应用于视觉、网络、时空数据和文本。Tony Jebara 已经创立了包括 Sense Networks、Agolo、Ninoh 和 Bookt 在内的几家初创公司，并为其提供咨询服务。他在会议、研讨会和期刊上发表了 100 多篇同行评议论文，包括 NIPS、ICML、UAI、COLT、JMLR、CVPR、ICCV 和 AISTAT。他是《机器学习：辨别与生成》一书的作者，也是视觉、学习和时空建模领域多项专利的共同发明人。

2004 年，Tony Jebara 获得了国家科学基金会的职业奖。他的作品在第 26 届机器学习国际会议上获得最佳论文奖，在第 20 届机器学习国际会议上获得最佳学生论文奖，并在 2001 年获得模式识别学会的杰出贡献奖。Jebara 的研究已经在电视上（ABC，BBC，New York One，TechTV 等）和大众媒体（纽约时报，斜线点，有线电视，商业周刊，IEEE 频谱等）上出现。《绅士》杂志将他评为 2008 年最优秀、最聪明的人物之一。Jebara 还是《机器学习研究》杂志和《机器学习》编辑委员会的副主编。2007 年至 2011 年，Jebara 任机器学习副主编，2010 年至 2012 年任 IEEE 模式分析和机器智能事务副主编。2006 年，他与人共同创立了 NYAS 机器学习研讨会，并从那时起一直担任该研讨会的指导委员会成员。Tony Jebara 还担任了 2014 年第 31 届机器学习国际会议（ICML）的项目主席。

- Max Welling





Max Welling，阿姆斯特丹大学的“研究主席”，加州大学欧文分校（UCI）计算机科学与统计学教授，加拿大高级研究所（CIFAR）副研究员，Scyfer BV 联合创始人。

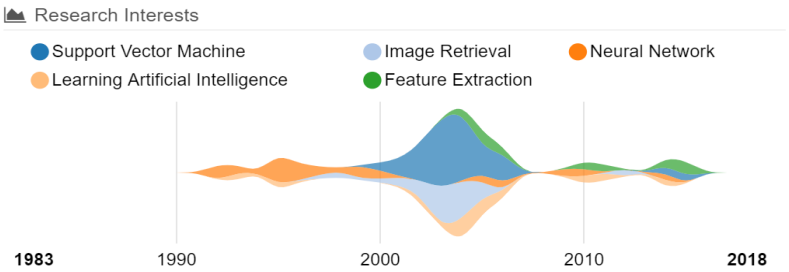
在过去，他曾在加州理工学院（Caltech）（1998-2000）、加州大学洛杉矶分校（UCL）（2000-2001）和多伦多大学（U.Toronto）（2001-2003）担任博士后。1998 年，他在诺贝尔奖获得者霍夫特教授的指导下获得了博士学位。

Max Welling 从 2011 年至 2015 年担任 IEEE TPAMI 的副主编。他自 2015 以来担任 NIPS 基金会的董事会成员（在机器学习方面规模最大的会议），分别担任 2013 和 2014 年度 NIPS 的计划主席和总主席。2009 年，他还是 AISTATS 和 2016 年 ECCV 的项目主席，2018 年 MIDL 的总主席。他曾在 JMLR 和 JML 的编辑委员会任职，并担任神经计算、JCGS 和 TPAMI 的副主编。他从谷歌、Facebook、雅虎、NSF、NIH、NWO 和 ONR-MURI 获得了多项资助，其中一项是 2005 年的 NSF 职业资助。他是 2010 年 ECCV Koenderink 奖的获得者。Welling 是阿姆斯特丹数据科学研究中心的董事会成员，他领导阿姆斯特丹机器学习实验室（AMLAB），并共同领导高通公司的 UvA 深度学习实验室（QUVA）和博世公司的 UvA 深度学习实验室（DELTA）。

2.6.2 国内知名学者

- 张钹





张钹，中国科学院院士，清华大学计算机科学与技术系教授，清华大学人工智能研究院院长。

张钹于 1958 年毕业于清华大学自动控制系，是国家第一批自动控制专业的毕业生。1995 年他当选为中国科学院院士。

他早期从事自动控制理论与系统研究，1979 年开始计算机科学与技术研究。从事人工智能理论、人工神经网络、遗传算法、分形和小波等理论研究；以及把上述理论应用于模式识别、知识工程、智能机器人与智能控制等领域的应用技术研究。

他针对人工智能问题求解计算复杂性、指数爆炸的主要困难，提出了问题分层求解的商空间理论，解决了不同粒度空间的描述、它们之间相互转换、复杂性分析等理论问题。在此基础上提出统计启发式搜索算法，基于拓扑的空间规划方法和关系矩阵的规划算法，对克服计算量的指数爆炸很有成效。还提出了研究不确定性处理、定性推理、模糊分析、证据合成等新原理。指导并参加建成了陆地自主车、图像与视频检索等实验平台。

张钹和同期同事成了国内最早接触到人工智能的研究者，并成为我国在这方面的首批专家。

在学术研究上的主要贡献是提出问题分层求解的商空间理论，通过代数的方法，系统地解决了不同层次求解空间的问题表达、复杂性分析、不同层次空间之间信息、算子及推理机制等的相互转换关系。在上述理论基础上，他进一步提出了统计启发式搜索算法，基于拓扑的空间规划方法以及基于关系矩阵的时间规划

算法等，极大降低了计算复杂性，具有重要的应用价值。其专著《问题求解理论及应用》全面总结了他在人工智能理论研究中的成果，其英文版于 1992 年由 Elsevier Science Publishers B.V.(Nortn-Holland)出版，中文版获国家教委颁发的高校出版社优秀学术专著特等奖。澳大利亚专家 Ronald Walts 在计算机杂志 *The Australian Computer Journal*（1995）对《问题求解理论及应用》（英文版）的评论为“这是一部在重要研究领域的优秀著作”。美国学者 Harold S.Stone 认为，张钹等在统计启发式搜索等方面的工作，是“最近几年中国学者作出的很有意义的贡献”，“将新一代计算技术的前沿向前推进了”。

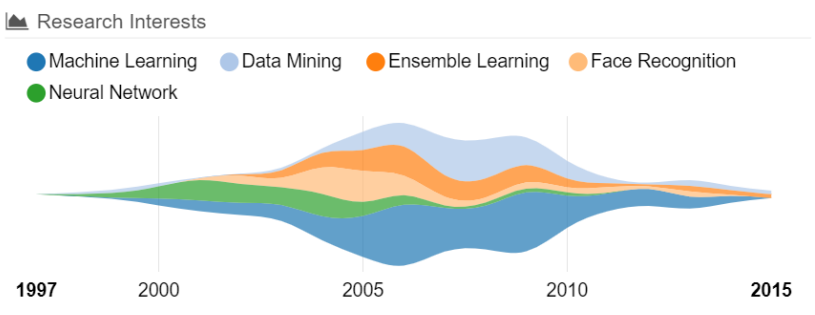
他在国内外共发表论文 100 多篇，中英文专著有《问题求解理论及应用》（中英版）以及《人工神经网络理论及应用》等。

他于 1994 年当选为俄罗斯自然科学院外籍院士；1995 年当选为中国科学院院士；2011 年德国汉堡大学授予自然科学名誉博士；2015 年 1 月 31 日，张钹获得 2014 CCF 终身成就奖。

他的社会任职有：智能技术与系统国家重点实验室主任、校学位委员会副主任、信息科学与技术学院学术委员会主任；中国自动化学会机器人专业委员会副主任及智能控制专业委员会副主任；《计算机学报》副主编；国家高技术“863”计划智能机器人主题专家组成员；河南科技大学兼职院士；计算机学学术委员会主任。

- 周志华





周志华，南京大学教授，博士生导师；教育部长江学者特聘教授，国家杰出青年基金获得者；南京大学计算机科学与技术系副主任、软件新技术国家重点实

验室常务副主任，机器学习与数据挖掘研究所（LAMDA）所长，校学术委员会委员、南京大学人工智能学院院长（兼）。

周志华于 2000 年获得南京大学计算机科学与技术系博士学位，2001 年 1 月起留校任教，2002 年 3 月被破格聘任为副教授，2003 年，在他 29 岁时获得国家杰出青年科学基金，随后被聘为教授。

他于 2006 年入选教育部长江学者特聘教授，2012 年当选 IEEE Fellow 和 IAPR Fellow（国际模式识别学会会士），2013 年当选 ACM Distinguished Scientist（ACM 杰出科学家）和中国计算机学会（CCF）会士，成为大陆高校首位当选 ACM 杰出科学家的学者。2007 年创建南京大学机器学习与数据挖掘研究所（LAMDA），2010 年 11 月任软件新技术国家重点实验室常务副主任，2013 年 5 月任计算机系副主任。

2016 年，他当选 AAAI Fellow（国际人工智能学会），成为我国大陆第一位，也是此次入选的唯一来自美欧之外的学者，并且是唯一在中国大陆取得博士学位的 AAAI Fellow。2016 年 11 月，当选美国科学促进会会士（AAAS Fellow）。2016 年 12 月，当选 ACM Fellow，成为第一位在中国大陆取得全部学位的 ACM Fellow。

2017 年 2 月，当选人工智能领域顶级学术会议 AAAI 2019 程序委员会主席，是该会议自 1980 年成立以来首位华人主席、也是首次由美欧之外国家的学者出任主席。

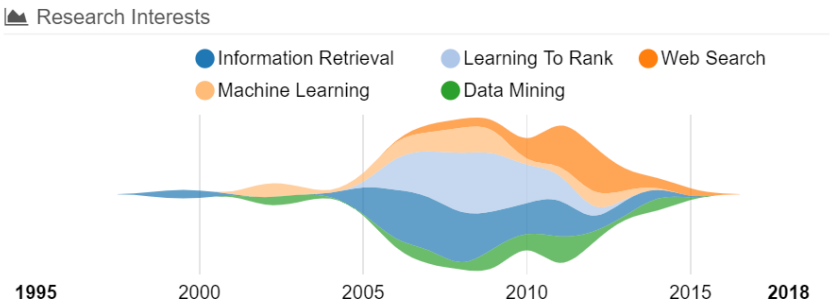
兼任 AAAI Fellow，IEEE Fellow，IAPR Fellow，ACM Fellow 和 AAAS Fellow，周志华成为国际上与人工智能相关的重要学会“大满贯”Fellow 华人第一人。

此外，他还担任 IJCAI 程序委员会主席，是中国内地首位任此职位学者。

周志华主要从事人工智能、机器学习、数据挖掘等领域的研究工作。他著有机器学习入门书籍《机器学习》。

- 李航





李航，北京大学、南京大学兼职教授。

李航毕业于日本京都大学电气电子工程系，于 1998 年获得日本东京大学计算机科学博士学位。曾任日本 NEC 公司中央研究所研究员，微软亚洲研究院高级研究员与主任研究员、华为技术有限公司诺亚方舟实验室主任。现任今日头条人工智能实验室主任。

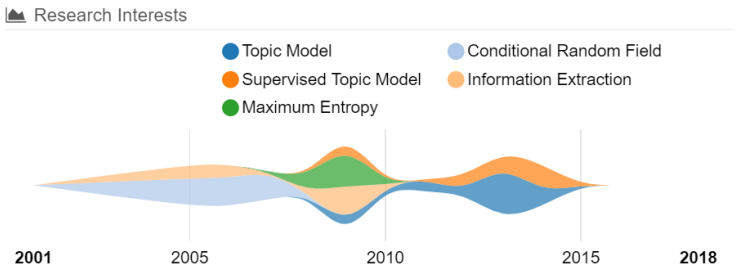
他的研究方向包括信息检索、自然语言处理、统计机器学习及数据挖掘。他一直活跃在相关学术领域，曾出版过三部学术专著，并在顶级国际学术会议和国际学术期刊上发表过上百篇学术论文，拥有 42 项授权美国专利。

他出版的三本技术书籍其中最广为人知的是 2012 年出版的《统计学习方法》，他发表超过 120 项技术论文，包括 SIGIR、WWW、WSDM、ACL、EMNLP、ICML、NeurIPS、SIGKDD、AAAI、IJCAI 等顶级国际会议以及包括 CL、NLE、JMLR、TOIS、IRJ、IPM、TKDE、TWEB、TIST。他和他同事的论文收到了 SIGKDD'08 最佳应用论文奖，SIGIR'08 最佳学生论文奖，ACL'12 最佳学生论文奖。

他是 ACM 杰出科学家。他的社会任职包括 AIRS-2008 程序委员会主席，SIGIR-2008 Poster & Demo 委员会主席，KDD-2009 宣传主席，EMNLP-2009 领域主席，ACM Transaction on Asian Language Information Processing 副主编，Journal of Computer Science and Technology 编委等。

- 朱军





朱军，清华大学计算机科学系教授，智能技术与系统国家重点实验室副主任，卡内基梅隆大学兼职教授。2013 年，入选 IEEE Intelligent Systems 的“人工智能 10 大新星”（AI’s 10 to Watch）。

朱军于 2009 年获得清华大学计算机博士学位，主要从事机器学习研究。

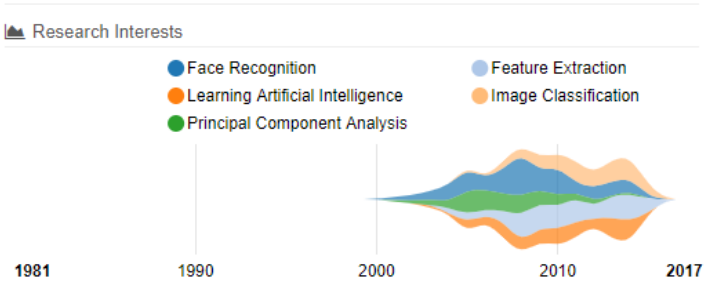
他在国际重要期刊与会议发表学术论文 80 余篇，他开源了 ZhuSuan，一个用于贝叶斯深度学习（贝叶斯方法和深度学习的结合）的 GPU 库，可以在 TensorFlow 上使用

他是 AAAI 2019，NeurIPS 2018，ICML 2018，UAI 2018，IJCAI 2018 的区域主席，担任国际期刊 IEEE TPAMI 和 Artificial Intelligence 的编委、国际会议 ICML 2014 地区联合主席、以及 ICML、NEURIPS 等国际会议的领域主席。

他于 2006 年被评为微软学者，2009 年入选卡内基梅隆大学 Innovation Fellow；中国计算机学会优秀博士论文奖获得者（2009）；清华大学 221 基础研究计划入选者（2012）；中国计算机学会青年科学家（2013）；IEEE Intelligent Systems 杂志评选的“AI’s 10 to Watch”（2013）；国家优秀青年科学基金获得者（2013），同年获得、中国计算机联合会（CCF）颁发的“CCF 青年科学家”奖；2014 年，他获得清华-MSRA 联合研究实验室颁发的最佳协作奖；2015 年获得全国青年顶尖人才支持计划的支持，同年收到了“CVIC SE 人才”奖；2017 年，他被麻省理工学院 TR35 中国选为“先驱者”之一。

- 颜水成





颜水成，新加坡国立大学副教授、360 集团副总裁、人工智能研究院院长、第十三批国家 “千人计划”专家。

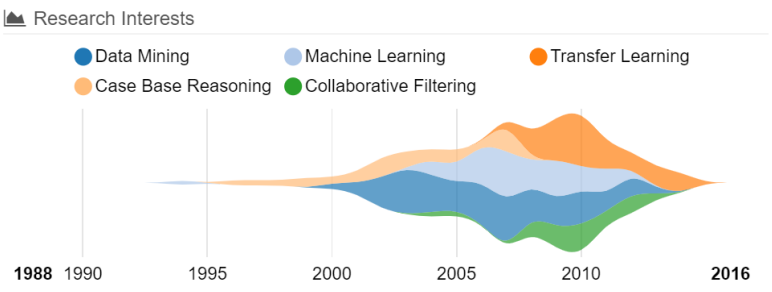
主要研究兴趣是为计算机视觉、多媒体和信息检索应用开发机器学习理论。

他在众多研究课题上撰写/合著了数百篇技术论文，其中谷歌学者引文 2 万余次。他是 2014 年、2015 年和 2016 年 ISI 被高度引用的研究员。颜水成博士率领的团队共获得了 10 次计算机视觉领域两大核心竞赛 Pascal VOC 和 ImageNet 大规模视觉识别（ILSVRC）冠军和荣誉奖，10 余次最佳（学生）论文奖。他的团队还曾获得多媒体领域顶会 ACM MM 最佳论文奖、最佳学生论文奖和最佳技术演示奖的大满贯。

颜水成博士团队提出的“Network in Network”（NIN）网络结构的核心 1x1 卷积是近年来几乎所有计算机视觉深度学习模型的标准模块，在学术界和工业界影响深远，其思想也被后期的 GoogleNet、残差网络（ResNet）等模型所采用。

- 杨强





杨强，香港科技大学新明工程学讲席教授，计算机科学和工程学系主任，大数据研究所所长。IEEE Fellow，IAPR Fellow，AAAS Fellow，ACM 杰出科学家，KDD 中国主席。

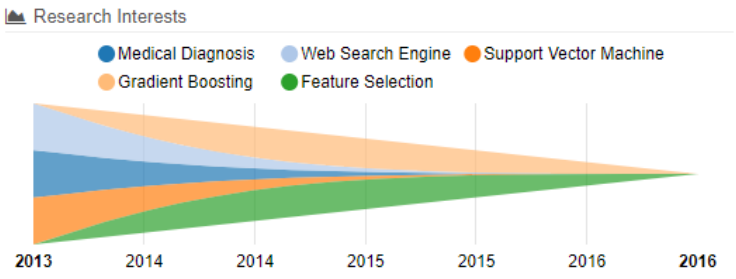
杨强于 1989 年获得马里兰大学计算机科学博士学位，之后直到 1995 年，于加拿大滑铁卢大学计算机系任助理教授及副教授。其主要研究领域为机器学习、数据挖掘和自动规划。他是人工智能研究的国际专家和领军人物，在学术界和工业界做出了杰出的服务和贡献，尤其近些年为中国人工智能（AI）和数据挖掘（KDD）的发展起了重要引导和推动作用。迄今为止，杨强已发表逾 400 篇关于人工智能和数据挖掘方面的论文，引用超过 20000 次。

2009 年，他创建了 ACM 刊物 *Transactions on Intelligent Systems and Technology（TIST）*并任首届主编。2012 年至 2015 年，出任华为诺亚方舟实验室创始主任；2015 年，任香港科技大学计算机与工程学系主任；2016 年，在香港科技大学创建大数据研究所。

他的社会兼职有：2013 年 7 月当选为国际人工智能协会（AAAI）院士，是第一位获此殊荣的华人，之后又于 2016 年 5 月当选为 AAAI 执行委员会委员，是首位也是至今为止唯一的 AAAI 华人执委，同年，任 ACM 数据挖掘中国分会（KDD China）主席。2017 年 8 月他当选为国际人工智能联合会（IJCAI，国际人工智能领域创立最早的顶级国际会议）理事会主席，是第一位担任 IJCAI 理事会主席的华人科学家。

- 黄高





黄高，清华大学助理教授，博士生导师。

主要研究领域为深度神经网络的结构设计与优化算法，以及深度学习在计算机视觉中的应用。

2009 年本科毕业于北京航空航天大学，2015 年获得清华大学控制科学与工程博士学位，2015 年至 2018 年为美国康奈尔大学计算机系博士后。其博士论文

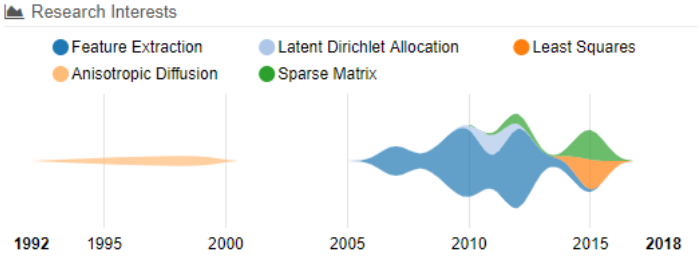
获选中国自动化学会优秀博士学位论文以及清华大学优秀博士论文一等奖。该获奖论文的主要贡献是提出了一种全新的卷积神经网络架构“密集链接卷积网络”（DenseNet），显著地提升了模型在图片识别任务上的准确率。

目前在 NIPS，ICML，CVPR 等国际顶级会议及 IEEE 多个汇刊共计发表学术论文 30 余篇。2016 年曾获得全国百篇最具国际影响学术论文、2017 年国际计算机视觉顶级会议 CVPR 最佳论文奖、2018 年世界人工智能创新大赛 SAIL 先锋奖和吴文俊人工智能自然科学一等奖等奖励和荣誉。

他是 AAAI 2018 高级程序委员，担任 NeurIPS、ICML、CVPR、ICCV、ECCV、ICLR、AAAI 等国际学术会议和 JMLR、TPAMI、TIP、TNNLS 等国际期刊审稿人。

- 林宙辰





林宙辰，北京大学信息科学技术学院机器感知与智能教育部重点实验室教授。

主要研究领域为机器学习、模式识别、计算机视觉、图像处理、数值优化。

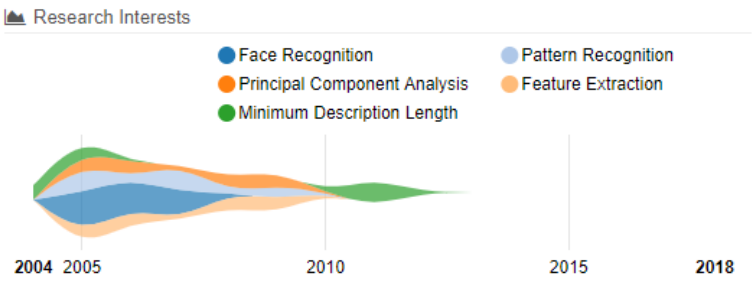
1993 年于南开大学获理学学士学位，1996 年于北京大学获理学硕士学位，1998 年于香港理工大学获得哲学硕士学位，2000 年于北京大学获得理学博士学位。

2007 年荣获 Microsoft SPOT Award，2015 年 ImageNet 大规模视觉识别竞赛（ILSVRC）场景分类项目冠军，2016 年获国家自然科学杰出青年基金资助。

他是 CVPR 2014/2016、ICCV 2015、NIPS 2015 的领域主席和 AAAI 2016/2017、IJCAI 2016 的高级程序委员。他也是 IEEE Transactions on Pattern Analysis And Machine Intelligence 和 International Journal of Computer Vision 的编委。

- 王立威





王立威，1999 年获清华大学学士学位，2002 年获清华大学硕士学位，2005 年获北京大学博士学位。现为北京大学信息科学技术学院教授。

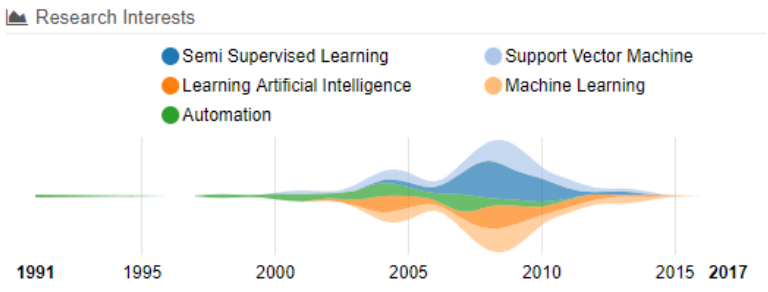
长期从事机器学习相关研究，目前主要致力于机器学习基础理论，即泛化理论的研究，差分隐私算法的设计与分析以及医疗影像诊断算法与系统的开发。

自 2002 年以来，在 PAMI、CVPR、ICML 等国际顶级期刊和和会议上发表论文 60 余篇，并参与编写《机器学习及其应用》2009 版“关于 Boosting 算法的 Margin 解释”及 2015 版“差分隐私保护的机器学习”相关章节。

曾获得第 11 届 Meeting on Image Recognition and Understanding 会议最佳论文奖，2010 年获得 *Pattern Recognition Letters* 期刊最高引用论文奖（2005-2010），2010 年入选 AI’s 10 to Watch，是首位获得该奖项的亚洲学者。2012 年获得首届国家自然科学基金优秀青年基金，新世纪优秀人才。任 NIPS 等权威会议 Area Chair，以及多家学术期刊编委。

- 张长水





张长水，男，1965 年生，河北人。智能技术与系统国家重点实验室学术委员会委员，清华大学自动化系教授、博士生导师，智能技术与系统国家重点实验室副主任，自动化系主任。

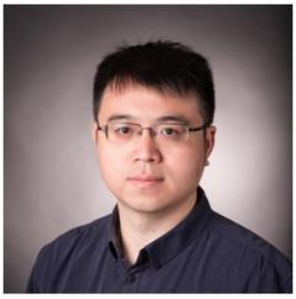
主要从事图像处理、信号处理、模式识别与人工智能、进化计算等研究领域以及和工业界的合作。

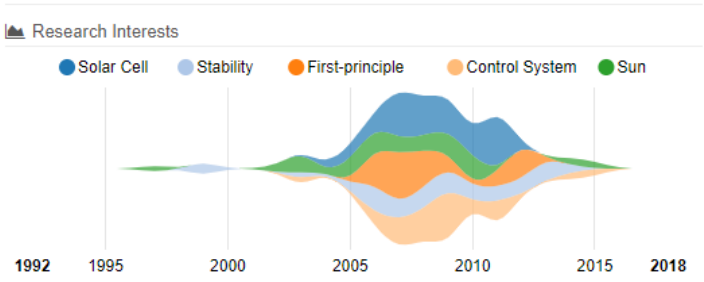
1986 年 7 月毕业于北京大学数学系，获得理学学士学位。1992 年 7 月毕业于清华大学自动化系，获得博士学位。

1992 年 7 月—1994 年 12 月， 在清华大学自动化系任讲师；1995 年 1 月—2000 年 8 月，在清华大学自动化系任副教授；2000 年 9 月起，在清华大学自动化系任教授；2001 年起，任清华大学博士生导师。

近几年在国际期刊和会议上发表学术论文超过 100 篇，其中包括国际权威期刊 Pattern Recognition、TNN、TKDE、IEEE Transaction on Multimedia 以及国际顶级会议 IJCAI、AAAI、NIPS、ICML、ECML、SIGIR、CVPR 等。他还是国际权威期刊 Pattern Recognition 编委。

- 孙剑





孙剑，男，前微软亚研院首席研究员，现就职于北京旷视科技有限公司，任旷视首席科学家、旷视研究院院长。

其主要研究方向是计算机视觉和深度学习。

1993 年—1997 年就读于西安交通大学自动控制专业，获工学学士学位。2000 年、2003 年在西安交通大学模式识别与智能控制专业研究生毕业，分别获得工学硕士学位和工学博士学位。2003 年孙剑加入微软亚洲研究院。

孙剑自 2002 年以来在 CVPR、ICCV、SIGGRAPH、PAMI 等顶级学术会议和期刊上发表学术论文 100 余篇，孙剑博士拥有超过 40 项美国或国际专利。

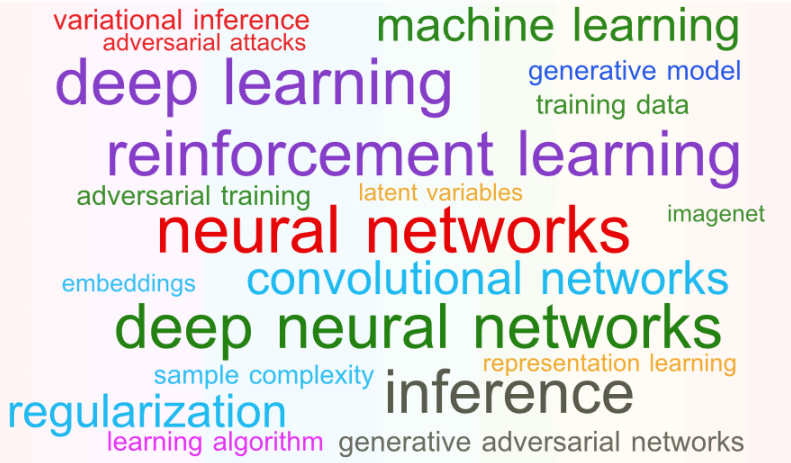
2009 年孙剑带领团队发表的论文 *Single Image Haze Removal Using Dark Channel Prior* 赢得了国际计算机视觉与模式识别会议（CVPR）的最佳论文奖（CVPR Best Paper），这是亚洲人第一次获得该奖；2010 年，孙剑被美国科技评论期刊《麻省理工科技评论》（*MIT Technology Review*）评选为“全球 35 岁以下杰出青年创新者”。2012 年至 2014 年，孙剑加入法国国家信息与自动化研究院(INRIA)/巴黎高等师范学院 Willow 组。2016 年，孙剑带领的团队凭借 *Deep Residual Learning for Image Recognition* 再次获得了国际计算机视觉与模式识别会议（CVPR）的最佳论文奖（CVPR Best Paper）。2016 年 7 月，孙剑正式加入旷视任首席科学家、旷视研究院院长。2017 年 8 月，孙剑担任中国自动化学会（Chinese Association of Automation, CAA）混合智能专委会副主任。2018 年 5 月，2018 年第一批国家重点研发计划公示孙剑博士担任变革性技术关键科学问题专项项目负责人。2019 年 1 月，孙剑出任西安交通大学人工智能学院首任院长。

2.7 论文解读

本节对本领域的高水平学术会议论文进行挖掘，解读这些会议在近年的部分代表性工作，会议具体包括：

International Conference on Machine Learning

Conference and Workshop on Neural Information Processing Systems



我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如上图所示。其中，出神经网络（neural networks）、深

度学习（deep learning）、强化学习（reinforcement learning）是本领域中最热的关键词。ICML 和 NeurIPS 是机器学习领域非常具有代表性的会议，限于报告篇幅，我们选取 ICML 和 NeurIPS 近十年若干最佳论文进行解读。

表 2-2 ICML 近 10 年 best paper

ICML (International Conference on Machine Learning)		
年份	论文标题	作者
2019	Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations	Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, Olivier Bachem
	Rates of Convergence for Sparse Variational Gaussian Process Regression	David R. Burt, Carl E. Rasmussen, Mark van der Wilk
2018	Delayed Impact of Fair Machine Learning	Lydia T. Liu, University of California Berkeley; et al.
	Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples	Anish Athalye, Massachusetts Institute of Technology; et al.
2017	Understanding Black-box Predictions via Influence Functions	Pang Wei Koh & Percy Liang, Stanford University
2016	Ensuring Rapid Mixing and Low Bias for Asynchronous Gibbs Sampling	Christopher De Sa, Stanford University; et al.
	Pixel Recurrent Neural Networks	Aaron Van den Oord, Google; et al.
	Dueling Network Architectures for Deep Reinforcement Learning	Ziyu Wang, Google; et al.
2015	A Nearly-Linear Time Framework for Graph-Structured Sparsity	Chinmay Hegde, Massachusetts Institute of Technology; et al.
	Optimal and Adaptive Algorithms for Online Boosting	Alina Beygelzimer, Yahoo! Research; et al.
2014	Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis	Jian Tang, Peking University; et al.
2013	Vanishing Component Analysis	Roi Livni, The Hebrew University of Jerusalum; et al.
	Fast Semidifferential-based Submodular Function Optimization	Rishabh Iyer, University of Washington; et al.
2012	Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring	Sungjin Ahn, University of California Irvine; et al.
2011	Computational Rationalization: The Inverse Equilibrium Problem	Kevin Waugh, Carnegie Mellon University; et al.
2010	Hilbert Space Embeddings of Hidden Markov Models	Le Song, Carnegie Mellon University; et al.
2009	Structure preserving embedding	Blake Shaw, Tony Jebara, Columbia University

表 2-3 NeurIPS 近 10 年 best paper

NeurIPS（Conference and Workshop on Neural Information Processing Systems）		
年份	论文标题	作者
2018	Non-delusional Q-learning and Value-iteration	Tyler Lu, Dale Schuurmans, Craig Boutilier
	Optimal Algorithms for Non-Smooth Distributed Optimization in Networks	Kevin Scaman, Francis Bach, Sebastien Bubeck, Laurent Massoulié, Yin Tat Lee

NeurIPS（Conference and Workshop on Neural Information Processing Systems）		
	Nearly Tight Sample Complexity Bounds for Learning Mixtures of Gaussians via Sample Compression Schemes	Hassan Ashtiani, Shai Ben-David, ick Harvey，Christopher Liaw, Abbas Mehrabian, Yaniv Plan
	Neural Ordinary Differential Equations	Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt，David Duvenaud
2017	Safe and Nested Subgame Solving for Imperfect-Information Games	Noam Brown, Tuomas Sandholm
	Variance-based Regularization with Convex Objectives	Hongseok Namkoong, John Duchi
	A Linear-Time Kernel Goodness-of-Fit Test	Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo，Kenji Fukumizu, Arthur Gretton
2016	Value Iteration Networks	Aviv Tamar, Yi Wu, Garrett Thomas, Sergey Levine, Pieter Abbeel
	Matrix Completion has No Spurious Local Minimum	Rong Ge, Jason Lee, Tengyu Ma
	Interactive musical improvisation with Magenta	Adam Roberts, Jesse Engel, Curtis Hawthorne, Ian Simon, Elliot Waite, Sageev Oore, Natasha Jaques, Cinjon Resnick, Douglas Eck
2015	Competitive Distribution Estimation: Why is Good-Turing Good	Alon Orlitsky, Ananda Theertha Suresh
	Fast Convergence of Regularized Learning in Games	Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo，Robert Schapire
2014	Asymmetric LSH (ALSH) for sublinear time Maximum Inner Product Search (MIPS)	Anshumali Shrivastava, Ping Li
	A* Sampling	Chris J. Maddison, Daniel Tarlow, Tom Minka
2013	A Memory Frontier for Complex Synapses	Subhaneil Lahiri, Surya Ganguli
	Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints	Rishabh Iyer, Jeff Bilmes
	Scalable Influence Estimation in Continuous-Time Diffusion Networks	Nan Du, Le Song, Manuel Gomez-Rodriguez，Hongyuan Zha
2012	No voodoo here! Learning discrete graphical models via inverse covariance estimation	Po-Ling Loh, Martin Wainwright
	Discriminative Learning of Sum-Product Networks	Robert Gens, Pedro Domingos
2011	Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials	Philipp Krähenbühl, Vladlen Koltun
	Priors Over Recurrent Continuous Time Processes	Ardavan Saeedi, Alexandre Bouchard-Côte
	Fast and Accurate K-means for Large Datasets	Michael Shindler, Alex Wong, Adam Meyerson
2010	Construction of dependent dirichlet Processes based on Poisson Processes	Dahua Lin, Eric Grimson, John Fisher
	A Theory of Multiclass Boosting	Indraneel Mukherje，Robert E Schapire
2009	An LP View of the M-Best MAP Problem	Menachem Fromer，Amir Globerson
	Fast Subtree Kernels on Graphs	Nino Shervashidze，Karsten Borgwardt

2.7.1 ICML 历年最佳论文解读

- 2019 年最佳论文

论文题目：*Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations*

中文题目：挑战无监督分离式表征的常见假设

论文作者：Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, Olivier Bachem

论文地址：https://aminer.cn/pub/5c04967517c44a2c74709162/challenging-common-assumptions-in-the-unsupervised-learning-of-disentangled-representations

论文解读：文章主要从理论和实践两方面对这一领域中的一些基本假设提出了挑战。文章从理论上证明，如果没有对所考虑的学习方法和数据集产生归纳偏置，那么解耦表示的无监督学习基本上是不可能的。文章还采用了完善的无监督解耦学习实验方案，进行了一个超级大规模的实验研究。最后还发布了 disentanglement_lib，这是一个用于训练和评估解耦表示的新库。由于复制这个结果需要大量的计算工作论文还发布了超过 10000 个预训练的模型，可以作为未来研究的基线方法。

论文题目：*Rates of Convergence for Sparse Variational Gaussian Process Regression*

中文题目：稀疏变分高斯过程回归的收敛速度

论文作者：David R. Burt, Carl E. Rasmussen,Mark van der Wilk

论文地址：https://www.aminer.cn/pub/5cede106da562983788e64b9/rates-of-convergence-for-sparse-variational-gaussian-process-regression

论文解读：这篇文章来自英国剑桥大学。自从许多研究人提出了对高斯过程后验的变分近似法后，避免了数据集大小为 N 时 $O(N^3)$ 的缩放。它们将计算成本降低到 $O(NM^2)$，其中 $M \leqslant N$ 是诱导变量的数量。虽然 N 的计算成本似乎是线性的，但算法的真正复杂性取决于 M 如何增加以确保一定的近似质量。论文

证明了稀疏 GP 回归变分近似到后验变分近似的 KL 散度的界限，该界限仅依赖于先验核的协方差算子的特征值的衰减。这些边界证明了直观的结果，平滑的核、训练数据集中在一个小区域，允许高质量、非常稀疏的近似。这些边界证明了用 $M \leqslant N$ 进行真正稀疏的非参数推理仍然可以提供可靠的边际似然估计和点后验估计。对非共轭概率模型的扩展，是未来研究的一个有前景的方向。

- 2018 年最佳论文

论文题目：*Delayed Impact of Fair Machine Learning*

中文题目：公正机器学习的滞后影响

论文作者：Lydia T.Liu, Sarah Dean, Esther Rolf, Max Simchowitz, Moritz Hardt

论文地址：https://arxiv.org/abs/1803.04383

论文解读：机器学习的公平性主要在静态分类设置中得到研究，但却没有关注这些决策如何随时间改变潜在的群体。传统的观点认为公平性准能提升他们想保护的群体的长期利益。本文研究了静态公平性标准如何与暂时的利益指标相互作用，例如利益变量的长期提升、停滞和下降。本文证实了即使在一步反馈模型中，常见的公平性准则没有随时间带来改善，实际上可能给特定案例带来了伤害。

论文题目：*Obfuscated Gradients Give a False Sense of Security: Circumventing Defenses to Adversarial Examples*

中文题目：混淆梯度的虚假安全感：对抗样本防御

论文作者：Anish Athalye, Nicholas Carlini, David Wagner

论文地址：https://arxiv.org/abs/1802.00420v1

论文解读：如果在一张图片添加干扰，可能就可以骗过分类器。为了抵御对抗样本的攻击，使得神经网络在受到迭代攻击时不受对抗样本干扰，研究人员在寻找强大的对抗样本防御器，使其在面对基于优化的攻击之下，可以实现对对抗样本的鲁棒性防御。

- 2017 年最佳论文

论文题目：*Understanding Black-box Predictions via Influence Functions*

中文题目：利用影响函数理解黑箱预测

论文作者：Pang Wei Koh, Percy Liang

论文地址：https://arxiv.org/abs/1703.04730

论文解读：这篇论文利用影响函数（稳健统计学中的经典技术），通过学习算法跟踪模型的预测并追溯到训练数据，从而确定对给定预测影响最大训练点来解释黑箱模型的预测。为了将影响函数扩展到现代机器学习中，论文中设计了一个简单高效的实验，仅需梯度 oracle 访问和 Hessian 矢量积。而且即使在非凸和非微分模型上，影响函数的近似值算法仍然可以提供有价值的信息。在线性模型和卷积神经网络中，论文中也证明，影响函数可用于理解模型行为，调试模型，检测数据集错误，甚至是生成视觉上无法区分的训练集攻击。

- 2016 年最佳论文

论文题目：*Ensuring Rapid Mixing and Low Bias for Asynchronous Gibbs Sampling*

中文题目：确保异步吉布斯采样的快速混合和低偏差

论文作者：Christopher De Sa, Kunle Olukotun, Christopher Ré

论文地址：https://arxiv.org/abs/1602.07415

论文解读：吉布斯采样（Gibbs Sampling）是一种常被用于估计边缘分布（marginal distribution）的马尔可夫链蒙特卡罗技术（Markov chain Monte Carlo technique）。为了加速吉布斯采样，人们最近产生了通过异步执行并行处理它的兴趣。尽管一些经验结果表明许多模型都可以有效地进行异步采样，但传统的马尔克夫链分析却无法应用于异步的情况，因此对异步吉布斯采样只有很少的了解。在这篇论文中，我们设法更好地了解了异步吉布斯的两个主要挑战：偏差（bias）和混合时间（mixing time）。我们通过实验证明了我们的理论结果是符合实际结果的。

论文题目：*Pixel Recurrent Neural Networks*

中文题目：像素循环神经网络

论文作者：Aaron van den Oord, Nal Kalchbrenner, Koray Kavukcuoglu

论文地址：https://arxiv.org/abs/1601.06759

论文解读：在无监督学习中，给自然图像分布建模是一个里程碑式的问题。这项任务要求得到可以同时表现图像、易于处理并且具备可扩展性的图像模型。我们展示了一个可以沿二维空间维度依次预测图像中像素的深度神经网络。我们的方法建立了原始像素值的离散概率模型，并且编码了图像中完整的依赖关系集合。该架构的不同之处在于它包括快速二维循环层（recurrent layers）和对深度循环网络中残差连接（residual connections）的有效利用。我们完成了自然图像上的对数似然分数，其比之前最先进的还要好很多。我们主要的成果还包括提供多样化的 ImageNet 数据集基准。从模型中生成了新鲜多样且全局同一的样本。此论文提出了一系列生成模型，可直接对像素的统计依赖关系进行建模。这些模型包括两个 PixelRNN：Row LSTM 和 Diagonal BiLSTM（区别主要在于它们进行预测使用到的条件信息所在的领域）；一个 PixelCNN，以及一个多尺度 PixelRNN。

论文题目：*Dueling Network Architectures for Deep Reinforcement Learning*

中文题目：深度强化学习中的竞争网络架构

论文作者：Ziyu Wang, Tom Schaul, Matteo Hessel, Hado van Hasselt, Marc Lanctot, Nando de Freitas

论文地址：https://arxiv.org/abs/1511.06581

论文解读：近几年，已经有很多在强化学习中使用深度表征获得成功的例子。然而，这些应用中的很多例子仍然使用传统的架构，比如卷积网络、LSTMs，或者是自动编码器。在此论文中，我们提出了一个新的用于无模型（model free）强化学习的神经网络架构。我们的竞争网络（dueling network)表示了两种独立的评估量：一个用于状态价值函数（state value function)，一个用于状态依存动作优势函数（state-dependent action advantage function）。这一分解的主要好处是在没有将任何变化强加于低层的强化学习算法的情况下，在动作（action）间归纳学习。

我们的结果显示，这一架构在多种价值相似的动作面前能引发更好的政策评估。此外，这一竞争架构使得我们的强化学习代理胜过 Atari 2600 领域最前沿的研究。在这篇论文中，作者基于分开建模状态值和动作优势的想法，提出了一款可供选择的用于深度 Q 网络（DQN）的架构和相关的学习方案。当被应用于 Atari 学习环境（Atari Learning Environment）基准时，这项技术显著推进了当前最先进的研究成果。

- 2015 年最佳论文

论文题目：*A Nearly-Linear Time Framework for Graph-Structured Sparsity*

中文题目：图结构稀疏性的近似线性时间框架

论文作者：Hegde, Chinmay, Indyk, Piotr, Schmidt, Ludwig

论文地址：http://proceedings.mlr.press/v37/hegde15.pdf

论文解读：本文引入了一个通过图定义的稀疏结构框架。其方法较灵活，并且推广到了以前研究过的几个稀疏模型。此外，本文还为该稀疏度模型提供了有效的投影算法，该模型几乎在线性时间内运行。在稀疏恢复的背景下，本文证明了该框架在理论上实现了广泛参数下的信息最优样本复杂性。本文用实验来补充该理论分析，证明该算法在实践中也改进了先前的工作。

论文题目：*Optimal and Adaptive Algorithms for Online Boosting*

中文题目：Online Boosting 的优化和自适应算法

论文作者：Alina Beygelzimer, Satyen Kale, Haipeng Luo

论文地址：https://arxiv.org/abs/1502.02651

论文解读：我们学习在线促进，这是一项将任何一个弱的在线学习者转变为强的在线学习者的任务。基于对网络学习能力弱的一个新的自然定义，我们开发了两种在线增强算法。第一种算法是在线版本的 Boost by Majority。通过证明一个匹配下界，我们证明了该算法对于弱学习者的数量和达到指定精度所需的样本复杂度是本质上最优的。然而，这种优化算法并不具有自适应性。利用在线损失最小化的工具，推导了一种无参数但非最优的自适应在线增强算法。这两种算法

都与基础学习者一起工作，基础学习者可以直接处理示例重要性权重，也可以使用升迁者定义的概率拒绝抽样示例。结果与广泛的实验研究相辅相成。

- 2014 年最佳论文

论文题目：*Understanding the Limiting Factors of Topic Modeling via Posterior Contraction Analysis*

中文题目：经由过程后验收缩分析理解主题建模的限制因素

论文作者：Ming Zhang, Jian Tang, Zhaoshi Meng

论文地址：http://proceedings.mlr.press/v32/tang14.pdf

论文解读：潜在狄利克雷分布（LDA）已经成为了机器学习建模工具箱中的一个标准工具。它们已被应用于各种不同程度的数据集、背景和任务，但是迄今为止，几乎没有正式的理论来解释 LDA 的行为，并且尽管对其很熟悉，但是对影响模型推理性能的数据的性质几乎没有系统性的分析和指导。本文试图通过对影响 LDA 的性能的因素进行系统分析来解决此问题。本文提出的定理阐明了随着数据量的增加后验概率，并使用综合和真实的数据集进行了全面的支持性实证研究。基于这些结果，本文对如何为主题模型识别合适的数据集以及如何制定特定的模型参数提供了实际指导。

- 2013 年最佳论文

论文题目：*Vanishing Component Analysis*

中文题目：经由过程后验收缩分析理解主题建模的限制因素

论文作者：Roi Livni, Shai Shalevshwartz, Amir Globerson

论文地址：http://proceedings.mlr.press/v28/livni13.pdf

论文解读：传统的特征选择方法通常是在采样中选择显著的特征，作者研究的是，在特征选择时，是否能够选择一些不变的特征。文章描述并分析了构造一组消失理想生成器的有效过程。该过程是数值稳定的，并且可以用于近似消失多项式。这由此得到的多项式捕捉数据中的非线性结构，例如可用于监督学习。与

核方法的实证比较表明，文章提出的方法构造了更紧凑的分类器，具有相当的精度。

论文题目：*Fast Semidifferential-based Submodular Function Optimization*

中文题目：基于半微分的快速子模块函数优化

论文作者：Iyer, Rishabh, Jegelka, Stefanie, Bilmes, Jeff

论文地址：http://export.arxiv.org/pdf/1308.1006

论文解读：本文提出了一种实用而强大的基于离散半微分（子微分和超微分）的无约束和约束子模函数优化新框架。所得到的算法反复计算并有效地优化了子模半梯度，为子模优化提供了新的、通用的方法。此外，本文的方法还采取步骤，提供适用于次模最小化和最大化的统一范式，这些问题在历史上得到了相当明显的处理。本文的算法的实用性很重要，因为子模性由于其自然和广泛的适用性，最近在机器学习中占据了优势。分析了本文的极小化和最大化算法的理论性质，表明许多最先进的最大化算法都是特殊情况。最后，本文将理论分析与实证实验相补充。

- 2012 年最佳论文

论文题目：*Bayesian Posterior Sampling via Stochastic Gradient Fisher Scoring*

中文题目：通过随机梯度 Fisher 得分进行贝叶斯后验采样

论文作者：S.Ahn, A.Korattikara, M.Welling

论文地址：https://arxiv.org/ftp/arxiv/papers/1206/1206.6380.pdf

论文解读：在本文中，讨论了以下问题：如果我们只允许对生成的每个样本接触一小批数据项，那么我们是否可以近似地从贝叶斯后验分布中提取样本？本文提出了一种基于随机梯度朗格文方程（SGLD）的混合算法，但是其混合速率慢。通过利用贝叶斯中心极限定理，我们扩展了 SGLD 算法，使其在高混合速率下从后验函数的正态近似中采样，而在慢混合速率下，它将使用预调节矩阵模拟 SGLD 的行为。作为一个额外的好处，该算法使人想起费希尔评分（随机梯度），因此在老化过程中是一个有效的优化器。

- 2011 年最佳论文

论文题目：*Computational Rationalization: The Inverse Equilibrium Problem*

中文题目：计算合理化：逆向平衡问题

论文作者：Kevin Waugh, Brian D.Ziebart, J. Andrew (Drew) Bagnell

论文地址：https://www.ri.cmu.edu/pub_files/2011/6/paper.pdf

论文解读：从少量的观察结果中模拟不完美因素的有目的行为是一项具有挑战性的任务。当限制在单智能体决策理论设置下时，逆最优控制技术假定观测行为是未知决策问题的近似最优解。这些技术学习解释示例行为的实用函数，然后可用于准确预测或模拟类似观察到或未观察到的情况下的未来行为。在这项工作中，我们考虑了竞争和合作多代理领域中的类似任务。在这里，不同于单一代理设置，玩家不能近似地最大化其回报-它必须推测其他代理如何行动，以影响游戏的结果。利用遗憾博弈论的概念和最大熵原理，提出了一种预测和概括行为的方法，并在此领域中恢复了奖励函数。

- 2010 年最佳论文

论文题目：*Hilbert Space Embeddings of Hidden Markov Models*

中文题目：隐马尔科夫模型的希尔伯特空间嵌入

论文作者：Le Song, Byron Boots, Sajid M. Siddiqi, Geoffrey J. Gordon

论文地址：https://storage.googleapis.com/pub-tools-public-publication-data/pdf/36408.pdf

论文解读：隐马尔可夫模型是序列数据建模的重要工具。然而，它们仅限于离散的潜在状态，并且主要限于高斯和离散观测。而且，HMM 的学习算法主要依赖于局部搜索启发式算法，除了下面描述的谱方法。本文提出了一个非参数 HMM，它将传统的 HMM 扩展到结构化和非高斯连续分布。此外，本文还导出了一个学习这些 HMM 的局部最小自由核谱算法。本文将该方法应用于机器人视觉数据、槽车惯性传感器数据和音频事件分类数据，结果表明，在这些应用中，嵌入式 HMM 的性能超过了以往的先进水平。

- 2009 年最佳论文

论文题目：*Structure preserving embedding*

中文题目：结构保留嵌入

论文作者：Blake Shaw, Tony Jebara

论文地址：http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.149.7500&rep=rep1&type=pdf

论文解读：结构保留嵌入（SPE）是一种用于在欧几里得空间中嵌入图形的算法，该嵌入是低维的，并保留了输入图的全局拓扑属性。如果诸如 k 最近邻之类的连通性算法可以在嵌入后仅从节点的坐标轻松恢复输入图的边缘，则可以保留拓扑。SPE 公式化为半定程序，该程序学习受一组线性不等式约束的低阶内核矩阵，该不等式捕获输入图的连通性结构。SPE 在图形的可视化和无损压缩方面提供了显着的改进，胜过了诸如光谱嵌入和 Laplacian 特征图之类的流行方法。我们发现，仅需使用几个维度就可以正确嵌入许多经典图和网络。

2.7.2 NeurIPS 历年最佳论文解读

- 2018 年最佳论文

论文题目：*Non-delusional Q-learning and Value-iteration*

中文题目：非妄想 Q 学习和价值迭代

论文作者：Tyler Lu, Dale Schuurmans, Craig Boutilier

论文地址：http://120.52.51.18/papers.NeurIPS.cc/paper/8200-non-delusional-q-learning-and-value-iteration.pdf

论文解读：本文用函数逼近法确定了 Q-学习和其他形式的动态规划中误差的根本来源。当近似结构限制了可表达的贪婪策略的类别时，就会产生偏差。由于标准 Q-updates 对可表达的策略类做出了全局不协调的动作选择，因此可能导致不一致甚至冲突的 Q 值估计，从而导致病态行为，例如过度/低估、不稳定甚至发散。为了解决这个问题，本文引入了策略一致性的新概念，并定义了一个本地备份流程，通过使用信息集，也就是记录与备份 Q 值一致的策略约束集，来确

保全局一致性。本文证明了使用这种备份的基于模型和无模型的算法都能消除妄想偏差，从而产生第一种已知算法，保证在一般条件下的最优结果。此外，这些算法只需要多项式的多个信息集（从潜在的指数支持）。最后，本文建议使用其他实用的启发式价值迭代和 Q 学习方法去尝试减少妄想偏差。

论文题目：*Optimal Algorithms for Non-Smooth Distributed Optimization in Networks*

中文题目：非光滑凸函数的分布式优化算法

论文作者：Kevin Scaman, Francis Bach, Sebastien Bubeck, Laurent Massoulié, Yin Tat Lee

论文地址：https://arxiv.org/pdf/1806.00291.pdf

论文解读：在本文中，我们考虑使用计算单元网络的非光滑凸函数的分布式优化。我们在两个正则性假设下研究这个问题：（1）全局目标函数的 Lipschitz 连续性；（2）局部个体函数的 Lipschitz 连续性。在局部正则性假设下，本文给出了称为多步原对偶（MSPD）的一阶最优分散算法及其相应的最优收敛速度。这个结果的一个显著特点是，对于非光滑函数，当误差的主要项在 O(1/t)中时，通信网络的结构仅影响 O(1/t)中的二阶项，其中 t 是时间。换言之，即使在非强凸目标函数的情况下，由于通信资源的限制而导致的误差也以快速率减小。在全局正则性假设下，基于目标函数的局部平滑，给出了一种简单而有效的分布式随机平滑（DRS）算法，并证明了 DRS 在最优收敛速度的 d1/4 乘因子内，其中 d 为底层。

论文题目：*Nearly Tight Sample Complexity Bounds for Learning Mixtures of Gaussians via Sample Compression Schemes*

中文题目：通过样本压缩方案学习混合高斯模型的近乎紧密的样本复杂性边界

论文作者：Hassan Ashtiani, Shai Ben-David, NicholasJ.A.Harvey, Chritopher Liaw, AbbasMehrabian, YanivPlan

论文地址：https://papers.NeurIPS.cc/paper/7601-nearly-tight-sample-complexity-bounds-for-learning-mixtures-of-gaussians-via-sample-compression-schemes.pdf

论文解读：本文证明了 $O(k\ d^2/\varepsilon^2)$样本对于学习 R^d 中 k 个高斯的混合，直至总变差距离中的误差 ε 来说，是充分必要条件。这改善了已知的上界和下界这一问题。对于轴对准高斯混合，本文证明了 $O(k\ d/\varepsilon^2)$样本匹配一个已知的下界是足够的。上限是基于样本压缩概念的分布学习新技术。任何允许这种样本压缩方案的分布类都可以用很少的样本来学习。此外，如果一类分布具有这样的压缩方案，那么这些产品和混合物的类也是如此。本文主要结果的核心是证明了 R^d 中的高斯类能有效的进行样本压缩。

论文题目：*Neural Ordinary Differential Equations*

中文题目：神经常微分方程

论文作者：Tian Qi Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud

论文地址：https://arxiv.org/pdf/1806.07366.pdf

论文解读：本文介绍了一系列新的深度神经网络模型。本文使用神经网络参数化隐藏状态的导数，而不是指定隐藏层的离散序列。使用黑盒微分方程求解器计算网络的输出。这些连续深度模型具有恒定的内存成本，使其评估策略适应每个输入，并且可以明确地交换数值精度以获得速度。本文在连续深度残差网络和连续时间潜变量模型中证明了这些性质。本文还构建了连续归一化流，这是一种可以通过最大似然进行训练的生成模型，无需对数据维度进行分区或排序。为了训练，本文展示了如何通过任何 ODE 求解器进行可扩展反向传播，而无需访问其内部操作。这允许在较大模型中对 ODE 进行端到端训练。

- 2017 年最佳论文

论文题目：*Safe and Nested Subgame Solving for Imperfect-Information Games*

中文题目：不完全信息博弈的安全嵌套子博弈求解

论文作者：Noam Brown, Tuomas Sandholm

论文地址：https://arxiv.org/pdf/1705.02955.pdf

论文解读：和完美信息博弈不同，不完美信息博弈不能通过将博弈分解为可独立求解的子博弈而求得占优策略。因此本文越来越多地使用计算密集的均衡判

定技术，并且所有的决策必须将博弈的策略当作一个整体。本文提出了一种无论在理论上还是在实践上都超越了之前方法的子博弈求解技术。本文还展示了如何对它们和以前的子博弈求解技术进行调整，以对超出初始行动提取（original action abstraction）的对手的行动做出应答；这远远超越了之前的顶尖方法，即行动转化（action translation）。最后，本文展示了当博弈沿着博弈树向下进行时，子博弈求解可能会重复进行，从而大大降低可利用性。

论文题目：*Variance-based Regularization with Convex Objectives*

中文题目：带有凸对象的基于方差的正则化方法

论文作者：Hongseok Namkoong, John Duchi

论文地址：https://arxiv.org/pdf/1610.02581.pdf

论文解读：本文研究了一种风险最小化和随机优化的方法，该方法可以为方差提供一个凸属性的替代项，并允许在逼近和估计误差间实现近似最优与高效计算间的权衡。本文的方法建立在分布鲁棒性优化和 Owen 经验性似然度的基础上，并提供了一些有限样本（finite-sample）和渐进结果以展示估计器的理论性能。具体来说，本文证明了该过程具有最优性保证（certificates of optimality），并通过逼近和最优估计误差间良好的权衡在更一般的设定下比经验风险最小化方法有更快的收敛率。本文还给出了确凿的经验性证据，表明估计器在实践中会在训练样本的方差和绝对性能之间进行权衡。此外，估计器也会提升标准经验风险最小化方法在许多分类问题上的测试性能。

论文题目：*A Linear-Time Kernel Goodness-of-Fit Test*

中文题目：一种线性时间核的拟合优度测试方法

论文作者：Wittawat Jitkrittum, Wenkai Xu, Zoltan Szabo, Kenji Fukumizu, Arthur Gretton.

论文地址：https://arxiv.org/pdf/1705.07673.pdf

论文解读：本文提出了一个全新的拟合优度（goodness-of-fit）的适应性测试法，其中计算资源的消耗与样本数呈线性关系。本文通过最小化假负类率来学习

最能展示观察样本和参考模型之间差异的测试特征。这些特征是通过 Stein 法构造的——这意味着没有必要计算模型的归一化常数。本文分析了新测试的 Bahadur 渐进效率，并证明了在均值偏移（mean-shift）的情况下，无论选择哪个测试参数，本文的测试总是比先前的线性时间核测试具有更高的相对效率。在高维和模型结构可用的情况下，本文的拟合优度测试在模型中抽取样本，表现远远超越基于最大平均差异（Maximum Mean Discrepancy）的二次时序双样本测试。

- 2016 年最佳论文

论文题目：*Value Iteration Networks*

中文题目：价值迭代网络

论文作者：Aviv Tamar, Yi Wu, Garrett Thomas, Sergey Levine, Pieter Abbeel

论文地址：https://arxiv.org/pdf/1602.02867.pdf

论文解读：本文介绍了一个价值迭代网络（VIN）：一种完全可微分的神经网络，内置“规划模块”。VIN 可以学习计划，并且适用于预测涉及基于计划的推理的结果，例如加强学习的政策。我们的方法的关键是一种新的可微近似值迭代算法，它可以表示为卷积神经网络，并使用标准的反向传播训练端到端。本文基于离散和连续路径规划域以及基于自然语言的搜索任务评估基于 VIN 的策略。本文表明，通过学习一个明确的规划计算，VIN 策略可以更好地推广到新的、未发现的领域。

论文题目：*Matrix Completion has No Spurious Local Minimum*

中文题目：矩阵填充没有假的局部最小值

论文作者：Rong Ge, Jason Lee, Tengyu Ma

论文地址：https://arxiv.org/pdf/1605.07272.pdf

论文题目：矩阵填充是一个基本的机器学习问题，具有广泛的应用，尤其是在协作过滤和推荐系统中。简单的非凸优化算法在实践中很流行且有效。我们证明了用于矩阵填充的常用非凸目标函数没有假的局部最小值——所有局部最小值也必须是全局的。因此，许多流行的优化算法（例如随机梯度下降）可以通过

多项式时间内的任意初始化可证明地解决矩阵填充问题。当观察到的条目包含噪声时，结果可以推广到该设置。我们认为，我们的主要证明策略对于理解其他涉及部分或嘈杂观测值的统计问题的几何性质很有用。

论文题目：*Interactive musical improvisation with Magenta*

文题目：基于 Magenta 的即兴音乐交互体验

论文作者：Adam Roberts, Jesse Engel, Curtis Hawthorne, Ian Simon, Elliot Waite, Sageev Oore, Natasha Jaques, Cinjon Resnick, Douglas Eck

论文地址：https://nips.cc/Conferences/2016/ScheduleMultitrack?event=6307

论文解读：作者结合了基于 LSTM 的循环神经网络和 Deep Q-learning 建立了实时生成音乐序列。LSTM 的任务是学习音乐评分（编码为 MIDI，而不是音频文件）的一般结构。Deep Q-learning 用来改进基于奖励的序列，如期望的类型，组成正确性和预测人类协作者演奏的内容。基于 RNN 模型的生成与强化学习的结合是一种生成音乐的全新方式。这种方式比单独使用 LSTM 更为稳定，生成的音乐更加好听。该方法有两个任务：生成对短旋律输入的响应，以及实时生成对旋律输入的伴奏，持续对未来输出进行预测。本方法在 TensorFlow 中加入了一个全新的 MIDI 接口产生即兴的音乐体验，让使用者可以与神经网络实时交互。

- 2015 年最佳论文

论文题目：*Competitive Distribution Estimation: Why is Good-Turing Good*

中文题目：竞争分布估计：为什么 Good-Turing 好

论文作者：Alon Orlitsky, Ananda Theertha Suresh

论文地址：http://120.52.51.17/papers.NeurIPS.cc/paper/5762-competitive-distribution-estimation-why-is-good-turing-good.pdf

论文解读：该论文属于统计学习的理论研究范畴，它对估计离散变量的分布律这一普遍问题，提出了基于 Good-Turing 估计量的两种改进方法，借助对先验

的最优估计量，给出了针对任意分布律的近似最优的高效估计。论文不仅指出这两种方法可以快速收敛，同时还给出相应的理论分析。

论文题目：*Fast Convergence of Regularized Learning in Games*

中文题目：博弈中正则化学习的快速收敛

论文作者：Vasilis Syrgkanis, Alekh Agarwal, Haipeng Luo, Robert Schapire

论文地址：http://120.52.51.17/papers.NeurIPS.cc/paper/5763-fast-convergence-of-regularized-learning-in-games.pdf

论文解读：我们证明了具有新近偏置形式的自然类正则化学习算法，可以在多人正常形式博弈中达到更快的收敛速度，从而有效地近似并达到粗略的相关均衡。当博弈中的每个玩家使用我们类中的算法时，它们会在 $O(T^{-3/4})$处衰减，而效用的总和会在 $O(T^{-1})$处收敛至最佳值——在最差的情况 $O(T^{-1/2})$比率下有所改善情况。我们展示了该类中任何算法的黑盒衰减，以针对对手达到 $\tilde{O}(T^{-1/2})$的速率，同时保持该类中算法的较快速率。我们的结果扩展了 Rakhlin 和 Shridharan 和 Daskalakis 等人的结果，它们只针对特定算法分析了两人零和博弈。

- 2014 年最佳论文

论文题目：*Asymmetric LSH (ALSH) for sublinear time Maximum Inner Product Search (MIPS)*

中文题目：次线性时间的不对称 LSH（ALSH）最大内积检索（MIPS）

论文作者：Anshumali Shrivastava, Ping Li

论文地址：https://arxiv.org/pdf/1405.5869.pdf

论文解读：我们提出了第一个可证明的次线性时间哈希算法，用于近似最大内积检索（MIPS）。使用（未归一化的）内积作为基础相似性度量进行检索是一个已知的难题，并且为 MIPS 查找哈希方案是很困难。虽然现有的本地敏感哈希（LSH）框架不足以解决 MIPS，但在本文中，我们将 LSH 框架扩展为允许非对称哈希方案。我们的方法基于一个关键的观察，即在独立的不对称变换之后，找

到最大内积的问题可以转化为经典设置中的近似近邻搜索问题。这个关键的发现使针对 MIPS 的高效亚线性哈希方案成为可能。我们提出的算法简单易实现。所提出的散列方案与协作过滤中的两种流行的常规 LSH 方案相比，显着节省了计算量：（i）符号随机投影（SRP）和（ii）基于 L-2 范数的 p 稳定分布（L2LSH），在 Netflix 和 Movielens（10M）数据集上的项目推荐任务。

论文题目：*A* Sampling*

中文题目：A*采样

论文作者：Chris J. Maddison, Daniel Tarlow, Tom Minka

论文地址：https://arxiv.org/pdf/1411.0030.pdf

论文解读：从离散分布中提取样本的问题可以转化为离散优化问题。在这项工作中，本文展示了如何将连续分布的采样转化为连续空间上的优化问题。该方法的核心是最近在数学统计学中描述的一个随机过程，本文称之为 Gumbel 过程。本文提出了一种新的 Gumbel 过程和 A*采样结构，这是一种实用的通用采样算法，它使用 A*搜索来搜索 Gumbel 过程的最大值。本文分析了 A*抽样的正确性和收敛时间，并从经验上证明了它比最相关的自适应拒绝抽样算法更有效地利用了边界和似然估计。

- 2013 年最佳论文

论文题目：*A Memory Frontier for Complex Synapses*

中文题目：复杂突触的记忆边界

论文作者：Subhaneil Lahiri, Surya Ganguli

论文地址：http://120.52.51.14/papers.NeurIPS.cc/paper/4872-a-memory-frontier-for-complex-synapses.pdf

论文解读：一个令人难以置信的鸿沟将突触的理论模型分开，通常仅由表示突触后电位大小的单个标量值描述，来自真实突触下的分子信号传导途径的巨大复杂性。为了理解这种分子复杂性对学习和记忆的功能贡献，必须将突触的理论概念从单个标量扩展到具有许多内部分子功能状态的整个动力系统。这里产生了

一个基本问题，突触复杂性如何产生记忆？为了解决这个问题，本文开发了新的数学定理，阐明了复杂突触的结构组织和记忆特性之间的关系，这些突触本身就是分子网络。此外，在证明这些定理时，本文发现了一个基于第一次通过时间理论的框架，对复杂突触模型的内部状态施加顺序，从而简化了突触结构和功能之间的关系。

论文题目：*Submodular Optimization with Submodular Cover and Submodular Knapsack Constraints*

中文题目：具有子模块覆盖和子模块背包约束的子模块优化

论文作者：Rishabh Iyer, Jeff Bilmes

论文地址：https://static.aminer.org/pdf/20160902/web-conf/NEURIPS/NEURIPS-2013-2874.pdf

论文解读：我们研究了两个新的优化问题——最小化受子模块化下界约束（子模块化覆盖）的子模函数和最大化子模块函数受下模块化上限约束（子模块化背包）的约束。我们受到机器学习中许多实际应用的启发，这些应用包括传感器放置和数据子集选择，这些应用要求最大化某个子模块功能（例如覆盖范围或分集），同时最小化另一个子模块功能（例如合作成本）。我们发现通过将这些问题表述为约束优化（对于许多应用程序而言更自然），可以实现许多有界逼近的保证。我们还表明，这两个问题都是密切相关的，可以使用求解一个问题的近似算法来获得对另一个问题的近似保证。我们提供了两个问题的结果，从而表明我们的逼近因子严格到对数因子。最后，我们通过实验证明了算法的性能和良好的可伸缩性。

论文题目：*Scalable Influence Estimation in Continuous-Time Diffusion Networks*

中文题目：连续时间扩散网络中的可扩展影响估计

论文作者：Nan Du, Le Song, Manuel Gomez-Rodriguez, Hongyuan Zha

论文地址：https://static.aminer.org/pdf/20160902/web- conf/NEURIPS/NEURIPS-2013-2951.pdf

论文解读：如果从媒体站点发布一条信息，我们能否预测它是否可以在一个月内传播到一百万个网页？由于需要同时处理任务的时间敏感性和可伸缩性要求，因此影响估计问题非常具有挑战性。在本文中，我们提出了一种用于连续时间扩散网络中影响估计的随机算法。我们的算法可以用|V|估计网络中每个节点的影响。节点和|ε|使用 n = $O(1/\varepsilon^2)$随机化并以对数因子 O(n|ε|+n|V|)计算边缘到 ε 的精度。当在贪婪影响最大化方法中用作子例程时，我们提出的算法可确保找到一个至少受（1-1/ε|）OPT-2$C\varepsilon$|影响的 C 节点集，其中 OPT 是最佳值。对合成数据和实际数据进行的实验均表明，该算法可以轻松扩展至数百万个节点的网络，同时在估计影响力的准确性和质量方面都大大优于以前的最新技术，选择节点以最大程度地发挥影响力。

- 2012 年最佳论文

论文题目：*No voodoo here! Learning discrete graphical models via inverse covariance estimation*

中文题目：通过逆协方差估计学习离散图模型

论文作者：Po-Ling Loh, Martin Wainwright

论文地址：https://pdfs.semanticscholar.org/85fb/8ff18d1a588b7b314faaa6a17f6b68818683.pdf

论文解读：本文研究了广义协方差矩阵的逆的支持与离散图形模型的结构之间的关系。本文证明了对于某些图结构，指标变量的逆协方差矩阵对图的顶点的支持反映了图的条件独立结构。本文的工作扩展了以前仅针对多元高斯分布建立的结果，并且部分地回答了关于非高斯分布的逆协方差矩阵含义的开放问题。本文提出了基于可能损坏的观测值的具有有界度的一般离散图形模型的图选择方法，并通过模拟验证本文的理论结果。在此过程中，本文还在基于损坏和缺失观测的稀疏高维线性回归设置中建立支持恢复的新结果。

论文题目：*Discriminative Learning of Sum-Product Networks*

中文题目：和积网络的判别学习

论文作者：Robert Gens, Pedro Domingos

论文地址：http://papers.nips.cc/paper/4516-discriminative-learning-of-sum-product-networks.pdf

论文解读：Sum-product 网络是一种新的深度架构，可以对高树宽模型进行快速、准确的推断。迄今为止，仅提出了用于生成 SPN 的生成方法。在本文中，我们提出了第一种针对 SPN 的判别式训练算法，将前者的高精度与后者的表示能力和易处理性相结合。我们表明，可分辨的判别式 SPN 的类别比可处理的可区分性 SPN 的类别更广泛，并提出了一种有效的反向传播算法来计算条件对数似然度的梯度。我们在标准图像分类任务上测试判别式 SPN。我们使用迄今在 CIFAR-10 数据集上获得最佳结果的方法，其性能比具有 SPN 架构的方法（具有判别性学习本地图像结构）的性能要少。即使仅使用数据集的标记部分，我们也报告了 STL-10 上公布的最高测试准确性。

- 2011 年最佳论文

论文题目：*Efficient Inference in Fully Connected CRFs with Gaussian Edge Potentials*

中文题目：具有高斯边缘电位的完全连接 CRF 中的有效推理

论文作者：Philipp Krähenbühl, Vladlen Koltun

论文地址：https://static.aminer.org/pdf/20160902/web- conf/NEURIPS/NEURIPS-2011-1998.pdf

论文解读：用于多类图像分割和标记的大多数最新技术使用在像素或图像区域上定义的条件随机字段。尽管区域级模型通常具有密集的成对连通性，但像素级模型却要大得多，并且只允许稀疏图结构。在本文中，我们考虑在图像的完整像素集上定义的完全连接的 CRF 模型。生成的图具有数十亿条边，这使得传统的推理算法不切实际。我们的主要贡献是针对全连接 CRF 模型的高效近似推理算法，其中成对边缘势能由高斯核的线性组合定义。我们的实验表明，像素级的密集连接性可显着改善分割和标记的准确性。

论文题目：*Priors Over Recurrent Continuous Time Processes*

中文题目：连续时间过程的优先级

论文作者：Ardavan Saeedi, Alexandre Bouchard-Côte

论文地址：https://static.aminer.org/pdf/20160902/web- conf/NEURIPS/NEURIPS-2011-2195.pdf

论文解读：本文引入 Gamma 指数过程（GEP），这是一个大型连续时间过程系列的先验。该先验的分层版本（HGEP; Hierarchical GEP）产生用于分析复杂时间序列的有用模型。基于 HGEP 的模型显示出许多有吸引力的特性：等待时间的共轭性，可交换性和封闭形式预测分布，以及时间尺度参数的精确 Gibbs 更新。在建立这些属性之后，本文展示了如何使用粒子 MCMC 方法有效地进行后验推理。本文将本文的模型应用于估计多发性硬化症的疾病进展和 RNA 进化建模的问题。在这两个领域，本文发现本文的模型优于标准的速率矩阵估计方法。

论文题目：*Fast and Accurate K-means for Large Datasets*

中文题目：大型数据集的快速准确 K-means

论文作者：Michael Shindler, Alex Wong, Adam Meyerson

论文地址：http://120.52.51.15/papers.NeurIPS.cc/paper/4362-fast-and-accurate-k-means-for-large-datasets.pdf

论文解读：群集是许多应用程序中的一个普遍问题。在数据太大而无法存储在主内存中、并且必须顺序访问（例如从磁盘）、必须使用尽可能少的内存的情况下，我们考虑 k 均值问题。我们的算法基于最新的理论结果，并进行了重大改进以使其实用。然后，我们合并近似最近邻搜索以计算 $o(nk)$中的 k 均值（其中 n 是数据点的数量；请注意，在给定解的情况下，计算成本需要 $\Theta(nk)$时间）。我们证明了我们的算法在理论上和实验上都优于现有算法，从而在理论和实践上均提供了最先进的性能。

- 2010 年最佳论文

论文题目：*Construction of dependent dirichlet Processes based on Poisson Processes*

中文题目：基于泊松过程的 DDP 构建

论文作者：Dahua Lin, Eric Grimson, John Fisher

论文地址：https://static.aminer.org/pdf/20160902/web- conf/NEURIPS/NEURIPS-2010-3901.pdf

论文解读：本文提出了一种构造依赖 Dirichlet 过程的方法。新的方法揭示了 Dirichlet 和泊松过程之间的内在关系，以便创建一个适合用作先前演化混合模型的 Dirichlet 过程的马尔可夫链。该方法允许组件模型随时间的创建、移除和位置变化，同时保持随机测量略微 DP 分布的属性。此外，本文推导出用于模型推理的 Gibbs 采样算法，并在合成和实际数据上进行测试。实证结果表明该方法可有效地估算动态变化的混合模型。

论文题目：*A Theory of Multiclass Boosting*

中文题目：多类别 Boosting 算法的理论

论文作者：Indraneel Mukherje, Robert E Schapire

论文地址：https://static.aminer.org/pdf/20160902/web- conf/NEURIPS/NEURIPS-2010-3934.pdf

论文解读：Boosting 将弱分类器组合在一起，以形成高度准确的预测器。尽管二进制分类的情况已广为人知，但在多类设置中，缺少对弱分类器的“正确”要求或最有效的增强算法的概念。在本文中，我们创建了一个广泛而通用的框架，在此框架内，我们可以对弱分类器进行精确确定并确定最佳要求，并在某种意义上设计最有效的 Boosting 算法来满足此类要求。

- 2009 年最佳论文

论文题目：*An LP View of the M-Best MAP Problem*

论文作者：Menachem Fromer, Amir Globerson

论文地址：https://static.aminer.org/pdf/20160902/web- conf/NEURIPS/NEURIPS-2009-4089.pdf

论文解读：本文考虑在概率图形模型中以最大概率找到 M 指派的问题。本文展示了如何将这个问题表述为特定多面体上的线性程序（LP）。本文证明，对于树形图（和一般的交叉树），这个多面体具有特别简单的形式，并且与单个不

等式约束中的边际多面体不同。本文使用这种表征来为非树图提供近似方案，通过使用这些图上的生成树集。本文提出的方法在 LP 松弛的背景下提出了 M-最佳推理问题，LP 松弛最近得到了相当多的关注，并且已经证明在解决困难的推理问题方面是有用的。本文凭经验证明，本文的方法经常为高树宽度的问题找到可证明的精确 M 最佳配置。

论文题目：*Fast Subtree Kernels on Graphs*

中文题目：图的快速子树核

论文作者：Nino Shervashidze, Karsten Borgwardt

论文地址：https://static.aminer.org/pdf/20160902/web-conf/ NEURIPS/NEURIPS-2009-4199.pdf

论文解读：在本文中，我们提出了图的快速子树核。在具有 n 个节点和 m 个边，且最大度为 d 的图上，这些高度为 h 的比较子树核可以用 $O(mh)$计算，而 Ramon＆Gartner 经典子树核的缩放比例为 $O(n^2 4^d h)$。效率性的关键是观察到，根据图论进行的 Weisfeiler-Lehman 同构检验很好地计算了作为副产品的子树核。我们的快速子树核可以处理带标签的图形，可以轻松扩展到大型图形，并且可以在准确性和运行时间方面在多个分类基准数据集上胜过最新的图形核。

3 计算机视觉

3.1 计算机视觉概念

计算机视觉（computer vision），顾名思义，是分析、研究让计算机智能化的达到类似人类的双眼“看”的一门研究科学[3]。即对于客观存在的三维立体化的世界的理解以及识别依靠智能化的计算机去实现。确切地说，计算机视觉技术就是利用了摄像机以及电脑替代人眼使得计算机拥有人类的双眼所具有的分割、分类、识别、跟踪、判别决策等功能。总之，计算机视觉系统就是创建了能够在2D 的平面图像或者 3D 的三维立体图像的数据中，以获取所需要的“信息”的一个完整的人工智能系统。

计算机视觉技术是一门包括了计算机科学与工程、神经生理学、物理学、信号处理、认知科学、应用数学与统计等多门科学学科的综合性科学技术。由于计算机视觉技术系统在基于高性能的计算机的基础上，其能够快速的获取大量的数据信息并且基于智能算法能够快速的进行处理信息，也易于同设计信息和加工控制信息集成。

计算机视觉本身包括了诸多不同的研究方向，比较基础和热门的方向包括：物体识别和检测（Object Detection），语义分割（Semantic Segmentation），运动和跟踪（Motion & Tracking），视觉问答（Visual Question & Answering）等[4]。

- 物体识别和检测

物体检测一直是计算机视觉中非常基础且重要的一个研究方向，大多数新的算法或深度学习网络结构都首先在物体检测中得以应用如 VGG-net，GoogLeNet，ResNet 等等，每年在 imagenet 数据集上面都不断有新的算法涌现，一次次突破历史，创下新的记录，而这些新的算法或网络结构很快就会成为这一年的热点，并被改进应用到计算机视觉中的其它应用中去。

物体识别和检测，顾名思义，即给定一张输入图片，算法能够自动找出图片中的常见物体，并将其所属类别及位置输出出来。当然也就衍生出了诸如人脸检测（Face Detection），车辆检测（Viechle Detection）等细分类的检测算法。

- 语义分割

语义分割是近年来非常热门的方向，简单来说，它其实可以看作一种特殊的分类——将输入图像的每一个像素点进行归类，用一张图就可以很清晰地描述出来。很清楚地就可以看出，物体检测和识别通常是将物体在原图像上框出，可以说是“宏观”上的物体，而语义分割是从每一个像素上进行分类，图像中的每一个像素都有属于自己的类别。

- 运动和跟踪

跟踪也属于计算机视觉领域内的基础问题之一，在近年来也得到了非常充足的发展，方法也由过去的非深度算法跨越向了深度学习算法，精度也越来越高，不过实时的深度学习跟踪算法精度一直难以提升，而精度非常高的跟踪算法的速度又十分之慢，因此在实际应用中也很难派上用场。

学术界对待跟踪的评判标准主要是在一段给定的视频中，在第一帧给出被跟踪物体的位置及尺度大小，在后续的视频当中，跟踪算法需要从视频中去寻找到被跟踪物体的位置，并适应各类光照变换，运动模糊以及表观的变化等。但实际上跟踪是一个不适定问题（ill posed problem），比如跟踪一辆车，如果从车的尾部开始跟踪，若是车辆在行进过程中表观发生了非常大的变化，如旋转了 180 度变成了侧面，那么现有的跟踪算法很大的可能性是跟踪不到的，因为它们的模型大多基于第一帧的学习，虽然在随后的跟踪过程中也会更新，但受限于训练样本过少，所以难以得到一个良好的跟踪模型，在被跟踪物体的表观发生巨大变化时，就难以适应了。所以，就目前而言，跟踪算不上是计算机视觉内特别热门的一个研究方向，很多算法都改进自检测或识别算法。

- 视觉问答

视觉问答也简称 VQA（Visual Question Answering），是近年来非常热门的一个方向，其研究目的旨在根据输入图像，由用户进行提问，而算法自动根据提问内容进行回答。除了问答以外，还有一种算法被称为标题生成算法（Caption Generation），即计算机根据图像自动生成一段描述该图像的文本，而不进行问答。对于这类跨越两种数据形态（如文本和图像）的算法，有时候也可以称之为多模态，或跨模态问题。

3.2 计算机视觉发展历史

尽管人们对计算机视觉这门学科的起始时间和发展历史有不同的看法，但应该说，1982 年马尔（David Marr）《视觉》（Marr，1982）一书的问世，标志着计算机视觉成为了一门独立学科。计算机视觉的研究内容，大体可以分为物体视觉（object vision）和空间视觉（spatial vision）二大部分。物体视觉在于对物体进行精细分类和鉴别，而空间视觉在于确定物体的位置和形状，为“动作（action）”服务。正像著名的认知心理学家 J.J.Gibson 所言，视觉的主要功能在于“适应外界环境，控制自身运动”。适应外界环境和控制自身运动，是生物生存的需求，这些功能的实现需要靠物体视觉和空间视觉协调完成。

计算机视觉 40 多年的发展中，尽管人们提出了大量的理论和方法，但总体上说，计算机视觉经历了三个主要历程。即：马尔计算视觉、多视几何与分层三维重建和基于学习的视觉。下面将对这三项主要内容进行简要介绍[5]。

- 马尔计算视觉（Computational Vision）

现在很多计算机视觉的研究人员，恐怕对“马尔计算视觉”根本不了解，这不能不说是一件非常遗憾的事。目前，在计算机上调“深度网络”来提高物体识别的精度似乎就等于从事“视觉研究”。事实上，马尔的计算视觉的提出，不论在理论上还是研究视觉的方法论上，均具有划时代的意义。

马尔的计算视觉分为三个层次：计算理论、表达和算法以及算法实现。由于马尔认为算法实现并不影响算法的功能和效果，所以，马尔计算视觉理论主要讨论“计算理论”和“表达与算法”二部分内容。马尔认为，大脑的神经计算和计算机的数值计算没有本质区别，所以马尔没有对“算法实现”进行任何探讨。从现在神经科学的进展看，“神经计算”与数值计算在有些情况下会产生本质区别，如目前兴起的神经形态计算（Neuromorphological computing），但总体上说，“数值计算”可以“模拟神经计算”。至少从现在看，“算法的不同实现途径”，并不影响马尔计算视觉理论的本质属性。

- 多视几何与分层三维重建

上世纪 90 年代初计算机视觉从“萧条”走向进一步“繁荣”，主要得益于以下二方面的因素：首先，瞄准的应用领域从精度和鲁棒性要求太高的“工业应用”转到要求不太高，特别是仅仅需要“视觉效果”的应用领域，如远程视频会议（teleconference），考古，虚拟现实，视频监控等。另一方面，人们发现，多视几何理论下的分层三维重建能有效提高三维重建的鲁棒性和精度。

多视几何的代表性人物首数法国 INRIA 的 O.Faugeras，美国 GE 研究院的 R.Hartely 和英国牛津大学的 A.Zisserman。应该说，多视几何的理论于 2000 年已基本完善。2000 年 Hartley 和 Zisserman 合著的书（Hartley & Zisserman2000）对这方面的内容给出了比较系统的总结，而后这方面的工作主要集中在如何提高“大数据下鲁棒性重建的计算效率”。大数据需要全自动重建，而全自动重建需要反复优化，而反复优化需要花费大量计算资源。所以，如何在保证鲁棒性的前提下快速进行大场景的三维重建是后期研究的重点。举一个简单例子，假如要三维重建北京中关村地区，为了保证重建的完整性，需要获取大量的地面和无人机图像。假如获取了 1 万幅地面高分辨率图像（4000×3000），5 千幅高分辨率无人机图像（8000×7000）（这样的图像规模是当前的典型规模），三维重建要匹配这些图像，从中选取合适的图像集，然后对相机位置信息进行标定并重建出场景的三维结构，如此大的数据量，人工干预是不可能的，所以整个三维重建流程必须全自动进行。这样需要重建算法和系统具有非常高的鲁棒性，否则根本无法全自动三维重建。在鲁棒性保证的情况下，三维重建效率也是一个巨大的挑战。所以，目前在这方面的研究重点是如何快速、鲁棒地重建大场景。

- 基于学习的视觉

基于学习的视觉，是指以机器学习为主要技术手段的计算机视觉研究。基于学习的视觉研究，文献中大体上分为二个阶段：本世纪初的以流形学习为代表的子空间法和目前以深度学习为代表的视觉方法。

物体表达是物体识别的核心问题，给定图像物体，如人脸图像，不同的表达，物体的分类和识别率不同。另外，直接将图像像素作为表达是一种“过表达”，也不是一种好的表达。流形学习理论认为，一种图像物体存在其“内在流形”

（intrinsic manifold），这种内在流形是该物体的一种优质表达。所以，流形学习就是从图像表达学习其内在流形表达的过程，这种内在流形的学习过程一般是一种非线性优化过程。深度学习的成功，主要得益于数据积累和计算能力的提高。深度网络的概念上世纪 80 年代就已提出来了，只是因为当时发现“深度网络”性能还不如“浅层网络”，所以没有得到大的发展。目前似乎有点计算机视觉就是深度学习的应用之势，这可以从计算机视觉的三大国际会议：国际计算机视觉会议（ICCV），欧洲计算机视觉会议（ECCV）和计算机视觉和模式识别会议（CVPR）上近年来发表的论文可见一般。目前的基本状况是，人们都在利用深度学习来“取代”计算机视觉中的传统方法。“研究人员”成了“调程序的机器”，这实在是一种不正常的“群众式运动”。牛顿的万有引力定律，麦克斯韦的电磁方程，爱因斯坦的质能方程，量子力学中的薛定谔方程，似乎还是人们应该追求的目标。

3.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为计算机视觉领域全球学者分布情况：



图 3-1 计算机视觉领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布在其东西海岸；亚

洲也有较多的人才分布，主要集中在我国东部及日韩地区；欧洲的人才主要分布在欧洲中西部；其他诸如非洲、南美洲等地区的学者非常稀少；计算机视觉领域的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，计算机视觉中男性学者占比 91.0%，女性学者占比 9.0%，男性学者占比远高于女性学者。

计算机视觉学者的 h-index 分布如下图所示，大部分学者的 h-index 分布在中间区域，其中 h-index 在 20-30 区间的人数最多，有 706 人，占比 34.7%，小于 20 的区间人数最少，有 81 人。

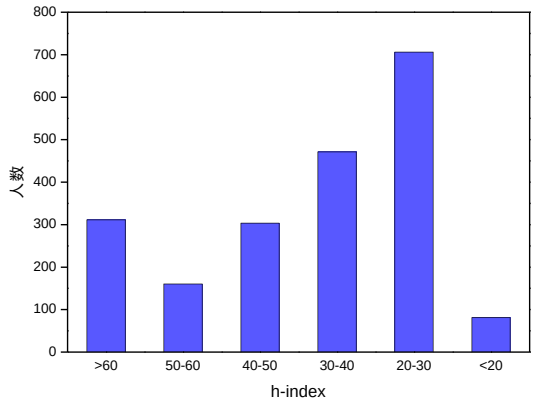


图 3-2 计算机视觉领域学者 h-index 分布

- 中国人才分布



图 3-3 计算机视觉领域中国学者分布

我国专家学者在计算机视觉领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是珠三角和长三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等亚洲国家相比，中国在计算机视觉领域学者数量相对较多。

中国与其他国家在计算机视觉的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 3-1 计算机视觉领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	1034	88585	86	1459
中国-新加坡	210	20194	96	283
中国-澳大利亚	110	6815	62	147
中国-英国	101	7769	77	148
中国-加拿大	70	7070	101	109
中国-日本	36	2093	58	69
中国-巴基斯坦	26	1933	74	35
中国-瑞士	25	2071	83	46
中国-德国	23	655	28	42
中国-韩国	22	1325	60	51

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在计算机视觉领域合作之密切；同时，中国与世界各地区之间的合作非常广泛，前 10 名合作关系里包含了欧洲、亚洲、北美洲以及大洋洲等；中国与加拿大合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中加合作达到了较高的水平。

3.4 论文解读

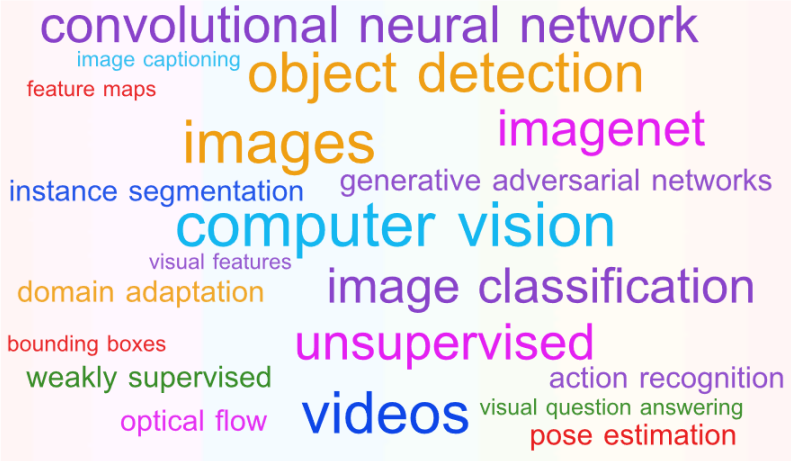
本节对本领域的高水平学术会议论文进行挖掘，解读这些会议在 2018-2019 年的部分代表性工作。会议具体包括：

IEEE Conference on Computer Vision and Pattern Recognition

IEEE International Conference on Computer Vision

European Conference on Computer Vision

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图。其中，计算机视觉（computer vision）、图像（images）、视频（videos）是本领域中最热的关键词。



论文题目：*Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation*

中文题目：具有空洞分离卷积的编码-解码器用于语义图像分割

论文作者：Liang-Chieh Chen，Yukun Zhu，George Papandreou，Florian Schroff，Hartwig Adam

论文出处：Proceedings of the European conference on computer vision (ECCV). 2018: 801-818.

论文地址：https://link.springer.com/chapter/10.1007%2F978-3-030-01234-2_49

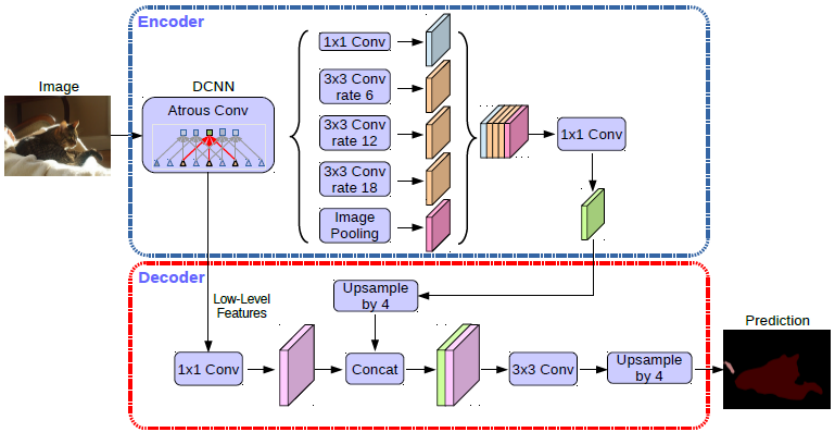
研究问题：

语义分割是计算机视觉中一项基本且重要的研究内容，它是为图像中的每个像素分配语义标签。在深度学习语义分割任务中经常会使用空间金字塔池化和编码-解码器结构。空间金字塔池化可以通过不同分辨率的池化特征捕捉丰富的上下文信息，但网络中具有步进操作的池化或卷积会导致与对象边界有关的详细信息丢失。这可以通过空洞卷积提取更密集的特征图来缓解，但大大增加了计算资

源的消耗。而编码-解码器结构则可以通过逐渐恢复空间信息来捕获更清晰的对象边界。通过组合两种方法的优点,提出新的模型—DeepLabv3+。

研究方法：

如下图是DeepLabv3+的网络结构,通过添加一个简单但有效的解码器模块来优化分割结果,尤其是对象边界的分割结果,扩展了 DeepLabv3。编码器模块（DeepLabv3）通过在多个尺度上应用空洞卷积，编码多尺度上下文信息。空洞卷积可以明确控制由深度卷积神经网络所提特征的分辨率，并调整滤波器的感受野以捕获多尺度信息。而简单而有效的解码器模块则沿对象边界调整分割结果。为了进一步提高模型的性能和速度，将深度分离卷积应用于 ASPP（空洞空间金字塔池化）和解码器模块。深度分离卷积将传统的卷积分解为一个深度卷积和一个 1×1 的逐点卷积，在深度卷积操作时应用膨胀率不同的空洞卷积，以获取不同的尺度信息。



研究结果：

以用 ImageNet-1k 预训练的 ResNet-101 和修改的对齐 Xception（更多的层、步进深度分离卷积替代最大池化、额外的 BN 和 ReLU）为骨架网络，通过空洞卷积提取稠密特征。在 PASCAL VOC 2012 和 Cityscapes 数据集上证明了DeepLabv3+的有效性和先进性，无需任何后处理即可实现 89%和 82.1%的测试集性能。但是对非常相近的物体（例如椅子和沙发）、严重遮挡的物体和视野极小的物体较难进行分割。

论文题目：*MobileNetV2: Inverted Residuals and Linear Bottlenecks*

中文题目：MobileNetV2: 反向残差和线性瓶颈

论文作者：Sandler Mark，Howard Andrew，Zhu Menglong，Zhmoginov Andrey，Chen Liang-Chieh

论文出处：2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018

论文地址：https://ieeexplore.ieee.org/document/8578572

研究问题：

在众多计算机视觉领域中，深度神经网络正扮演越来越重要的角色。但是优秀性能的获得通常是以高昂计算资源为代价的，从而大大限制了在计算资源严重受限的移动端或嵌入式设备中使用。因此轻量化网络的研究在近期收到了大量关注，本文提出了一种新的移动端轻量化模型—MobileNetV2，在保持相同精度的同时显着减少了所需的操作和内存需求，关键是设计了具有线性瓶颈的反向残差模块。将上述模型应用于移动端目标检测，介绍了一种有效的方法—SSDLite。此外，通过简化的 DeepLabv3 构建移动端语义分割模型—Mobile DeepLabv3。

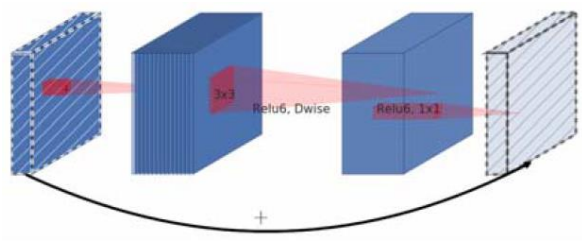
研究方法：

MobileNetV2 的关键是具有线性瓶颈的反向残差模块，该模块以低维压缩表示作为输入，首先将其扩张到高维，然后使用轻量级的深度卷积进行过滤，最后使用线性卷积将特征投影回低维表示。其包含两个主要的技术：深度分离卷积和残差模块。

深度分离卷积是很多有效的神经网络结构中关键的组成部分，其基本思想是将传统卷积分解为两部分：第一层称为深度卷积，它通过对每个输入通道应用单个卷积滤波器来执行轻量化滤波；第二层是 1×1 卷积，称为逐点卷积，它通过计算输入通道的线性组合来构建新特征。深度分离卷积的计算量相对于传统卷积减少了大约 k^2（k 是卷积核大小），但是性能只有极小的降低。

我们可以认为深度神经网络中任意层的激活组成一个“感兴趣流形”，它可以嵌入到低维子空间中。也就是说，深度卷积层中所有单个通道的像素，其中编

码的信息实际上位于某种流形中，而该流形可以嵌入到低维子空间中。通过分析作者得到两个属性：（1）如果感兴趣流形在 ReLU 变换后仍保持非零值，则对应于线性变换；（2）ReLU 能够保留输入流形的完整信息，但前提是输入流形位于输入空间的一个低维子空间中。基于以上两个观点，帮助作者优化现有的神经网络结构：假设感兴趣流形是低维的，可以通过向卷积块插入线性瓶颈获得该流形，即本文核心具有线性瓶颈的反向残差模块，其结构如下图所示。先使用逐点卷积扩大通道数+ReLU 激活，然后使用逐深度卷积提取特征+ReLU 激活，最后使用逐点卷积降低通道数+线性激活，并且使用了 shortcut 连接。



研究结果：

研究者首先通过实验验证了反向残差连接和线性瓶颈的有效性，然后在图像分类、目标检测和语义分割三个任务上证明了本文网络结构的先进性。ImageNet 图像分类任务上 MobileNetV2 的 Top1 最好可达 74.7，优于 MobileNetV1、ShuffleNet 和 NASNet-A。在目标检测任务上，MNetV2+SSDLite 与 MNetV1+SSDLite 的 mAP 很接近，但参数量和计算时间都明显减少。在语义分割任务上保持较好性能的同时减少了参数量和计算资源的消耗。

论文题目：*The Unreasonable Effectiveness of Deep Features as a Perceptual Metric*

中文题目：深度特征作为感知度量的有效性

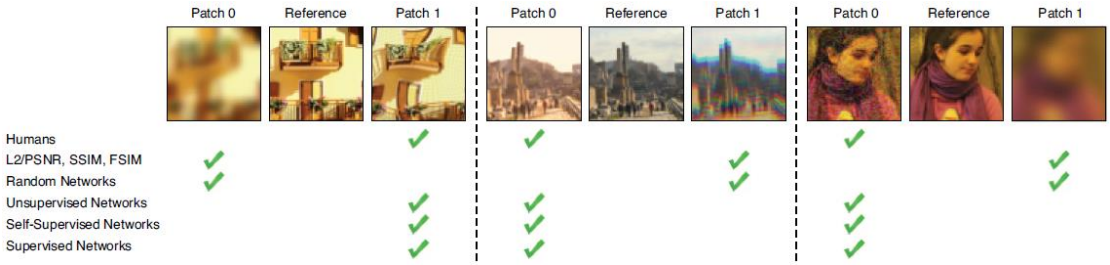
论文作者：Zhang Richard，Isola Phillip，Efros Alexei A.，Shechtman Eli，Wang Oliver

论文出处：2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018

论文地址：https://ieeexplore.ieee.org/document/8578166

研究方法：

对于人类来说，评估两幅图像之间的感知相似度几乎是毫不费力且快速的，但其潜在过程却被认为是相当复杂的。视觉模式是高维且高度相关的，视觉相似性的概念十分主观。例如在图像压缩领域，压缩图像是为了人类看来与原始图像没有很大区别，而不关注它们在像素值上可能有很大的差别。当今最广泛使用的、传统的基于像素值的度量方法（例如 L2 欧式距离、PSNR）或感知距离度量（如 SSIM、MSSIM 等）是简单的浅层函数，无法解决人类感知的许多细微差别，一个最典型的例子就是模糊会造成图像在感知上的很大不同，但是在 L2 范数上却差别不大。如下图所示，传统的评价指标与人类的感知判断是完全相反的。近期深度学习社区发现，将在 ImageNet 分类中训练的 VGG 网络模型所提取的深度特征，用作图像合成的训练损失是非常有用，一般将这种损失称为“感知损失”（perceptual losses）。但是这些感知损失的作用有多大？哪些要素对其成功至关重要？本文研究者们尝试探讨了这些问题。



研究方法：

为了研究将深度神经网络提取的深度特征作为感知损失的有效性，本文研究者们构造了一个人类感知相似性判断的新数据集——Berkeley-Adobe Perceptual Patch Similarity Dataset（BAPPS 数据集）。该数据集包括 484K 个人类判断，具有大量传统失真，如对比度、饱和度和噪声等；还有基于 CNN 模型的失真，例如自编码、降噪等造成的失真；以及一些真实算法的失真，如超分辨率重建、去模糊等真实应用。

论文用如下公式计算在给到一个网络 F 时，参考和失真图像块的距离。首先提取特征，然后将通道维度的激活归一化，用向量 w 缩放每个通道，并采用 2 距离。最后对空间维度的所有层取平均。

$$d(x,x_0)=\sum_l \frac{1}{H_l W_l}\sum_{h,w}\left\| w_l \square \left(\hat{y}^l_{hw}-\hat{y}^l_{0hw}\right)\right\|_2^2$$

研究结果：

作者进行了大量的实验，系统地评估了不同网络结构和任务中的深度特征，并将它们与经典指标进行比较，发现深度特征是一种非常好的感知度量指标。更令人惊讶的是，该结果不仅限于 ImageNet 训练的 VGG 提取的深度特征，而且还适用于不同的深度网络结构和不同的训练方式（监督，自监督，甚至无监督）。

论文题目：*Residual Dense Network for Image Super-Resolution*

中文题目：基于残差密集网络的图像超分辨率重建

论文作者：Yulun Zhang，Yapeng Tian，Yu Kong，Bineng Zhong，Yun Fu

论文出处：2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2018

论文地址：https://ieeexplore.ieee.org/document/8578360

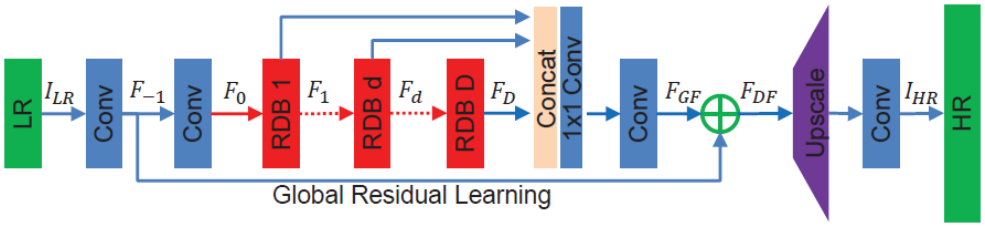
研究内容：

单幅图像超分辨率（SISR）旨在通过其退化的低分辨率（LR）观测结果生成视觉上令人愉悦的高分辨率（HR）图像。最近，深度卷积神经网络在图像超分辨率重建方面取得了巨大的成功，网络的不断加深使模型能提供丰富分层特征，图像中的目标具有不同的比例、视角和宽高比，来自非常深的网络的分层特征能为重建提供更多线索。但是，大多数基于卷积神经网络的深层图像超分辨率模型都没有充分利用原始低分辨率（LR）图像中的分层特征，以致获得了相对较低的性能。在本文中，研究者提出了一种新颖的残差密集网络（RDN）来解决图像超分辨率中的上述问题，使模型能充分利用所有卷积层提取的分层特征。

研究方法：

如下图是残差密集网络 RDN，主要包含四部分：浅层特征提取网络（SFEnet）、残差密集块（RDBs）、密集特征融合（DFF）和上采样网络（UPNet）。一个非常深的网络直接提取 LR 空间中每个卷积层的输出是困难且不切实际的，所以使用残差密集块（RDB）作为 RDN 的构建模块。RDB 由密集连接层和具有局部残

差学习能力的局部特征融合（LFF）组成。RDB 还支持 RDB 之间的连续存储，一个 RDB 的输出可以直接访问下一个 RDB 中每一层，形成连续的状态传递。RDB 中的每个卷积层都可以访问所有后续层，并传递需要保留的信息。局部特征融合将先前的 RDB 和当前 RDB 中所有先前层的状态连接在一起，通过自适应保留信息来提取局部密集特征。LFF 通过更高的增长率来稳定更宽网络的训练。在提取多层局部密集特征后，进一步进行全局特征融合（GFF），以全局方式自适应地保留分层特征。在 RDN 中每个卷积层卷积核大小为 3×3，局部和全局特征融合卷积核大小为 1×1。在上采样部分使用 ESPCNN 提升图像的分辨率。



研究结果：

使用 DIV2K 数据集中全部的 800 幅训练图像训练模型，测试选用 5 个标准基准数据集：Set5、Set14、B100、Urban 和 Manga109。为了全面地说明所提方法的有效性，模拟了三种图像退化过程：（1）双三次下采样（BI）；（2）高斯核模糊 HR 图像，再下采样（BD）；（3）先双三次下采样，再加入高斯噪声（DN）。作者进行了大量的实验发现：（1）RDB 数量或 RDB 中卷积层数量越多，模型性能越好；增长率越大也会获得更好的性能。当上述模块使用数量较少时 RDN 依然比 SRCNN 性能好。（2）进行了消融实验，验证了所提模型中连续存储、局部残差学习和全局特征融合的有效性。（3）在三种退化模型上与六种先进的模型进行了对比：SRCNN、LapSRN、DRNN、SRDenseNet、MemNet 和 MDSR。在不同比例因子、退化模型和数据集中，RDN 都表现出了相近甚至更好的性能。

论文题目：*ShuffleNet V2: Practical guidelines for efficient cnn architecture design*

中文题目：ShuffleNet V2: 高效 CNN 网络结构设计实用指南

论文作者：Ma Ningning，Zhang Xiangyu，Zheng Hai-Tao，Sun Jian

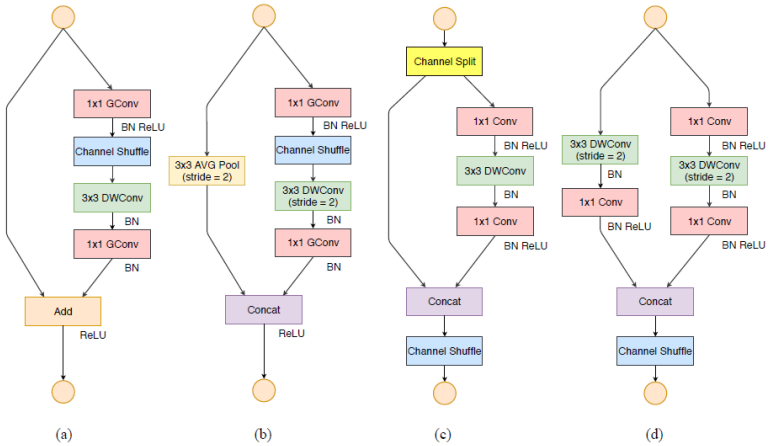
论文出处：Lecture Notes in Computer Science （including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics）, v 11218 LNCS, p 122-138, 2018, Computer Vision – ECCV 2018 - 15th European Conference, 2018, Proceedings

论文链接：https://link.springer.com/chapter/10.1007%2F978-3-030-01264-9_8

研究内容：

自 AlexNet 之后，ImageNet 图像分类准确率被很多新的网络结构如 ResNet 和 DenseNet 等不断提高，但是除准确率外，计算复杂度也是 CNN 网络需要考虑的重要指标。实际任务通常是要在有限的计算资源下获得最佳的精度，过复杂的网络由于速度原因难以在移动端等设备中应用。为此，研究者们提出了很多轻量化的 CNN 网络如 MobileNet 和 ShuffleNet 等，在速度和准确度之间做了较好地平衡。以往的移动端 CNN 网络结构设计在考虑计算复杂度时，直接致力于优化整体网络计算所需的 FLOPs，并没有考虑真正关心的速度和延迟，且具有相似 FLOPs 的网络速度也是不同的。像内存访问开销（MAC）、计算平台等也是需要考虑的方面。为了实际需求，本文研究者不局限于追求理论 FLOPs 的减少，从更直接的角度为轻量化网络设计提供指导意见。

研究方法：



作者建议有效的网络结构设计应考虑两个原则。首先，应使用直接指标（例如速度）代替间接指标（例如 FLOP）。其次，应在目标平台上评估此类指标。通过对两个代表性最新网络的分析，作者得出了关于有效网络设计的四项准则：

（1）卷积层的输入和输出特征通道数相等时 MAC 最小；（2）过多的组卷积会增大 MAC；（3）网络碎片化会降低并行度；（4）元素级的操作（element-wise）会增加时间消耗。遵循以上准则提出了一个更有效的网络结构——ShuffleNet V2。下图是 ShuffleNet V1（图中 a 和 b）和 ShuffleNet V2（图中 c 和 d）组成模块的对比。对比（a）和（b），ShuffleNet V2 首先用 Channel Split 操作将输入按通道分成两部分，一部分直接向下传递，另外一部分则用于计算；然后弃用了 1x1 的组卷积，将通道混洗操作（Channel Shuffle）移到了最后，并将前面的 Add 操作用 Concat 代替。

研究结果：

论文进行了大量的实验，与 MobileNet V1/V2、ShuffleNet V1、DenseNet、Xception、IGCV3-D、NASNet-A 等模型在速度、精度、FLOPs 上进行了详细的对比。实验中不少结果都和前面几点发现吻合，ShuffleNet V2 在准确率和速度方面达到了很好的平衡。

论文题目：*A Theory of Fermat Paths for Non-Line-of-Sight Shape Reconstruction*

中文题目：非视距形状重建的费马路径理论

论文作者：Shumian Xin, Sotiris Nousias, Kiriakos N. Kutulakos, Aswin C. Sankaranarayanan,Srinivasa G. Narasimhan, and Ioannis Gkioulekas.

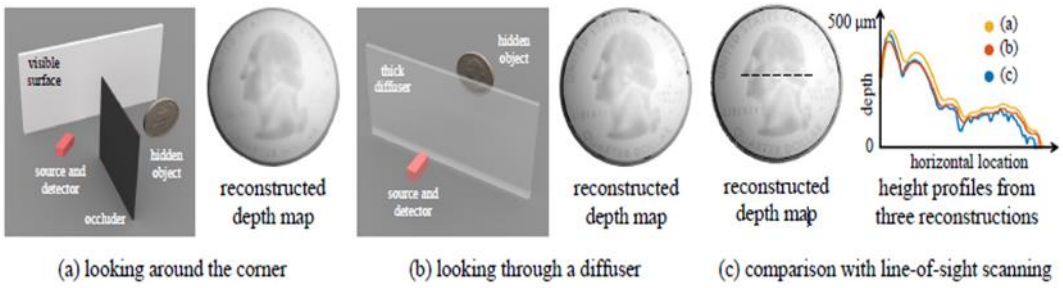
论文出处：CVPR 2019: IEEE Conference on Computer Vision and Pattern Recognition.

论文地址：https://www.ri.cmu.edu/wp-content/uploads/2019/05/cvpr2019.pdf

研究问题：

很多时候摄像头可能无法拍摄全部场景或物体，例如，面对摄像机的物体背面，拐角处的物体或通过漫射器观察到的物体。非视距（non-line-of-sight, NLOS）成像对于许多安全保护应用至关重要。一些传统方法通过分析隐藏场景投射阴影的细微本影和半影，以估计粗糙的运动和结构，或使用光的相干特性来定位隐藏的对象，但很难重建任意隐藏场景的 3D 形状。基于主动照明的瞬态 NLOS 成像大多采用快速调制光源和时间分辨传感器，但现有的 SPAD 强度估计不理想，而

且重建 NLOS 对象的朗伯反射率假设。作者使用 NLOS 瞬态测量得出几何约束而非强度约束的方法来克服上述限制。



(a) looking around the corner (b) looking through a diffuser (c) comparison with line-of-sight scanning

上图为非视距成像示例：被遮光板遮挡（a）和被漫射板遮挡（b）的物体表面重建结果与视距扫描结果（c）对比。

研究方法：

作者提出了一个新的光费马路径（Fermat path）理论，即光在已知的可见场景和不处于瞬态相机视线范围内的未知物体之间，这些光要么从镜面反射，要么被物体的边界反射，从而编码了隐藏物体的形状。作者证明，费马路径对应于瞬态测量中的不连续性，间断点的位置仅与 NLOS 对象的形状有关，与其反射率（BRDF）无关。并推导出一个新的约束条件，它将这些不连续处的路径长度的空间导数与曲面的曲率相关联。基于此理论，作者提出了一种称为费马流（Fermat Flow）的算法，用于估计非视距物体的形状。其关键在于，费马路径长度的空间导数可唯一确定隐藏场景点的深度和法线，再拟合和估算平滑路径长度函数，进一步结合深度和法线获得光滑的网格，从而精确恢复了对复杂对象（从漫反射到镜面反射）形状，范围从隐藏在拐角处以及隐藏在漫射器后面的漫反射到镜面反射。最后，该方法与用于瞬态成像的特定技术无关。

研究结果：

作者使用了一些不同 BRDF 的凹凸几何形状的日常物品，包括半透明（塑料壶），光滑（碗，花瓶），粗糙镜面（水壶）和光滑镜面（球形）等。分别开展了使用 SPAD 和超快激光从皮秒级瞬态中恢复毫米级形状，以及使用干涉法实现从飞秒级瞬态中恢复毫米级形状的两种实验，实验结果显示重建细节与 groundtruth 形状非常吻合。

论文题目：*Implicit 3D Orientation Learning for 6D Object Detection from RGB Images*

中文题目：从 RGB 图像检测 6 维位姿的隐式三维朝向学习

论文作者：Martin Sundermeyer , Zoltan-Csaba Marton , Maximilian Durner , Rudolph Triebel

论文出处：ECCV 2018: European Conference on Computer Vision.

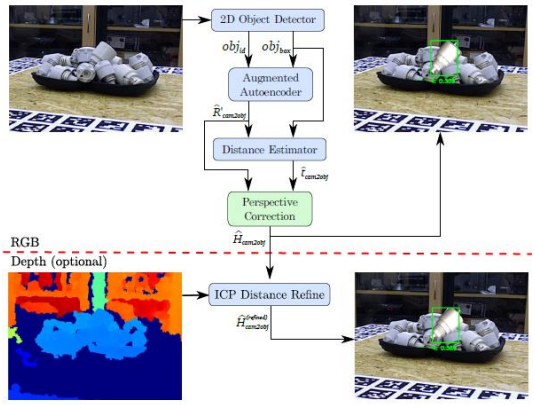
论文地址：

http://openaccess.thecvf.com/content_ECCV_2018/papers/Martin_Sundermeyer_Implicit_3D_Orientation_ECCV_2018_paper.pdf

研究问题：

对于诸如移动机器人控制和增强现实之类的应用而言，现代计算机视觉系统中最重要的组件之一就是可靠且快速的 6D 目标检测模块。至今尚无通用，易于应用，强大且快速的解决方案。原因是多方面的:首先，当前的解决方案通常不足以有效处理典型的挑战；其次，现有方法通常需要某些目标属性。而且，当前的方法在运行时间以及所需带标注的训练数据的数量和种类方面效率都不高。作者提出对单个 RGB 图像进行操作，可在很大程度上不需要深度信息，显著增加可用性。

研究方法：



上图为 6D 目标检测管道具有齐次坐标变换$H_{cam2obj}$（右上）和深度细化结果$H_{cam2obj}^{(refined)}$（右下）。作者提出了一种基于 RGB 的实时目标检测和 6D 姿态估

计流程。首先使用 SSD（Single Shot Multibox Detector）来提供目标边界框和标识符。其次，在此基础上，采用新颖的 3D 方向估计算法，该算法基于之前的降噪自动编码器（Denoising Autoencoder）的通用版本，增强型自动编码器（AAE）。AAE 使用一种新颖的域随机化策略，模型学到的并不是从输入图像到物体位姿的显式映射，而是会根据图像样本在隐含空间内建立一个隐式的物体位姿表征。因而，训练独立于目标方向的具体表示（例如四元数），避免从图像到方向的一对多映射，由此 AAE 可处理由对称视图引起的模糊姿态。另外学习专门编码 3D 方向的表征，同时实现对遮挡，杂乱背景的鲁棒性，并可推广到对不同环境和测试传感器。而且，AAE 不需要任何真实的姿势标注训练数据。相反，它被训练为以自我监督的方式编码 3D 模型视图，克服了对大型姿势标注数据集的需要。下图为 AAE 训练过程。

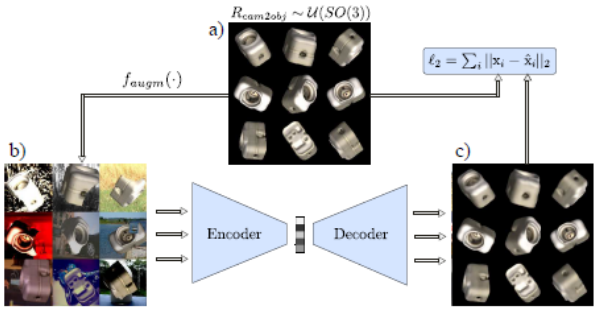


Fig. 4: Training process for the AAE; a) reconstruction target batch $\boldsymbol{x}$ of uniformly sampled SO(3) object views; b) geometric and color augmented input; c) reconstruction $\hat{\boldsymbol{x}}$ after 30000 iterations

研究结果：

作者在 T-LESS 和 LineMOD 数据集上评估了 AAE 和整个 6D 检测管道，仅包括 2D 检测，3D 方向估计和投影距离估计。与最先进的深度学习方法相比，AAE 准确性更好，同时效率更高。另外，作者也分析了一些失败案例，主要源于检测失败或强遮挡。

论文题目：*SinGAN: Learning a Generative Model from a Single Natural Image*

中文题目：SinGAN:从单张图像学习生成模型

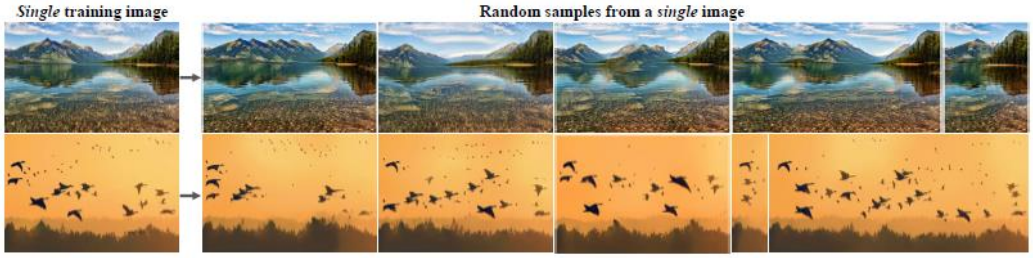
论文作者：Tamar Rott Shaham ,Technion Tali Dekel ,Google Research ,Tomer Michaeli ,Technion

论文出处：ICCV 2019 : IEEE International Conference on Computer Vision.

论文地址：https://arxiv.org/pdf/1905.01164.pdf

研究问题：

生成对抗网络（Generative Adversarial Nets, GAN）在模拟视觉数据的高维分布方面取得了巨大飞跃。特别是用特定类别的数据集（如人脸、卧室）进行训练时，非条件 GAN 在生成逼真的、高质量的样本方面取得了显著成功。但对高度多样化、多种类别的数据集（如 ImageNet）的模拟仍然是一项重大挑战，而且通常需要根据另一输入信号来调整生成或为特定任务训练模型。对单个自然图像中各种图像块的内部分布进行建模已被公认为是许多计算机视觉任务的有用先验。作者将 GAN 带入到一个新领域—从单个自然图像中学习非条件生成模型。单个自然图像通常具有足够的内部统计信息，可学习到强大的生成模型，而不必依赖某个相同类别的数据集。为此，作者提出了一个新的单图像生成模型 SinGAN，能够处理包含复杂结构和纹理的普通自然图像的神经网络。



相对于左边的原始图像，SinGAN 生成新的逼真的图像样本，该样本在创建新的对象配置和结构的同时保留原始图像块分布。

研究方法：

作者的目标是学习一个非条件生成模型，该模型可捕获单个训练图像的内部统计数据。 此任务在概念上与常规 GAN 设置相似，不同之处在于，训练样本是单个图像的多尺度的图像块，而非整个图像样本。为此，SinGAN 生成框架由具有层级结构的 patch-GANs（马尔可夫判别器）组成，其中每个判别器负责捕获不同尺度的分布，这是第一个为从单个图像进行内部学习而探索的网络结构。图像样本从最粗尺度开始，然后依次通过所有的生成器直到最细尺度，且每个尺度都注入噪声。所有生成器和判别器具有相同的感受野，因此，随着生成过程推进可

以捕获更细尺寸的结构。在训练时，对抗损失采用 WGAN-GP 损失，以增加训练稳定性。并设计了一种重建损失来确保可以生成原始图像的特定噪声图谱集合。

研究结果：

作者在图像场景跨度很大的数据集上进行了测试。直观上，SinGAN 很好地保留目标的全局结构和纹理信息，很真实地合成了反射和阴影效果。再使用 AMT 真假用户调研和 FID 的单幅图像版本进行量化。AMT 测试结果表明可以生成很真实的样本，对于细节保留的也更多，人类判别的混淆率较高。FID 结果与 AMT 一致。

3.5 计算机视觉进展

近年来，巨量数据的不断涌现与计算能力的快速提升，给以非结构化视觉数据为研究对象的计算机视觉带来了巨大的发展机遇与挑战性难题，计算机视觉也因此成为学术界和工业界公认的前瞻性研究领域，部分研究成果已实际应用，催生出人脸识别、智能视频监控等多个极具显示度的商业化应用。

计算机视觉的研究目标是使计算机具备人类的视觉能力，能看懂图像内容、理解动态场景，期望计算机能自动提取图像、视频等视觉数据中蕴含的层次化语义概念及多语义概念间的时空关联等。计算机视觉领域不断涌现出很多激动人心的研究成果，例如，人脸识别、物体识别与分类等方面的性能已接近甚至超过人类视觉系统。本文根据近两年计算机视觉领域顶级会议最佳论文及高引论文，对该领域中的技术现状和研究前沿进行了综合分析。

近两年大多数研究都集中在深度学习、检测和分类以及面部/手势/姿势、3D 传感技术等方面。随着计算机视觉研究的不断推进，研究人员开始挑战更加困难的计算机视觉问题，例如，图像描述、事件推理、场景理解等。单纯从图像或视频出发很难解决更加复杂的图像理解任务，一个重要的趋势是多学科的融合，例如，融合自然语言处理领域的技术来完成图像描述的任务。图像描述是一个融合计算机视觉、自然语言处理和机器学习的综合问题，其目标是翻译一幅图片为一段描述文字。目前主流框架为基于递归神经网络的编码器解码器结构其核心思想类似于自然语言机器翻译。但是，由于递归网络不易提取输入图像和文本的空间

以及层次化约束关系，层次化的卷积神经网络以及启发自认知模型的注意力机制受到关注。如何进一步从认知等多学科汲取知识，构建多模态多层次的描述模型是当前图像描述问题研究的重点。

事件推理目标是识别复杂视频中的事件类别并对其因果关系进行合理的推理和预测。与一般视频分析相比，其难点在于事件视频更加复杂，更加多样化，而最终目标也更具挑战性。不同于大规模图像识别任务，事件推理任务受限于训练数据的规模，还无法构建端到端的事件推理系统。目前主要使用图像深度网络作为视频的特征提取器，利用多模态特征融合模型，并利用记忆网络的推理能力，实现对事件的识别和推理认知。当前研究起源于视频的识别和检测，其方法并未充分考虑事件数据的复杂和多样性。如何利用视频数据丰富的时空关系以及事件之间的语义相关性，应是今后的关注重点。

场景理解的目的是计算机视觉系统通过分析处理自身所配置的传感器采集的环境感知数据，获得周围场景的几何/拓扑结构、组成要素（人、车及物体等）及其时空变化，并进行语义推理，形成行为决策与运动控制的时间、空间约束。近年来，场景理解已经从一个初期难以实现的目标成为目前几乎所有先进计算机视觉系统正在不断寻求新突破的重要研究方向。利用社会-长短记忆网络（Social-LSTM）实现多个行人之间的状态联系建模，结合各自运动历史状态，决策出未来时间内的运动走向。此外神经网络压缩方向也是是目前深度学习研究的一个热门的方向，其主要的研究技术有压缩，蒸馏，网络架构搜索，量化等。

综上所述，视觉的发展需要设计新的模型，它们需要能考虑到空间和时间信息；弱监督训练如果能做出好的结果，下一步就是自监督学习；需要高质量的人类检测和视频对象检测数据集；结合文本和声音的跨模态集成；在与世界的交互中学习。

4 知识工程

4.1 知识工程概念

1994 年图灵奖获得者、知识工程的建立者费根鲍姆给出知识工程定义—将知识集成到计算机系统从而完成只有特定领域专家才能完成的复杂任务。在大数据时代，知识工程是从大数据中自动或半自动获取知识，建立基于知识的系统，以提供互联网智能知识服务。大数据对智能服务的需求，已经从单纯的搜集获取信息，转变为自动化的知识服务。我们需要利用知识工程为大数据添加语义/知识，使数据产生智慧（Smart Data），完成从数据到信息到知识，最终到智能应用的转变过程，从而实现对大数据的洞察、提供用户关心问题的答案、为决策提供支持、改进用户体验等目标。知识图谱在以下应用中已经凸显出越来越重要的应用价值：

- 知识融合：当前互联网大数据具有分布异构的特点，通过知识图谱可以对这些数据资源进行语义标注和链接，建立以知识为中心的资源语义集成服务；
- 语义搜索和推荐：知识图谱可以将用户搜索输入的关键词，映射为知识图谱中客观世界的概念和实体，搜索结果直接显示出满足用户需求的结构化信息内容，而不是互联网网页；
- 问答和对话系统：基于知识的问答系统将知识图谱看成一个大规模知识库，通过理解将用户的问题转化为对知识图谱的查询，直接得到用户关心问题的答案；
- 大数据分析与决策：知识图谱通过语义链接可以帮助理解大数据，获得对大数据的洞察，提供决策支持。

我们根据知识工程生命周期各个阶段的关键技术，利用 AMiner 中近年来知识图谱领域的高水平学术论文，挖掘出了包括知识表示（knowledge representation）、知识获取（knowledge acquisition）、知识推理（knowledge reasoning）、知识集成（knowledge integration）和知识存储（knowledge storage）等相关关键词近年来全球活跃的学术研究。此外，结合知识图谱技术，本报告将以上研究领域表示为三级图谱结构，具体分析和处理的方法如下：

1. 使用自然语言处理技术，提取每篇论文文献的关键词，据此，结合学科领域知识图谱，将文章分配到相应领域；

2. 依据学科领域对论文进行聚类，并统计论文数量作为领域的研究热度；

3. 领域专家按照领域层级对学科领域划分等级，设计了三级图谱结构，最后根据概念热度定义当前研究热点。

知识工程三级知识图谱的详细数据可以参见本报告附录，或到 https://www.aminer.cn/data 中直接下载原始数据。鉴于自动分析技术和论文采集的局限性，图谱还可以进一步完善，欢迎读者批评指正，我们会根据根据读者的反馈定期更新。

4.2 知识工程发展历史

回顾知识工程四十年多来发展历程，总结知识工程的演进过程和技术进展，可以将知识工程分成五个标志性的阶段，前知识工程时期、专家系统时期、万维网 1.0 时期，群体智能时期以及知识图谱时期，如下图所示。



图 4-1 知识工程发展历程

- 1950-1970 时期：图灵测试—知识工程诞生前期

人工智能旨在让机器能够像人一样解决复杂问题，图灵测试是评测智能的是手段。这一阶段主要有两个方法：符号主义和连结主义。符号主义认为物理符号系统是智能行为的充要条件，连结主义则认为大脑（神经元及其连接机制）是一切智能活动的基础。这一阶段具有代表性的工作是通用问题求解程序（GPS）：将问题进行形式化表达，通过搜索，从问题初始状态，结合规则或表示得到目标状态。其中最成功应用是博弈论和机器定理证明等。这一时期的知识表示方法主要有逻辑知识表示、产生式规则、语义网络等。这一时代人工智能和知识工程的

先驱 Minsky，Mccarthy 和 Newell 以 Simon 四位学者因为他们在感知机、人工智能语言和通用问题求解和形式化语言方面的杰出工作分别获得了 1969 年、1971 年、1975 年的图灵奖。

- 1970-1990 时期：专家系统—知识工程蓬勃发展期

通用问题求解强调利用人的求解问题的能力建立智能系统，而忽略了知识对智能的支持，使人工智能难以在实际应用中发挥作用。70 年开始，人工智能开始转向建立基于知识的系统，通过“知识库+推理机”实现机器智能，这一时期涌现出很多成功的限定领域专家系统，如 MYCIN 医疗诊断专家系统、识别分子结构的 DENRAL 专家系统以及计算机故障诊断 XCON 专家系统等。斯坦福人工智能实验室的奠基人 Feigenbaum 教授在 1980 年的一个项目报告《Knowledge Engineering：The Applied Side of Artificial Intelligence》中提出知识工程的概念，从此确立了知识工程在人工智能中的核心地位。这一时期知识表示方法有新的演进，包括框架和脚本等。80 年代后期出现了很多专家系统的开发平台，可以帮助将专家的领域知识转变成计算机可以处理的知识。

- 1990-2000 时期：万维网

在 1990 年到 2000 年，出现了很多人工构建大规模知识库，包括广泛应用的英文 WordNet，采用一阶谓词逻辑知识表示的 Cyc 常识知识库，以及中文的 HowNet。Web1.0 万维网的产生为人们提供了一个开放平台，使用 HTML 定义文本的内容，通过超链接把文本连接起来，使得大众可以共享信息。W3C 提出的可扩展标记语言 XML，实现对互联网文档内容的结构通过定义标签进行标记，为互联网环境下大规模知识表示和共享奠定了基础。这一时期在知识表示研究中还提出了本体的知识表示方法。

- 2000-2006 时期：群体智能

在 2001 年，万维网发明人、2016 年图灵奖获得者 Tim Berners-Lee 在科学美国人杂志中发表的论文《The Semantic Web》正式提出语义 Web 的概念，旨在对互联网内容进行结构化语义表示，利用本体描述互联网内容的语义结构，通过对网页进行语义标识得到网页语义信息，从而获得网页内容的语义信息，使人和机器能够更好地协同工作。W3C 进一步提出万维网上语义标识语言 RDF（资源描

述框架）和 OWL（万维网本体表述语言）等描述万维网内容语义的知识描述规范。

万维网的出现使得知识从封闭知识走向开放知识，从集中构建知识成为分布群体智能知识。原来专家系统是系统内部定义的知识，现在可以实现知识源之间相互链接，可以通过关联来产生更多的知识而非完全由固定人生产。这个过程中出现了群体智能，最典型的代表就是维基百科，实际上是用户去建立知识，体现了互联网大众用户对知识的贡献，成为今天大规模结构化知识图谱的重要基础。

- 2006 年至今：知识图谱—知识工程新发展时期

从 2006 年开始，大规模维基百科类富结构知识资源的出现和网络规模信息提取方法的进步，使得大规模知识获取方法取得了巨大进展。与 Cyc、WordNet 和 HowNet 等手工研制的知识库和本体的开创性项目不同，这一时期知识获取是自动化的，并且在网络规模下运行。当前自动构建的知识库已成为语义搜索、大数据分析、智能推荐和数据集成的强大资产，在大型行业和领域中正在得到广泛使用。典型的例子是谷歌收购 Freebase 后在 2012 年推出的知识图谱（Knowledge Graph），Facebook 的图谱搜索，Microsoft Satori 以及商业、金融、生命科学等领域特定的知识库。最具代表性大规模网络知识获取的工作包括 DBpedia、Freebase、KnowItAll、WikiTaxonomy 和 YAGO，以及 BabelNet、ConceptNet、DeepDive、NELL、Probase、Wikidata、XLORE、Zhishi.me、CNDBpedia 等。这些知识图谱遵循 RDF 数据模型，包含数以千万级或者亿级规模的实体，以及数十亿或百亿事实（即属性值和与其他实体的关系），并且这些实体被组织在成千上万的由语义体现的客观世界的概念结构中。

目前知识图谱的发展和应用状况，除了通用的大规模知识图谱，各行业也在建立行业和领域的知识图谱，当前知识图谱的应用包括语义搜索、问答系统与聊天、大数据语义分析以及智能知识服务等，在智能客服、商业智能等真实场景体现出广泛的应用价值，而更多知识图谱的创新应用还有待开发。

在我国知识工程领域研究中，中科院系统所陆汝钤院士、计算所史忠植研究员等老一代知识工程研究学者为中国的知识工程研究和人才培养做出了突出贡献，例如，陆汝钤院士因在知识工程和基于知识的软件工程方面作出的系统和创

造性工作，以及在大知识领域的开创性贡献，荣获首届“吴文俊人工智能最高成就奖”[6]。

4.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为知识工程领域全球学者分布情况：



图 4-2 知识工程领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布在其东西海岸；欧洲及亚洲东部也有较多的人才分布；其他诸如非洲、南美洲等地区的学者非常稀少；知识工程领域的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，知识工程领域中男性学者占比 89.7%，女性学者占比 10.6%，男性学者占比远高于女性学者。

知识工程领域学者的 h-index 分布如下图所示，大部分学者的 h-index 分布在中低区域，其中 h-index 在 20-30 区间的人数最多，有 783 人，占比 38.9%，小于 20 区间的人数最少，有 90 人。

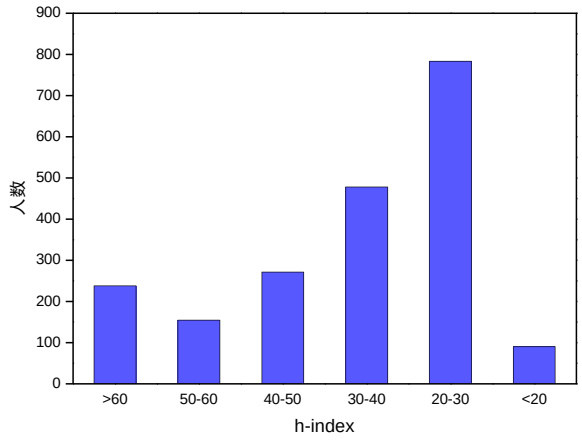


图 4-3 知识工程领域学者 h-index 分布

- 中国人才分布

我国专家学者在知识工程领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是珠三角和长三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等亚洲国家相比，中国在知识工程领域学者数量较多。



图 4-4 知识工程领域中国学者分布

中国与其他国家在知识工程领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 4-1 知识工程领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	541	17306	32	1092
中国-新加坡	116	4107	35	244
中国-澳大利亚	111	3634	33	237
中国-英国	27	352	13	52
中国-加拿大	24	632	26	58
中国-日本	21	572	27	56
中国-丹麦	14	328	23	23
中国-德国	10	344	34	20
中国-印度	10	76	8	22
中国-希腊	10	197	20	20

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在知识工程领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席；中国与新加坡合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中国与新加坡合作达到了较高的水平。

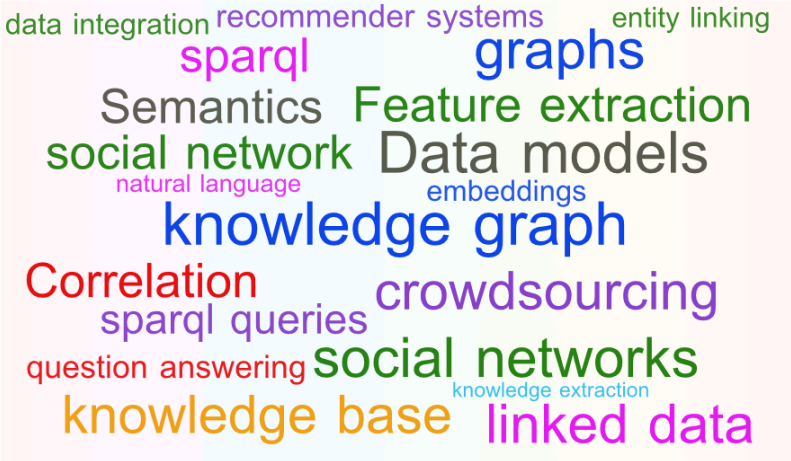
4.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

IEEE Transactions on Knowledge and Data Engineering

International Conference on Information and Knowledge Management

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，知识图谱（knowledge graph）、数据模型（Data models）、社交网络（social networks）是本领域中最热的关键词。



论文题目：*Convolutional 2D Knowledge Graph Embeddings*

中文题目：基于二维卷积的知识图谱嵌入表示学习

论文作者：Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, Sebastian Riedel

论文出处：The Thirty-Second AAAI Conference on Artificial Intelligence （AAAI 2018）

论文地址：

https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/download/17366/15884

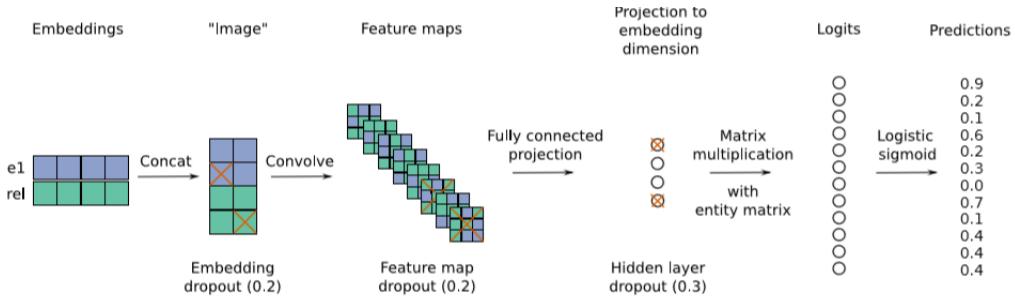
研究问题：

知识图谱的链接预测任务是预测节点之间潜在的关系。传统的链接预测方法专注于浅的、快速的模型，因为这样可以扩展到大规模的 KG 中。但是浅层模型学习到的特征比深沉模型少很多，大大限制了模型的性能。解决该问题的方法之一是增加 embedding 的维度，但是会增加模型参数量，不方便扩展到大规模 KG 中。此外，部分现有数据集中有测试集泄露问题：训练集中的三元组稍微翻转一下就可以得到测试集三元组，然后使用基于规则的模型就能达到最佳性能。文章通过构造一个简单的翻转来衡量这个问题严重性，并清洗了部分数据来解决该问题。

研究方法：

文章提出一种多层卷积神经网络模型用于知识图谱的链接预测任务。与自然

语言处理中常用的一维卷积不同，文章通过把多个向量堆叠成矩阵，就可以像图形一样用二维卷积核来抽取 embedding 之间的关系。



如上图所示，评分函数如下公式

$$\psi_r(\mathbf{e}_s, \mathbf{e}_o) = \quad f(\mathrm{vec}(f([\overline{\mathbf{e}_s}; \overline{\mathbf{r}_r}] * \omega))\mathbf{W})\mathbf{e}_o$$

模型的流程总结为：

- 经过 look-up embedding 得到实体和关系的向量表示，然后通过变形和堆叠转化为 2D 版本。
- 用多个卷积核对堆叠后的矩阵进行卷积操作，得到一个特征图 γ 。
- 把 γ 向量化，然后通过一个全连接层映射到 k 维空间中。
- 最后与目标实体的 embedding 相乘获得相应的得分。
- 将得分进行 sigmoid 操作得到概率 p，从而最小化交叉熵来训练模型。

值得一提的是，与传统模型对三元组关系（s, r, o）打分的 1-1 scoring 模式不同，ConvE 以实体关系对（s, r）作为输入，同时对所有实体 o 进行打分，即 1-N scoring。这种方式极大加快了计算速度。实验结果表明，即使实体个数扩大 10 倍，计算时间也只是增加了 25%.

研究结果：

作者在 4 个数据集 WN18、FB15K、YAGO3-10、Countries 上进行实验，与 DisMult、R-GCN 等模型进行了对比。实验结果表明：0.23M 个参数的 ConvE 就与 1.89M 个参数的 DistMult 有相近的性能表现，总的来说 ConvE 的参数效率是 R-GCN 的 17 倍以上，是 DistMult 的 8 倍以上。此外，作者还发现 ConvE 在

YAGO3-10 和 FB15k-237 上的表现比在 WN18RR 上好，因为前两者包含入度很大的结点，比如结点 United States 在“was born in”上的入度超过 10000，这种复杂的 KG 需要 deeper 模型，而浅层模型比如 DistMult 则在较简单的 KG 上有优势。

Model	Param. count	Emb. size	MRR	Hits @10	Hits @3	Hits @1
DistMult	1.89M	128	.23	.41	.25	.15
DistMult	0.95M	64	.22	.39	.25	.14
DistMult	0.23M	16	.16	.31	.17	.09
ConvE	5.05M	200	.32	.49	.35	.23
ConvE	1.89M	96	.32	.49	.35	.23
ConvE	0.95M	54	.30	.46	.33	.22
ConvE	0.46M	28	.28	.43	.30	.20
ConvE	0.23M	14	.26	.40	.28	.19

论文题目：*Explainable Reasoning over Knowledge Graphs for Recommendation*

中文题目：基于知识图谱路径推理的可解释推荐

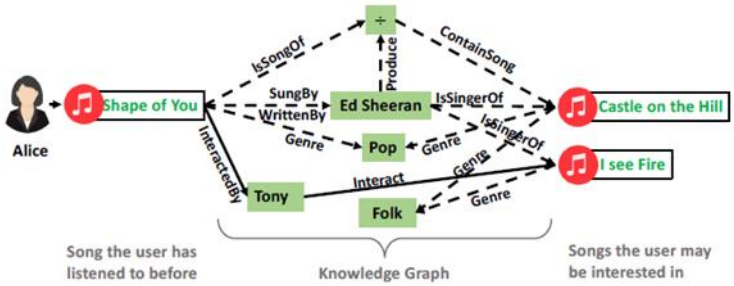
论文作者：Xiang Wang, DingxianWang, Canran Xu, Xiangnan He, Yixin Cao, Tat-Seng Chua[1]

论文出处：Proceedings of the AAAI Conference on Artificial Intelligence. 2019（AAAI'19）.

论文地址：https://www.aaai.org/ojs/index.php/AAAI/article/view/4470/4348

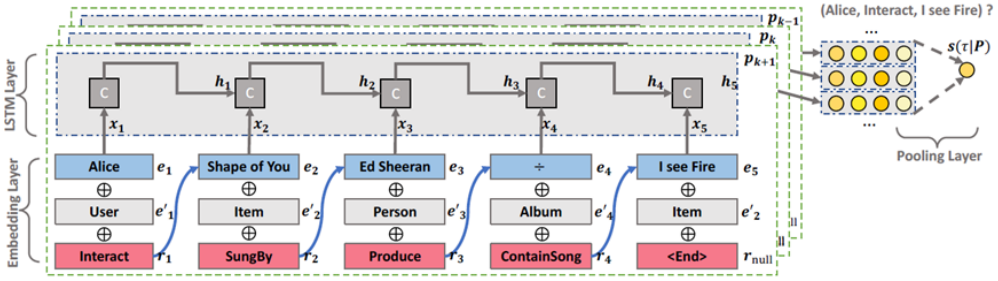
研究问题：

近年来如何将知识图谱融入推荐系统得到越来越多的关注，通过探索知识图谱中的用户到商品的路径，可以为用户与商品的交互行为提供丰富的补充信息。这些路径不仅揭示了实体和关系的语义，还能帮助理解用户的兴趣。然而现有的模型没能充分利用路径来推断用户偏好，尤其是在建模的顺序依赖关系和路径的整体语义方面。文章构建了知识感知路径递归网络模型（Knowledge aware Path Recurrent Network，KPRN），通过组合实体和关系的语义来生成路径表示。利用路径中的顺序依赖关系，可以基于路径进行有效推理，从而推断出用户-项目交互场景中的基本原理。此外，文章设计了一种新的权重池化操作，以区分用户与项目连接的不同路径的优势，赋予我们的模型一定的可解释性。下图为基于知识图谱的音乐推荐场景实例，虚线为关系，实线为用户-商品交互路径。



研究方法：

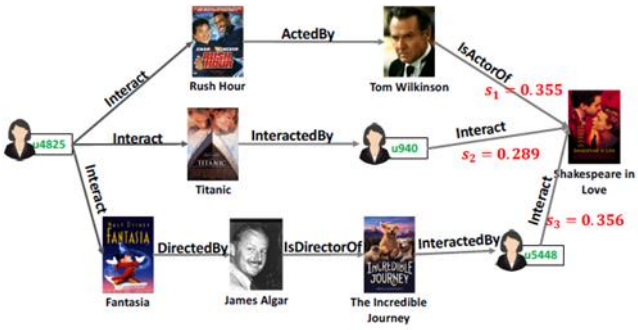
知识图谱和路径：知识图谱由一组三元组（h, r, t）组成，代表实体 h 和实体 t 构成关系 r。文章中知识图谱还融合了用户-商品的交互信息，即包含了三元组（用户，交互方式，商品），其中交互方式为预定义的关系。路径的定义是一个由实体或关系组成的序列，并且以用户作为起始点，以商品作为终点。给定一个用户，商品，以及连接该用户和商品的路径的集合，我们希望模型能够计算该用户和商品之间存在交互的可能性，即是否存在三元组（用户，交互方式，商品）。



模型一共分为三层：在 Embedding 层对路径的每一个行为做 embedding 的计算。对于给定三元组，分别计算实体名称、实体类型和关系（或交互方式）的 embedding 后再拼接得到最终特征表示。LSTM 层将路径上的每个单元的特征表示按照时间顺序输入并且将最后时刻的隐藏层状态作为该路径的特征表示。在 pooling 层，将所有路径的特征表示集合输入两层前馈神经网络，再对输出做带权重的池化操作得到最后的预测结果。

研究结果：

文章在公开电影数据集 MI 和音乐数据集 KKBox 上进行了实验，验证了所提出的模型的有效性，并且相对于仅将实体映射为一个向量表示的方法，KPRN 还能够从路径中挖掘用户和商品之间的交互关系，这提高了模型的可解释性。



如上图所示，在 MovieLens-1M 中随机选择的一个用户 u4825，并从她的交互记录中选择电影“恋爱中的莎士比亚”。然后，我们提取连接用户-项对的所有限定路径，得到每个路径的分数 $s_1 = 0.355$，$s_2 = 0.289$，$s_3 = 0.356$，即模型更倾向于认为用户 u4825 是通过路径 3 和电影“恋爱中的莎士比亚”产生交互关系。

论文题目：*Knowledge Graph Embedding with Iterative Guidance from Soft Rules*

中文题目：基于规则迭代引导的增强知识图谱表示学习

论文作者：Xiang Wang, DingxianWang, Canran Xu, Xiangnan He, Yixin Cao, Tat-Seng Chua

论文出处：Proceedings of the AAAI Conference on Artificial Intelligence. 2018（AAAI'18）.

论文地址：https://arxiv.org/abs/1711.11231v1

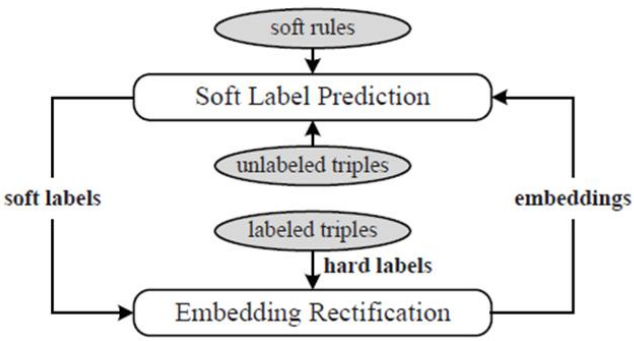
研究问题：

知识图谱表示学习旨在将实体和关系嵌入到向量空间，同时保留知识图谱的内在结构。传统方法主要基于关系三元组学习知识图谱的嵌入表示。本文认为，知识图谱中的逻辑规则对于表示学习也很有帮助，于是提出了一个新的表示学习模型 RUGE（Rule-Guided Embedding）。RUGE 的主要特点是，利用从知识图谱自动抽取的规则迭代地预测未标记三元组，并将其加入训练来增强表示学习。

研究方法：

为了在知识图谱进行分布式表示学习的过程中加入逻辑规则进行引导，RUGE 同时利用标注三元组（labeled Triples）、未标注三元组（Unlabeled Triples）、

自动抽取出的软规则（soft rules）这三种资源以迭代的方式进行知识图谱表示学习。软规则指的是不总是成立，带置信度的规则。每一轮迭代在软标签预测和 embedding 修正这两个步骤间交替进行。前者利用当前学到的 embedding 和软规则为未标注三元组预测软标签；后者进一步利用标注三元组（硬标签）和未标注三元组（软标签）对当前的 embedding 进行修正。通过这个迭代过程，RUGE 可以成功建模分布式知识表示学习和逻辑推理二者间的交互性，逻辑规则中蕴含的丰富知识也能被更好地传递到所学习的分布式表示中。



上图为 RUGE 框架图，可以看出，RUGE 使嵌入模型能够以迭代方式同时从标记的三元组、未标记的三元组和软规则中学习。在每次迭代中，模型交替在软标签预测阶段和 embeddings 校正阶段之间。

学习资源构建：文章假设在知识图谱中观测到的三元组集合为正三元组，使用随机替换头尾实体的方式构建负三元组，此外还考虑那些能被软规则（soft rules）编码的未标记三元组，其中规则为不同置信度水平的 FOL 规则。

三元组和规则建模：对于三元组建模，文章采用现有模型 ComplEx，三元组的真值可以直接计算得到。对于规则建模，也就是建模规则 groundings 的真值，文章采用 T-norm fuzzy logics，规则的真值等于其三元组真值的逻辑组合。

软标签预测：可以基于 embedding 表示来计算标记和未标记和三元组的“真值”；也可以基于规则 groundings 的真值来计算三元组的实际真值，即软标签（soft label）。文章希望这两个真值应该是接近的，且应当使得规则 groundings 为真为此设置了相应的优化目标。

Embedding 校正：得到了未标记三元组的软标签之后，文章结合已标记三元组，使用交叉熵进行统一优化。

研究结果：

文章的主要实验任务是传统的关系预测。数据集采用了 FB15K 和 YAGO37。实验结果如下表所示，可以看见，RUGE 相比基线方法取得了较好的结果。文章创新性在于提出了软规则，并可以成功建模分布式知识表示学习和逻辑推理二者间的交互性，逻辑规则中蕴含的丰富知识也能被更好地传递到所学习的分布式表示中。

	FB15K						YAGO37					
			HITS@N						HITS@N			
Method	MRR	MED	1	3	5	10	MRR	MED	1	3	5	10
TransE	0.400	4.0	0.246	0.495	0.576	0.662	0.303	13.0	0.218	0.336	0.387	0.475
DistMult	0.644	**1.0**	0.532	0.730	0.769	0.812	0.365	6.0	0.262	0.411	0.493	0.575
HolE	0.600	2.0	0.485	0.673	0.722	0.779	0.380	7.0	0.288	0.420	0.479	0.551
ComplEx	0.690	**1.0**	0.598	0.756	0.793	0.837	0.417	**4.0**	0.320	0.471	0.533	**0.603**
PTransE	0.679	**1.0**	0.565	0.768	0.810	0.855	0.403	9.0	0.339	0.444	0.473	0.506
KALE	0.523	2.0	0.383	0.616	0.683	0.762	0.321	9.0	0.215	0.372	0.438	0.522
RUGE	**0.768**	**1.0**	**0.703**	**0.815**	**0.836**	**0.865**	**0.431**	**4.0**	**0.340**	**0.482**	**0.541**	**0.603**

论文题目：*Variational Reasoning for Question Answering with Knowledge Graph*

中文题目：基于知识图谱的问答变分推理

论文作者：Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J. Smola, and Le Song

论文出处：Proceedings of the AAAI Conference on Artificial Intelligence. 2018（AAAI'18）.

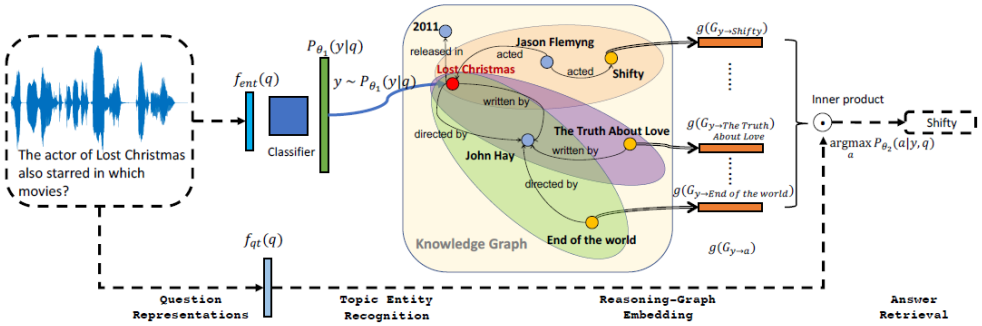
论文地址：https://arxiv.org/abs/1709.04071v1

研究问题：

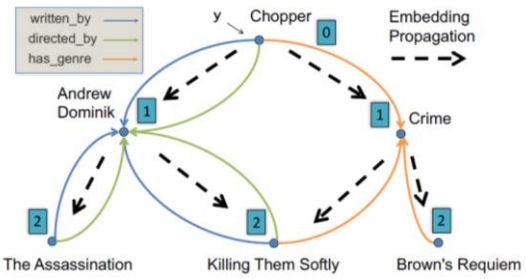
传统的知识图谱问答主要是基于语义解析的方法，这种方法通常是将问题映射到一个形式化的逻辑表达式，然后将这个逻辑表达转化为知识图谱的查询。问题答案可以从知识图谱中通过转化后的查询得到。然而传统的基于语义解析的知识库问答会存在一些挑战，如基于查询的方法只能获取一些明确的信息，对于知识库中需要多跳才能获取的答案则无法回答；此外实际的场景中用户的输入可能是通过语音识别转换而来或者是用户通过打字输入而来，很难确保不存在一定的噪声。在具有噪声的场景下，问句中的实体很难直接准确的匹配到知识库上。因此文章提出了一个端到端的知识库问答模型来解决以上两个问题。

研究方法：

论文提出了变分推理网络（VRN），模型分为两部分：通过概率模型来识别问句中的实体（得到图谱中每个实体是问句中实体的概率），这避免了语义解析带来的误差。具体而言，将问句（基于语音或者文本）映射成向量，然后对其做 softmax 多分类，以计算问句中的实体的概率。



问答时在知识图谱上做逻辑推理，且推理规则将被学习出来。给定 query 和问题实体 y，希望模型从知识图谱中找到答案 a。文章首先将 query 通过另一个网络编码成向量 q，然后从实体 y 沿着知识图谱向相邻实体扩展搜索答案，形成推理路径$\mathcal{G}_y$。下图是两跳推理路径的示意图：



研究结果：

实验结果显示在 Vanilla、NTM 和 Audio 数据集下，算法的效果都超过传统的 QA 系统，同时在需要推理的问题中性能更为显著。其中，在文章新发布的 KBQA 数据集 MetaQA 上相比对照模型提升较为明显，尤其是要求多跳推理的问题。另外，在问题的形式是语音，机器翻译后的结果，以及训练时不给定标注好的 topicentity 的情况下，都有较大的提升。

	NTM-EU 1-hop	NTM-EU 2-hop	NTM-EU 3-hop	Audio-EU 1-hop	Audio-EU 2-hop	Audio-EU 3-hop
VRN	**81.3**	**69.7**	**38.0**	**37.0**	**24.6**	**21.1**
Bordes et al. [22]'s QA system	32.5	32.3	25.3	18.5	19.3	15.3
KV-MemNN	33.9	8.7	10.2	4.3	7.0	15.3
Supervised embedding	16.1	22.8	24.2	4.1	6.1	12.1

	Vanilla 1-hop	Vanilla 2-hop	Vanilla 3-hop	Vanilla-EU 1-hop	Vanilla-EU 2-hop	Vanilla-EU 3-hop
VRN	**97.5**	**89.9**	**62.5**	**82.0**	**75.6**	**38.3**
Bordes et al. [22]'s QA system	95.7	81.8	28.4	39.5	38.3	26.9
KV-MemNN	95.8	25.1	10.1	35.8	10.3	10.5
Supervised embedding	54.4	29.1	28.9	18.1	23.2	25.3

论文题目：*TorusE: Knowledge Graph Embedding on a Lie Group*

中文题目：TorusE：一种基于李氏群的知识图谱嵌入表示学习

论文作者：Yuyu Zhang, Hanjun Dai, Zornitsa Kozareva, Alexander J. Smola, and Le Song

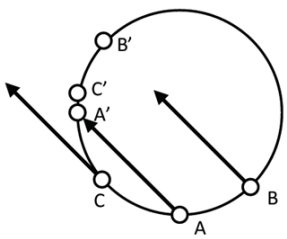
论文出处：Proceedings of the AAAI Conference on Artificial Intelligence. 2018（AAAI'18）

论文地址：

https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/viewPDFInterstitial/16227/15885

研究问题：

知识图谱在许多人工智能任务都扮演重要的角色。知识图谱通常是用一个三元组（h, r, t）来表示一条知识。为了获得它们的低维稠密表示，通常采用分布式向量表示，例如 TransE 模型。然而 TransE 的正则约束会迫使实体的向量表示在一个球面上，而这与之前的优化条件又是相互矛盾的。



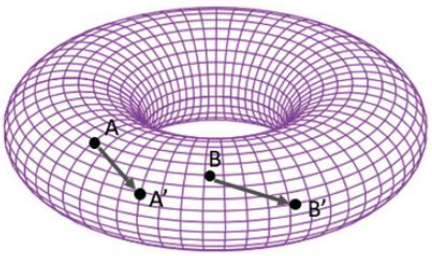
这种矛盾还会影响实体间连接预测的准确性。以下图为例，箭头方向表示关系 r，A, B, C及A', B', C'表示实体，对于(C, r, C')和(B, r, B')的正则约束和优化目标就是相互矛盾的。

研究方法：

文章提出的 TorusE 拥有类似 TransE 遵循的优化目标和正则项。为了避免上述的正则项带来的矛盾，TorusE 不再将特征学习到一个开流形（open manifold）的欧式空间，而是在紧空间（compact space）上学习知识图谱的嵌入表示。可以证明，紧李群可以满足 TransE 遵循的优化目标和正则项条件，即嵌入空间为可微的流形空间，满足阿贝尔群性质，群运算可微且能够定义距离函数。证明了任意一个阿贝尔李群符合嵌入空间的需求后，文章构建了一个紧李群的圆环空间T^n和圆环空间上的不同范式的距离函数d_{L_1}，d_{L_2}，d_{eL_2}。具体定义如下下左图所示。

Definition 2 *An n-dimensional torus T^n is a quotient space, $\mathbb{R}^n/\sim = \{[\boldsymbol{x}] | \boldsymbol{x} \in \mathbb{R}^n\} = \{\{\boldsymbol{y} \in \mathbb{R}^n | \boldsymbol{y} \sim \boldsymbol{x}\} | \boldsymbol{x} \in \mathbb{R}^n\}$, where $\sim$ is an equivalence relation and $\boldsymbol{y} \sim \boldsymbol{x}$ if and only if $\boldsymbol{y} - \boldsymbol{x} \in \mathbb{Z}^n$.*

- d_{L_1}: A distance function d_{L_1} on T^n is derived from the L_1 norm of the original vector space by defining $d_{L_1}([\boldsymbol{x}], [\boldsymbol{y}]) = \min_{(\boldsymbol{x}', \boldsymbol{y}') \in [\boldsymbol{x}] \times [\boldsymbol{y}]} \|\boldsymbol{x}' - \boldsymbol{y}'\|_1$.
- d_{L_2}: A distance function d_{L_2} on T^n is derived from the L_2 norm of the original vector space by defining $d_{L_2}([\boldsymbol{x}], [\boldsymbol{y}]) = \min_{(\boldsymbol{x}', \boldsymbol{y}') \in [\boldsymbol{x}] \times [\boldsymbol{y}]} \|\boldsymbol{x}' - \boldsymbol{y}'\|_2$.
- d_{eL_2}: T^n can be embedded on $\mathbb{C}^n$ by g. A distance function d_{eL_2} on T^n is derived from the L_2 norm of the $\mathbb{C}^n$ by defining $d_{eL_2}([\mathbf{x}], [\mathbf{y}]) = \|g([\mathbf{x}]) - g([\mathbf{y}])\|_2$.



类似于 TransE 在$\mathbb{R}^n$上的优化目标$h + r = t$，TorusE 在T^n上构建$[h] + [r] = [t]$，并根据距离函数的不同定义三个对应的评分函数：以 2 维的 TorusE 模型为例（上右图），箭头方向表示关系 r，对于三元组(A, r, A')和(B, r, B')，映射至嵌入空间后，我们仍可以得到$[A'] - [A]$与$[B'] - [B]$在圆环空间上是相似的。

研究结果：

文章在知识表示嵌入的可扩展性和连接预测上将 TorusE 和 TransE 进行对比。实验结果表明，TorusE 具有比 TransE 更快的计算表现，证明了 TorusE 具有更低的计算复杂度。在连接预测的任务上，TorusE 比当今最好的模型仍要表现出色（见下表）。

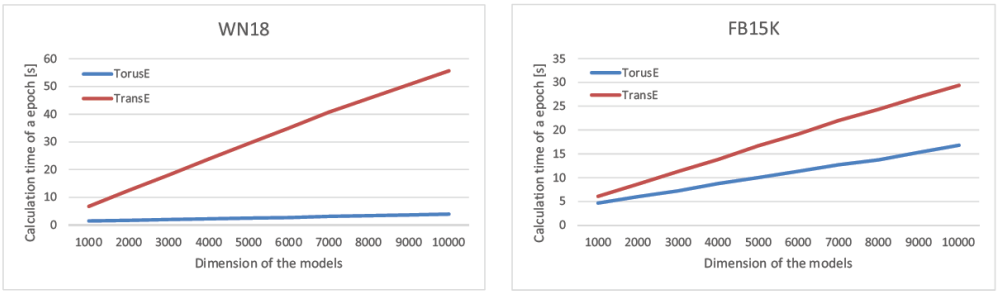


Figure 4: Calculation time of TorusE and TransE on WN18 and FB15K

	WN18					FB15K				
	MRR		HITS@			MRR		HITS@		
Model	Filtered	Raw	*1*	*3*	*10*	Filtered	Raw	*1*	*3*	*10*
TransE	0.397	0.306	0.040	0.745	0.923	0.414	0.235	0.247	0.534	0.688
TransR	0.605	0.427	0.335	0.876	0.940	0.346	0.198	0.218	0.404	0.582
RESCAL	0.890	0.603	0.842	0.904	0.928	0.354	0.189	0.235	0.409	0.587
DistMult	0.822	0.532	0.728	0.914	0.936	0.654	0.242	0.546	0.733	0.824
ComplEx	0.941	0.587	0.936	0.945	0.947	0.692	0.242	0.599	0.759	**0.840**
TorusE	**0.947**	**0.619**	**0.943**	**0.950**	**0.954**	**0.733**	**0.256**	**0.674**	**0.771**	0.832

论文题目：*Commonsense Knowledge Aware Conversation Generation with Graph Attention*

中文题目：基于图注意力机制的常识感知对话生成

论文作者：Hao Zhou, Tom Young, Minlie Huang, Haizhou Zhao, Jingfang Xu, Xiaoyan Zhu

论文出处：*Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*

论文地址：https://www.ijcai.org/proceedings/2018/0643.pdf

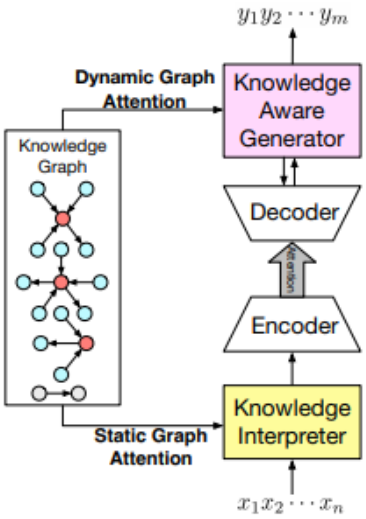
研究问题：

文章着重研究了基于常识知识库的开放域对话生成问题。具体来说，该问题的输入为一个常识知识库和开放域的问句，模型需要生成一个符合上下文语境的回答作为输出。

研究方法：

整体来说，对于一个输入的问题，文章提出的模型会从常识知识库中检索相应的知识图，然后基于静态图注意力机制对其进行编码，图注意力机制有助于提升语义信息，从而帮助系统更好地理解问题。接下来，在语句的生成过程中，模

型会逐个读取检索到的知识图以及每个图中的知识三元组，并通过动态的图注意力机制来优化语句的生成。



具体的，模型可以分为 3 个主要的部分，分别如下：（1）知识解析器，这部分模型旨在优化问题理解这一环节。它通过引入每个单词对应的图向量，来增强单词的语义。知识解析器会把问题中的每个单词作为关键实体，从整个常识知识库中检索图；（2）静态图注意力，静态图注意力机制可以看做是知识解析器的一部分，它可以为检索到的知识图提供一个表现形式；（3）知识感知生成器与动态图注意力，这部分模型会读取所有的知识图和每个图中的所有三元组，用来生成最终的回复。

研究结果：

该论文使用 ConceptNet 作为常识知识库，并基于 reddit 上一问一答形式的对话数据，构建了常识对话数据集。由于论文的目标是用常识知识优化语言理解和生成，所以数据集仅保留了带有知识三元组的原始语料数据。

论文使用自动评估和人工评估两种方法同时对模型的表现进行评价，并选取了 Seq2Seq、MemNet 和 CopyNet 三种基线模型进行对比。最终研究结果表明，本篇论文提出的模型在两种评价方法下，都比几个基线模型表现突出。此外，该论文还做了案例研究，对于一个具体的问题，本篇论文的模型可以生成更为合理，信息也更丰富的回复。

论文题目：*That's Interesting, Tell Me More! Finding Descriptive Support Passages for Knowledge Graph Relationships*

中文题目：知识图谱关系的描述性段落查找

论文作者：Sumit Bhatia, Purusharth Dwivedi, Avneet Kaur.

论文出处：*Proceedings of the International Semantic Web Conference 2018*

论文地址：https://link.springer.com/chapter/10.1007/978-3-030-00671-6_15

研究问题：

近年来，知识图谱的应用日益增多，其可靠性成为一个待解决的关键问题。文章研究如何从文本语料中为知识图谱的三元组寻找支持段落，从而增强用户对知识图谱的信任并帮助其做出更正确的决策。

研究方法：

文章提出了一种概率方法来表示一个段落描述了特定三元组的概率。通过使用贝叶斯公式，并假设三元组的头实体、谓语、尾实体三个元素条件独立，可以将目标概率拆解为段落描述头实体、段落描述谓语、段落描述尾实体三个概率的乘积。进一步，由于头实体、谓语、尾实体均由一个或多个单词组成，可以将这些单词提取出来，将目标概率继续拆解为给定段落描述了某个单词的概率的乘积。

为了表示段落 p 描述单词 w 的概率，文章基于语言模型进行改进，提出考虑三个层次的因素，分别是段落因素、文档因素和集合因素。其中段落因素考虑 w 直接在 p 中出现的概率；文档因素考虑 w 在 p 所在的文档中出现的概率，从而解决段落中使用代词来指代实体时直接概率过低的问题；集合因素考虑整个文本语料，相当于背景语言模型，与信息检索领域常用的 IDF 作用相当。文章使用统计频率来计算以上三个因素的概率，计算方法如下。

$$\text{Passage Evidence: } P(w|\theta_p) = \frac{count(w,p)+1}{|p|+|V|}$$

$$\text{Document Evidence: } P(w|\theta_d) = \frac{count(w,d)+1}{|d|+|V|}$$

$$\text{Collection Evidence: } P(w|\theta_c) = \frac{count(w,c)}{|C|}$$

将三个概率加权相加，得到段落描述单词的概率，再进行累乘从而得到段落描述三元组的概率。最终为每个三元组对所有候选段落根据该概率进行排序。

研究结果：

文章从 WikiData 中构造了 50 个三元组作为目标三元组，以 Wikipedia 作为文本语料，对每一个三元组选出了语料中排名最高的 5 个段落作为结果，并依靠 3 名标注人员对结果进行评价。评价分为三个等级，分别是该段落与三元组相关、部分相关和不相关。

	Evaluator 1	Evaluator 2	Evaluator 3	Final
non-relevant	406	438	444	449
partially-relevant	41	11	43	12
relevant	248	246	208	234

对于基线方法，使用相同的三元组进行抽取和评价。最终结果表明，文章提出的方法相比于基线方法有显著改进，排名第一的段落的准确率从 0.251 提高到了 0.860。

	P@1	Precision	MRR
Inf. N/w	0.251	0.156	0.272
Inf. N/w + Rel. Exp.	0.165	0.088	0.144
Proposed Approach	**0.860**	**0.727**	**0.805**

论文题目：*HighLife: Higher-arity Fact Harvesting*

中文题目：HighLife：高精度事实获取

论文作者：Patrick Ernst, Amy Siu, Gerhard Weikum

论文出处：Proceedings of the 2018 World Wide Web Conference

论文地址：https://dl.acm.org/citation.cfm?id=3186000

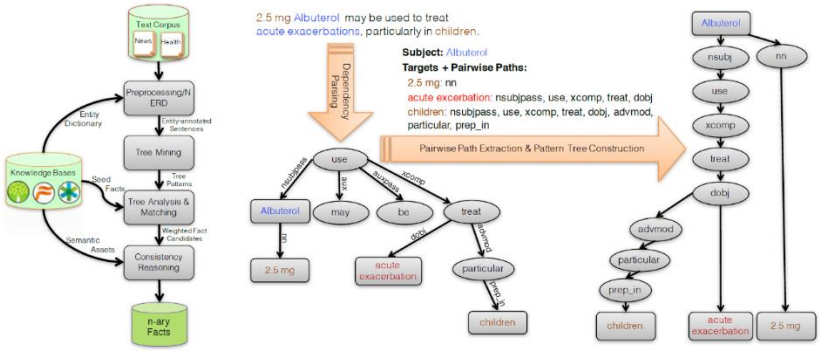
研究问题：

现有的知识抽取方法主要关注二元实事，即两个实体之间的关系。然而在实际中，我们往往需要考虑三元关系甚至是更加高阶的关系，这种高阶的关系能够更加精确地表达一个事实。比如在医疗领域，我们需要知道一种药品治疗哪一种

疾病，这种药品的使用剂量是什么以及适合哪种类型的患者（例如，儿童或者是成人）。这篇文章提出了一种从文本中获取高阶实事的方法，基于远程监督的假设，从一些种子集合出发。一方面为了提高召回率，利用二元的事实模板去发现尽可能多的事实；另一方面为了提高准确率，设计了一种基于约束的推理方法来去除错误的候选。主要的创新点是解决了高阶事实在文本中的表达不集中，分布不规律的问题。例如，一句话可以指一种药物、一种疾病和一组患者，而另一句话则是指药物、其剂量和目标人群，而没有提到疾病。这篇文章的方法在模式学习和约束推理阶段都能很好地处理这些部分观察到的事实。对健康相关文档和新闻文章的实验证明了这种方法的可行性。

研究方法：

这篇文章的框架名叫 HighLife，由两部分组成。首先使用种子事实作为远程监控来学习模板，使用这些模板来提取候选事实，并迭代这些步骤，保证召回率。进而将事实从二元扩展到更高阶的情况。在实现高召回率的同时，这种方法容易受到噪声和目标漂移的影响。因此，作者使用约束推理来消除虚假的候选事实，将基于 MaxSat 的推理器扩展到更高阶情况。例如，可以应用类型约束来确定何时赢得普利策奖的事实是电影或歌曲（而不是书籍）。



研究结果：

为了证明方法的有效性和通用性，文章在新闻和生物医学领域上分别进行了实验。将 HighLife 模型及其变种和语义角色标注的基线方法进行对比，结果表明 Highlife 模型显著优于已有的 SRL 方法，进一步分析了 Highlife 模型在两个数据集上的表现，发现在医学领域的效果更好。

4.5 知识工程最新进展

近两年知识获取、推理和应用研究取得了显著的进展，主要表现在如下几个方面：

- 资源匮乏情况下的知识获取

知识图谱的构建始终是知识图谱领域的核心问题之一，近年来除了传统的有监督的实体、关系、事件知识获取的研究外，也涌现了一批在弱资源情况下的知识获取方法。例如：在集合扩展（实体集扩展）研究中，Learning to Bootstrap for Entity Set Expansion 使用蒙特卡洛树搜索策略的 booststrap 方法有效地提升了实体集扩展方法的稳定性，尤其是在与分类体系相关任务的同时优化上。HiExpan: Task-Guided Taxonomy Construction by Hierarchical Tree Expansion 提出一个知识分类体系的扩展框架，模型利用弱监督关系抽取模型，从一个小型的上下位关系树出发，抽取扩展的节点并扩展成一个更加丰富的上下位体系。FewRel 2.0: Towards More Challenging Few-Shot Relation Classification 提出了少次学习任务，通过设计少次学习机制，能够利用从过往数据中学到的泛化知识，结合新类型数据的少量训练样本，实现快速迁移学习。COMET: Commonsense Transformers for Automatic Knowledge Graph Construction 提出常识 Transformer 架构，将 GPT-2 等语言模型与种子知识图谱相结合，学习其结构和关系，根据图表征形成语言模型，从而生成新的知识并将它们添加到种子图中。

- 知识图谱的知识补全和可解释推理

传统的表示学习缺乏可解释性，知识图谱推理越来越受到关注，其中既有使用强化学习方法寻找路径的方法，也有使用实体邻居和注意力权重做可解释性推理方法。Multi-Hop Knowledge Graph Reasoning with Reward Shaping 是基于多跳推理的知识库问答方法，基于强化学习扩展在知识图谱的推理路径，以获得问题的正确答案。Learning Attention-based Embeddings for Relation Prediction in Knowledge Graphs 提出一种基于注意力机制的特征嵌入方法，获取实体邻近范围内的实体和关系特征，引入关系聚类和多跳关系，有效提升了基于多跳推理的知识图谱补全的效果。Iteratively Learning Embeddings and Rules for Knowledge Graph

Reasoning 研究如何迭代地进行知识表示学习和规则学习，提出的 IterE 模型可以利用学习的规则改进稀疏实体的表示学习，进而提升规则学习和链接预测效果。

- 基于知识图谱的推荐和对话问答

将知识图谱作为辅助信息引入到推荐系统中可以有效地解决传统推荐系统存在的稀疏性和冷启动问题，近几年吸引大量研究人员在相关工作。随着图卷积神经网络，图注意力机制等技术的逐渐兴起，基于图表示学习的推荐模型达到了更高的表现效果，并为推荐系统的可解释性提供了帮助。KGAT: Knowledge Graph Attention Network for Recommendation 利用知识图谱中商品之间的关系，训练了一个端到端的含注意力机制的模型，用于提高推荐系统的能力。AKUPM: Attention-Enhanced Knowledge-Aware User Preference Model for Recommendation 使用注意力模型，利用知识图谱对用户进行建模，显著提升了推荐系统的效果。Reinforcement Knowledge Graph Reasoning for Explainable Recommendation 结合强化学习的框架和知识图谱推理来提供对推荐结果的解释。在对话问答方面，以前对话生成的信息源是文本与对话记录，但如果遇到词表之外的（Out-of-Vocabulary）的词，模型往往难以生成合适的、有信息量的回复，而会产生一些低质量的、模棱两可的回复。Commonsense Knowledge Aware Conversation Generation with Graph 提出一种基于常识知识图谱的对话模型 CCM 来理解对话，产生信息丰富且合适的回复。

5 自然语言处理

5.1 自然语言处理概念

自然语言是指汉语、英语、法语等人们日常使用的语言，是人类社会发展演变而来的语言，而不是人造的语言，它是人类学习生活的重要工具。概括说来，自然语言是指人类社会约定俗成的，区别于如程序设计的语言的人工语言。在整个人类历史上以语言文字形式记载和流传的知识占到知识总量的 80%以上。就计算机应用而言，据统计，用于数学计算的仅占 10%，用于过程控制的不到 5%，其余 85%左右都是用于语言文字的信息处理。

处理包含理解、转化、生成等过程。自然语言处理，是指用计算机对自然语言的形、音、义等信息进行处理，即对字、词、句、篇章的输入、输出、识别、分析、理解、生成等的操作和加工。实现人机间的信息交流，是人工智能、计算机科学和语言学所共同关注的重要问题。自然语言处理的具体表现形式包括机器翻译、文本摘要、文本分类、文本校对、信息抽取、语音合成、语音识别等。可以说，自然语言处理就是要计算机理解自然语言，自然语言处理机制涉及两个流程，包括自然语言理解和自然语言生成。自然语言理解是指计算机能够理解自然语言文本的意义，自然语言生成则是指能以自然语言文本来表达给定的意图。



图 5-1 自然语言理解层次

自然语言的理解和分析是一个层次化的过程，许多语言学家把这一过程分为五个层次，可以更好地体现语言本身的构成，五个层次分别是语音分析、词法分析、句法分析、语义分析和语用分析。

- 语音分析是要根据音位规则，从语音流中区分出一个个独立的音素，再根据音位形态规则找出音节及其对应的词素或词。
- 词法分析是找出词汇的各个词素，从中获得语言学的信息。

- 句法分析是对句子和短语的结构进行分析，目的是要找出词、短语等的相互关系以及各自在句中的作用。
- 语义分析是找出词义、结构意义及其结合意义，从而确定语言所表达的真正含义或概念。
- 语用分析是研究语言所存在的外界环境对语言使用者所产生的影响。

在人工智能领域或者是语音信息处理领域中，学者们普遍认为采用图灵试验可以判断计算机是否理解了某种自然语言，具体的判别标准有以下几条：

第一， 问答，机器人能正确回答输入文本中的有关问题；

第二， 文摘生成，机器有能力生成输入文本的摘要；

第三， 释义，机器能用不同的词语和句型来复述其输入的文本；

第四， 翻译，机器具有把一种语言翻译成另一种语言的能力。

5.2 自然语言的理解发展历史

自然语言处理是包括了计算机科学、语言学心理认知学等一系列学科的一门交叉学科，这些学科性质不同但又彼此相互交叉。因此，梳理自然语言处理的发展历程对于我们更好地了解自然语言处理这一学科有着重要的意义。

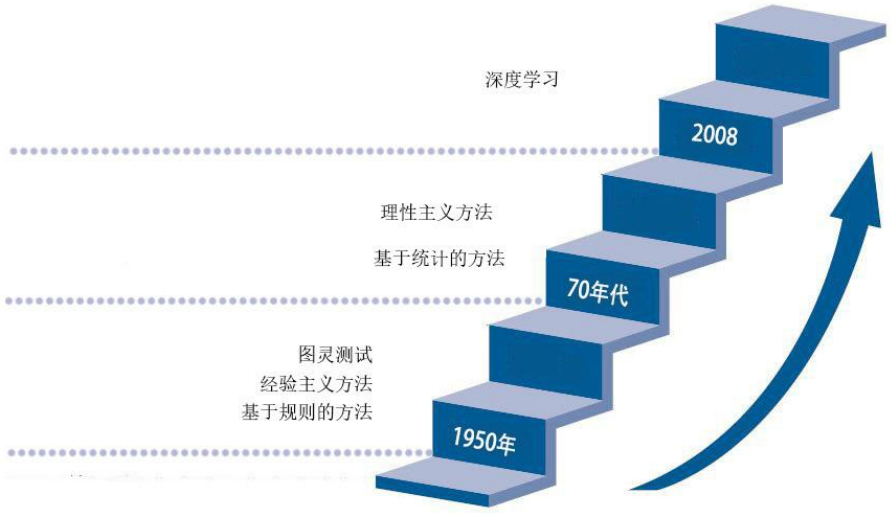


图 5-2 自然语言发展历程

1950 年图灵提出了著名的“图灵测试”，这一般被认为是自然语言处理思想的开端，20 世纪 50 年代到 70 年代自然语言处理主要采用基于规则的方法，研究人员们认为自然语言处理的过程和人类学习认知一门语言的过程是类似的，所以大量的研究员基于这个观点来进行研究，这时的自然语言处理停留在理性主义思潮阶段，以基于规则的方法为代表。但是基于规则的方法具有不可避免的缺点，首先规则不可能覆盖所有语句，其次这种方法对开发者的要求极高，开发者不仅要精通计算机还要精通语言学，因此，这一阶段虽然解决了一些简单的问题，但是无法从根本上将自然语言理解实用化。

70 年代以后随着互联网的高速发展，丰富的语料库成为现实以及硬件不断更新完善，自然语言处理思潮由经验主义向理性主义过渡，基于统计的方法逐渐代替了基于规则的方法。贾里尼克和他领导的 IBM 华生实验室是推动这一转变的关键，他们采用基于统计的方法，将当时的语音识别率从 70%提升到 90%。在这一阶段，自然语言处理基于数学模型和统计的方法取得了实质性的突破，从实验室走向实际应用。

从 2008 年到现在，在图像识别和语音识别领域的成果激励下，人们也逐渐开始引入深度学习来做自然语言处理研究，由最初的词向量到 2013 年的 word2vec，将深度学习与自然语言处理的结合推向了高潮，并在机器翻译、问答系统、阅读理解等领域取得了一定成功。深度学习是一个多层的神经网络，从输入层开始经过逐层非线性的变化得到输出。从输入到输出做端到端的训练。把输入到输出对的数据准备好，设计并训练一个神经网络，即可执行预想的任务。RNN 已经是自然语言处理最常用的方法之一，GRU、LSTM 等模型相继引发了一轮又一轮的热潮。

近年自然语言处理在词向量（word embedding）表示、文本的（编码）encoder 和 decoder（反编码）技术以及大规模预训练模型（pre-trained）上的方法极大地促进了自然语言处理的研究[7]。

5.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为自然语言处理领域全球学者分布情况：



图 5-3 自然语言处理领域全球人才分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布在其东西海岸；欧洲也有较多的人才分布，主要集中在欧洲中西部；亚洲的人才主要分布在我国东部及日韩地区；其他诸如非洲、南美洲等地区的学者非常稀少；自然语言处理领域的人才分布与各地区的科技、经济实力情况大体一致。此外，在性别比例方面，自然语言处理领域中男性学者占比 89.3%，女性学者占比 10.7%，男性学者占比远高于女性学者。

自然语言处理领域学者的 h-index 分布如下图所示，分布情况大体呈阶梯状，大部分学者的 h-index 分布在中低区域，其中 h-index 在小于 20 区间的人数最多，有 929 人，占比 45.3%，50-60 区间的人数最少，有 98 人。

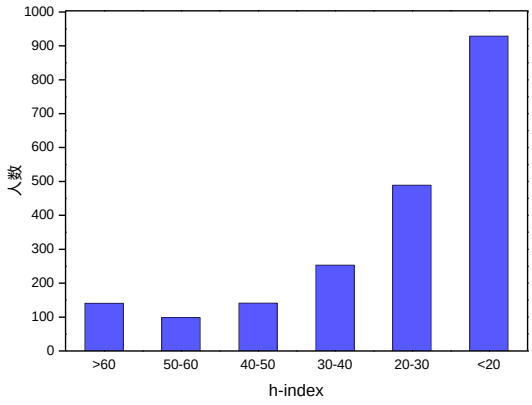


图 5-4 自然语言处理领域学者 h-index 分布

- 中国人才分布

我国专家学者在自然语言处理领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等亚洲国家相比，中国在自然语言处理领域学者数量较多。



图 5-5 自然语言处理领域中国学者分布

中国与其他国家在自然语言处理领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 5-1 自然语言处理领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	250	7589	30	472
中国-爱尔兰	41	826	20	34
中国-新加坡	37	1537	42	77
中国-英国	34	1223	36	61
中国-日本	24	513	21	41

中国-印度	23	1368	59	32
中国-加拿大	19	307	16	35
中国-德国	15	124	8	30
中国-澳大利亚	8	120	15	20
中国-柬埔寨	8	101	13	9

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在自然语言处理领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 3 席；中国与印度合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中印合作达到了较高的水平。

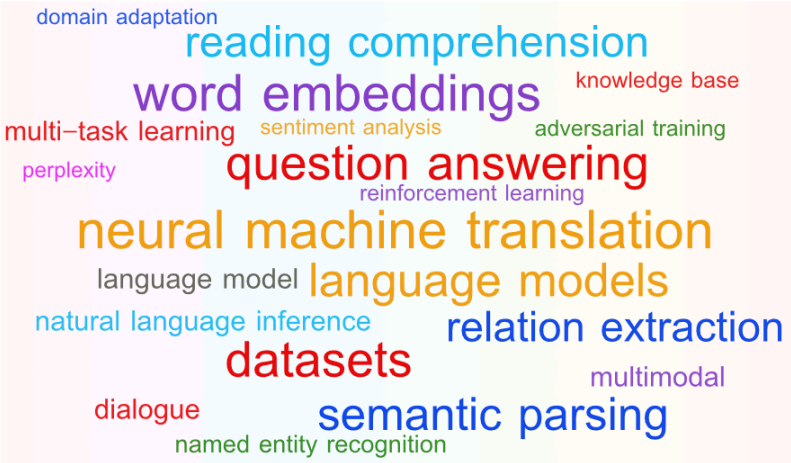
5.4 论文解读

本节对本领域的高水平学术会议论文进行挖掘，解读这些会议在 2018-2019 年的部分代表性工作。会议具体包括：

Annual Meeting of the Association for Computational Linguistics

Conference on Empirical Methods in Natural Language Processing

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，神经机器翻译（neural machine translation）、词嵌入（word embeddings）、数据集（datasets）是本领域中最热的关键词。



论文题目：*BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*

中文题目：BERT: 语言理解的深层双向转换器的预训练

论文作者：Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova

论文出处：In Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics.

论文地址：https://arxiv.org/abs/1810.04805

研究问题：

文章介绍一种新的语言表示模型 BERT（Bidirectional Encoder Representations from Transformers），通过联合上下文信息从未标记文本中预训练深层双向表示形式，只需一个额外的输出层，就可以对预训练模型进行调整，在不需要对特定任务的体系结构进行大量修改的前提下，在多种语言相关任务上获得。

研究内容：

模型包含预训练和微调两个步骤：在预训练阶段，对不同训练任务的未标记数据进行训练。在微调阶段，首先用预训练参数初始化 BERT 模型，然后，使用来自下游任务的标记数据对预训练的参数进行微调。

BERT 是一个多层的双向 Transformer，输入包括三个部分，分别为词向量、单词所属句子向量和单词的位置向量，形象的表示如下图所示，其中[CLS]和[SEP]是放在每个输入最前和用户分隔句子的特殊符号。

文章提出两种无监督任务来预训练 BERT，分别是屏蔽语言模型（Masked Language Model, MLM）和下句预测模型（Next Sentence Prediction, NSP）：MLM 通过屏蔽一句话中部分词然后让模型来预测屏蔽词来训练模型。在实验设置中，大约 15%的词被随机屏蔽。但是这样的训练方法也有缺陷，屏蔽词相当于从数据集中抹去，且可能预训练阶段与微调阶段不一致。因此，对于屏蔽词有如下三种处理方式：80%用[MASK]替换，10%用随机的词语替换，另外 10%不做改变。NSP 任务是为了增强模型对句子间关系的理解能力，训练时选择的句对 A、B 中，

B 有 50%的概率真的是 A 的下一句，50%的概率不是 A 的下一句。预训练语料使用 BooksCorpus 和英语维基百科的文本段落。

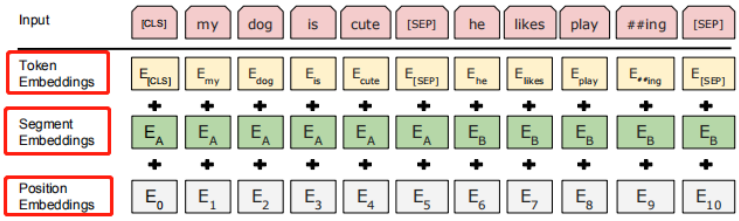


Figure 2: BERT input representation. The input embeddings are the sum of the token embeddings, the segmentation embeddings and the position embeddings.

模型微调测试了 11 个自然语言处理任务上的效果，包括 General Language Understanding Evaluation(GLUE)基准测试集中的 8 项评测、SQuAD 1.1 和 SQuAD 2.0 两个阅读理解数据集和 Situations With Adversarial Generations (SWAG)数据集。BERT 均稳定优于基线方法，下表展示了 GLUE 上的对比结果。

System	MNLI-(m/mm) 392k	QQP 363k	QNLI 108k	SST-2 67k	CoLA 8.5k	STS-B 5.7k	MRPC 3.5k	RTE 2.5k	**Average** -
Pre-OpenAI SOTA	80.6/80.1	66.1	82.3	93.2	35.0	81.0	86.0	61.7	74.0
BiLSTM+ELMo+Attn	76.4/76.1	64.8	79.8	90.4	36.0	73.3	84.9	56.8	71.0
OpenAI GPT	82.1/81.4	70.3	87.4	91.3	45.4	80.0	82.3	56.0	75.1
$BERT_{BASE}$	84.6/83.4	71.2	90.5	93.5	52.1	85.8	88.9	66.4	79.6
$BERT_{LARGE}$	**86.7/85.9**	**72.1**	**92.7**	**94.9**	**60.5**	**86.5**	**89.3**	**70.1**	**82.1**

Table 1: GLUE Test results, scored by the evaluation server (https://gluebenchmark.com/leaderboard). The number below each task denotes the number of training examples. The "Average" column is slightly different than the official GLUE score, since we exclude the problematic WNLI set.[8] BERT and OpenAI GPT are single-model, single task. F1 scores are reported for QQP and MRPC, Spearman correlations are reported for STS-B, and accuracy scores are reported for the other tasks. We exclude entries that use BERT as one of their components.

研究结论：

文章提出的 BERT 模型在 11 项自然语言处理任务上取得了最先进的效果。由语言模型转移学习带来的模型效果改进表明，丰富的、无监督的预训练是许多语言理解系统的组成部分。特别地，即使是资源匮乏的任务也可以从深层的单向架构中获益。文章主要贡献是进一步将这些发现推广到深层的双向架构，允许相同的预训练模型成功地应用于广泛的 NLP 任务。

论文题目：*Semi-Supervised Learning for Neural Machine Translation*

中文题目：神经机器翻译的半监督学习机制

论文作者：Yong Cheng, Wei Xu, Zhongjun He, Wei He, Hua Wu, Maosong Sun and Yang Liu

论文出处：Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics

论文地址：https://link.springer.com/chapter/10.1007/978-981-32-9748-7_3

研究问题：

近年来，端到端神经机器翻译(neural machine translation, NMT)取得了显著的进展，但 NMT 系统仅依靠并行语料库进行参数估计。由于平行语料库在数量、质量和覆盖范围等方面都存在一定的局限性，尤其是对资源相对较少的语言而言。所以利用单语语料库来提高网络机器翻译的性能就变得很有吸引力了。文章就提出了一种半监督的方法来训练 NMT 模型。其核心思想是使用一个自编码器重建单语语料库，其中源-目标和目标-源转换模型分别充当编码器和解码器。该方法不仅可以利用目标语的单语语料库，而且还可以利用源语的单语语料库。

研究内容：

首先，将观察到的目标句编码为潜在的源句（图中蓝色箭头的过程）。然后，使用源到源的翻译模型，对源句进行译码（图中黄色箭头的过程），利用源到目标的模型重构所观察到的目标句。

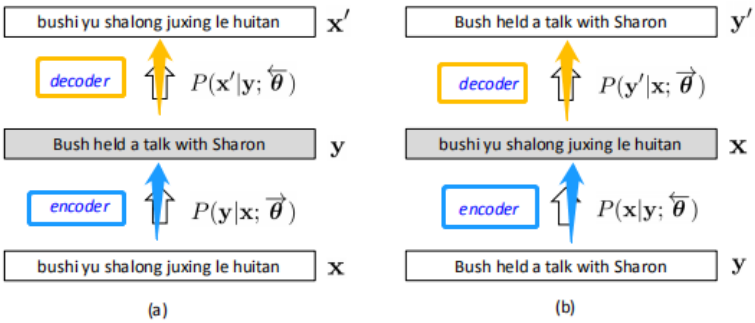


Figure 1: Examples of (a) source autoencoder and (b) target autoencoder on monolingual corpora. Our idea is to leverage autoencoders to exploit monolingual corpora for NMT. In a source autoencoder, the source-to-target model $P(\mathbf{y}|\mathbf{x};\overrightarrow{\boldsymbol{\theta}})$ serves as an encoder to transform the observed source sentence $\mathbf{x}$ into a latent target sentence $\mathbf{y}$ (highlighted in grey), from which the target-to-source model $P(\mathbf{x}'|\mathbf{y};\overleftarrow{\boldsymbol{\theta}})$ reconstructs a copy of the observed source sentence $\mathbf{x}'$ from the latent target sentence. As a result, monolingual corpora can be combined with parallel corpora to train bidirectional NMT models in a semi-supervised setting.

目标编码器公式如下：

$$\begin{aligned} & P(\mathbf{y}'|\mathbf{y};\overrightarrow{\boldsymbol{\theta}},\overleftarrow{\boldsymbol{\theta}}) \\ = & \sum_{\mathbf{x}} P(\mathbf{y}',\mathbf{x}|\mathbf{y};\overrightarrow{\boldsymbol{\theta}},\overleftarrow{\boldsymbol{\theta}}) \\ = & \sum_{\mathbf{x}} \underbrace{P(\mathbf{x}|\mathbf{y};\overleftarrow{\boldsymbol{\theta}})}_{encoder} \underbrace{P(\mathbf{y}'|\mathbf{x};\overrightarrow{\boldsymbol{\theta}})}_{decoder}, \end{aligned}$$

x 是潜在的原句，y 是目标句，y'是要重构的目标句的副本。同样的，源句的编码器公式如下：

$$\begin{aligned} & P(\mathbf{x}'|\mathbf{x};\overrightarrow{\boldsymbol{\theta}},\overleftarrow{\boldsymbol{\theta}}) \\ = & \sum_{\mathbf{y}} P(\mathbf{x}',\mathbf{y}|\mathbf{x};\overleftarrow{\boldsymbol{\theta}}) \\ = & \sum_{\mathbf{y}} \underbrace{P(\mathbf{y}|\mathbf{x};\overrightarrow{\boldsymbol{\theta}})}_{encoder} \underbrace{P(\mathbf{x}'|\mathbf{y};\overleftarrow{\boldsymbol{\theta}})}_{decoder}. \end{aligned}$$

由于自动编码器同时涉及到源到目标和目标到源的模型，所以很自然地要将并行语料库和单语语料库结合起来，在半监督的环境中学习双向 NMT 翻译模型。

平行语料库：$D = \{(x^n, y^n)\}_{n=1}^{N}$

目标语言的单语料库：$T = \{y^t\}_{n=1}^{T}$

源语言的单语料库：$S = \{x^s\}_{s=1}^{S}$

训练的目标函数为：

$$\begin{aligned} & J(\overrightarrow{\boldsymbol{\theta}},\overleftarrow{\boldsymbol{\theta}}) \\ = & \underbrace{\sum_{n=1}^{N} \log P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)};\overrightarrow{\boldsymbol{\theta}})}_{source\text{-}to\text{-}target\ likelihood} \\ & + \underbrace{\sum_{n=1}^{N} \log P(\mathbf{x}^{(n)}|\mathbf{y}^{(n)};\overleftarrow{\boldsymbol{\theta}})}_{target\text{-}to\text{-}source\ likelihood} \\ & + \lambda_1 \underbrace{\sum_{t=1}^{T} \log P(\mathbf{y}'|\mathbf{y}^{(t)};\overrightarrow{\boldsymbol{\theta}},\overleftarrow{\boldsymbol{\theta}})}_{target\ autoencoder} \\ & + \lambda_2 \underbrace{\sum_{s=1}^{S} \log P(\mathbf{x}'|\mathbf{x}^{(s)};\overrightarrow{\boldsymbol{\theta}},\overleftarrow{\boldsymbol{\theta}})}_{source\ autoencoder}, \end{aligned}$$

其中，λ_1和λ_2是超参数。很明显，源到目标和目标到源的模型是通过自动编码器连接起来的，并且有望在联合训练中相互受益。

模型用小批量随机梯度下降算法进行优化。$\vec{\theta}$的导数如下所示：

$$\begin{aligned}&\frac{\partial J(\overrightarrow{\boldsymbol{\theta}}, \overleftarrow{\boldsymbol{\theta}})}{\partial \overrightarrow{\boldsymbol{\theta}}}\\ =&\sum_{n=1}^{N} \frac{\partial \log P(\mathbf{y}^{(n)}|\mathbf{x}^{(n)}; \overrightarrow{\boldsymbol{\theta}})}{\partial \overrightarrow{\boldsymbol{\theta}}}\\ &+\lambda_1 \sum_{t=1}^{T} \frac{\partial \log P(\mathbf{y}'|\mathbf{y}^{(t)}; \overrightarrow{\boldsymbol{\theta}}, \overleftarrow{\boldsymbol{\theta}})}{\partial \overrightarrow{\boldsymbol{\theta}}}\\ &+\lambda_2 \sum_{s=1}^{S} \frac{\partial \log P(\mathbf{x}'|\mathbf{x}^{(s)}; \overrightarrow{\boldsymbol{\theta}}, \overleftarrow{\boldsymbol{\theta}})}{\partial \overrightarrow{\boldsymbol{\theta}}}.\end{aligned}$$

由于第二部分和第三部分依赖于搜索空间难以计算，文章又提出了一种近似计算的方法，缩小 y 的搜索空间。用$\tilde{x}$（y）近似替代x（y），即用x（y）候选翻译的前 top-k 个作为$\tilde{x}$（y）。

用文章的方法和最先进的 SMT 和 NMT 方法进行比较，实验结果图如下：

System	Training Data CE	C	E	Direction	NIST06	NIST02	NIST03	NIST04	NIST05
MOSES	√	×	×	C → E	32.48	32.69	32.39	33.62	30.23
				E → C	14.27	18.28	15.36	13.96	14.11
	√	×	√	C → E	34.59	35.21	35.71	35.56	33.74
	√	√	×	E → C	20.69	25.85	19.76	18.77	19.74
RNNSEARCH	√	×	×	C →E	30.74	35.16	33.75	34.63	31.74
				E →C	15.71	20.76	16.56	16.85	15.14
	√	×	√	C → E	35.61^{**++}	38.78^{**++}	38.32^{**++}	38.49^{**++}	36.45^{**++}
				E →C	17.59^{++}	23.99^{++}	18.95^{++}	18.85^{++}	17.91^{++}
	√	√	×	C →E	35.01^{++}	38.20^{**++}	37.99^{**++}	38.16^{**++}	36.07^{**++}
				E →C	21.12^{*++}	29.52^{**++}	20.49^{**++}	21.59^{**++}	19.97^{++}

Table 2: Comparison with MOSES and RNNSEARCH. MOSES is a phrase-based statistical machine translation system (Koehn et al., 2007). RNNSEARCH is an attention-based neural machine translation system (Bahdanau et al., 2015). "CE" donates Chinese-English parallel corpus, "C" donates Chinese monolingual corpus, and "E" donates English monolingual corpus. "√" means the corpus is included in the training data and × means not included. "NIST06" is the validation set and "NIST02-05" are test sets. The BLEU scores are case-insensitive. "*": significantly better than MOSES ($p < 0.05$); "**": significantly better than MOSES ($p < 0.01$);"+": significantly better than RNNSEARCH ($p < 0.05$); "++": significantly better than RNNSEARCH ($p < 0.01$).

Method	Training Data CE	C	E	Direction	NIST06	NIST02	NIST03	NIST04	NIST05
Sennrich et al. (2015)	√	×	√	C →E	34.10	36.95	36.80	37.99	35.33
	√	√	×	E →C	19.85	28.83	20.61	20.54	19.17
this work	√	×	√	C →E	35.61^{**}	38.78^{**}	38.32^{**}	38.49^{*}	36.45^{**}
				E →C	17.59	23.99	18.95	18.85	17.91
	√	√	×	C →E	35.01^{**}	38.20^{**}	37.99^{**}	38.16	36.07^{**}
				E →C	21.12^{**}	29.52^{**}	20.49	21.59^{**}	19.97^{**}

Table 3: Comparison with Sennrich et al. (2015). Both Sennrich et al. (2015) and our approach build on top of RNNSEARCH to exploit monolingual corpora. The BLEU scores are case-insensitive. "*": significantly better than Sennrich et al. (2015) ($p < 0.05$); "**": significantly better than Sennrich et al. (2015) ($p < 0.01$).

研究结果：

文章提出了一种训练神经机器翻译模型的半监督方法。其核心思想是在单语语料库上引入自动编码器，采用源对目标和目标对源的翻译模型作为编码器和译

码器。在汉英 NIST 数据集上的实验表明，与最先进的 SMT 和 NMT 方法进行，该方法带来了显著的改善。

论文题目：*Know What You Don't Know: Unanswerable Questions for SquAD*

中文题目：知道你所不知道的：针对 SquAD 中不可回答的问题

论文作者：Pranav Rajpurkar, Robin Jia, Percy Liang

论文出处：Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. 2018

论文地址：https://arxiv.org/abs/1806.03822

研究问题：

阅读理解系统（模型）通常可以在上下文文档中找到问题的正确答案，但对于没有在上下文中说明正确答案的问题，它们给出的答案也不那么可靠。现有的数据集要么只关注可回答的问题，要么使用容易识别的自动生成的不可回答的问题作为数据集。为了弥补这些不足，文章介绍了斯坦福问答数据集(SQuAD)的最新版本——SQuAD 2.0，它整合了现有的 SQuAD 中可回答的问题和 50000 多个由大众工作者编写的难以回答的问题，其中那些难以回答的问题与可回答的问题题目设置相似。为了在 SQuAD 2.0 中表现的更好，系统不仅要在可能的情况下回答问题，还要确定什么时候段落的上下文不支持回答，并且避免回答问题。SQuAD 2.0 数据集是自然语言理解任务中对现有模型的一个挑战。

研究方法：

数据集：在 Daemo 平台上雇佣了众包工作人员来编写无法回答的问题。每个任务由来自 SQuAD 1.1 的一整篇文章组成。对于文章中的每个段落，工作人员最多可提出 5 个仅凭段落是无法回答的问题，同时还要参考段落中出现的实体并给出一个合理的答案。同时给工作人员展示 SQuAD 1.1 中的问题作为参考，尽量使难以回答的那些问题与可回答的问题相似。

文章评估了三种现有的模型架构在两个数据集上的表现，让这些模型不仅去学习答案的分布，而且也去预测一个问题是不可回答问题的概率。当模型预测某

个问题无法回答的概率超过某个阈值时，模型就放弃学习答案分布。下表展示了三个模型在两个数据集（SQuAD 1.1 和 SQuAD 2.0）上的表现，结果显示：

表现最好的模型（DocQA + ELMo）在 SQuAD 2.0 上与人类仍有 23.2 的差距，意味着模型有很大的改进空间；

在两个数据集上运用相同模型架构，相比于 SQuAD1.1，最优模型和人的 F1 值差距在 SQuAD 2.0 上更大，说明对现有模型来说 SQuAD 2.0 是一个更难学习的数据。

System	SQuAD 1.1 test		SQuAD 2.0 dev		SQuAD 2.0 test	
	EM	F1	EM	F1	EM	F1
BNA	68.0	77.3	59.8	62.6	59.2	62.1
DocQA	72.1	81.0	61.9	64.8	59.3	62.3
DocQA + ELMo	**78.6**	**85.8**	**65.1**	**67.6**	**63.4**	**66.3**
Human	82.3	91.2	86.3	89.0	86.9	89.5
Human–Machine Gap	3.7	5.4	**21.2**	**21.4**	**23.5**	**23.2**

文章在 SQuAD 1.1 数据集上利用 TFIDF 和规则随机生成了一些难以回答的问题，仍采用相同的模型进行对比。结果显示（如下表）最好的模型还是在 SQuAD 2.0 数据集上表现最低，再次证明了 SQuAD 2.0 对现有的语言理解模型来说是一个有难度的挑战。

System	SQuAD 1.1 + TFIDF		SQuAD 1.1 + RULEBASED		SQuAD 2.0 dev	
	EM	F1	EM	F1	EM	F1
BNA	72.7	76.6	80.1	84.8	59.8	62.6
DocQA	75.6	79.2	80.8	84.8	61.9	64.8
DocQA + ELMo	**79.4**	**83.0**	**85.7**	**89.6**	**65.1**	**67.6**

研究结果：

文章证明了 SQuAD 2.0 是一个具有挑战性的、多样化的、大规模的数据集，它迫使模型去学习什么情况下一个问题在给定的环境中是无法回答的。我们有理由相信，SQuAD 2.0 将会促进新的阅读理解模型的发展，这些模型能够知道他们不知道的东西是什么，从而能在更深层次上理解语言文字。

论文题目：*GLUE: A MULTI-TASK BENCHMARK AND ANALYSIS PLATFORM FOR NATURAL LANGUAGE UNDERSTANDING*

中文题目：GLUE: 一个用于自然语言理解的多任务基准测试和分析平台

论文作者：Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill，Omer Levy& Samuel R. Bowman

论文出处：Proceedings of the International Conference on Learning Representations (ICLR). 2019

论文地址：https://arxiv.org/abs/1804.07461

研究问题：

人类理解语言的能力是灵活的、强健的。相比之下，单词级以上的大多数自然语言理解（Natural Language Understanding, NLU）模型都是为特定的任务设计的。我们期望开发一个能够学习在不同领域执行一系列不同语言任务的更统一的模型，它必须能够以一种不局限于单个任务、类型或数据集的方式来理解人类的语言。为了实现这一目标，文章设计了一个通用语言理解评估基准（General Language Understanding Evaluation, GLUE）用于评估模型在不同的现有 NLU 任务集上的性能。

研究方法：

文章设计了一个通用语言理解评估基准（General Language Understanding Evaluation, GLUE)，它包含一组 NLU 任务，包括问答系统、情感分析和文本蕴涵，以及一个用于模型评估、比较和分析的在线平台。GLUE 倾向于让模型在任务之间共享一般的语言知识。GLUE 还提供了一个人工设计的测试集（诊断集），可以对模型进行详细的分析。

Corpus	|Train|	|Test|	Task	Metrics	Domain
Single-Sentence Tasks					
CoLA	8.5k	**1k**	acceptability	Matthews corr.	misc.
SST-2	67k	1.8k	sentiment	acc.	movie reviews
Similarity and Paraphrase Tasks					
MRPC	3.7k	1.7k	paraphrase	acc./F1	news
STS-B	7k	1.4k	sentence similarity	Pearson/Spearman corr.	misc.
QQP	364k	**391k**	paraphrase	acc./F1	social QA questions
Inference Tasks					
MNLI	393k	**20k**	NLI	matched acc./mismatched acc.	misc.
QNLI	105k	5.4k	QA/NLI	acc.	Wikipedia
RTE	2.5k	3k	NLI	acc.	news, Wikipedia
WNLI	634	**146**	coreference/NLI	acc.	fiction books

为了评测这个 GLUE 评估基准，文章在公共数据集上评估了句子表示的 baseline 和最优模型，下表展示了数据集的任务表述和相关的统计信息。数据集的任务表述和统计信息如上表所示。

结果显示如下表所示，针对所有任务的多任务训练比针对每个任务单独训练模型的效果更好。然而，最佳模型的低性能表明模型仍存在改进空间。

		Single Sentence		Similarity and Paraphrase			Natural Language Inference			
Model	**Avg**	**CoLA**	**SST-2**	**MRPC**	**QQP**	**STS-B**	**MNLI**	**QNLI**	**RTE**	**WNLI**
Single-Task Training										
BiLSTM	63.9	15.7	85.9	69.3/79.4	81.7/61.4	66.0/62.8	70.3/70.8	75.7	52.8	**65.1**
+ELMo	66.4	**35.0**	90.2	69.0/80.8	85.7/65.6	64.0/60.2	72.9/73.4	71.7	50.1	**65.1**
+CoVe	64.0	14.5	88.5	73.4/81.4	83.3/59.4	67.2/64.1	64.5/64.8	75.4	53.5	**65.1**
+Attn	63.9	15.7	85.9	68.5/80.3	83.5/62.9	59.3/55.8	74.2/73.8	77.2	51.9	**65.1**
+Attn, ELMo	66.5	**35.0**	90.2	68.8/80.2	**86.5/66.1**	55.5/52.5	**76.9/76.7**	76.7	50.4	**65.1**
+Attn, CoVe	63.2	14.5	88.5	68.6/79.7	84.1/60.1	57.2/53.6	71.6/71.5	74.5	52.7	**65.1**
Multi-Task Training										
BiLSTM	64.2	11.6	82.8	74.3/81.8	84.2/62.5	70.3/67.8	65.4/66.1	74.6	57.4	**65.1**
+ELMo	67.7	32.1	89.3	**78.0/84.7**	82.6/61.1	67.2/67.9	70.3/67.8	75.5	57.4	**65.1**
+CoVe	62.9	18.5	81.9	71.5/78.7	84.9/60.6	64.4/62.7	65.4/65.7	70.8	52.7	**65.1**
+Attn	65.6	18.6	83.0	76.2/83.9	82.4/60.1	72.8/70.5	67.6/68.3	74.3	58.4	**65.1**
+Attn, ELMo	**70.0**	33.6	**90.4**	**78.0**/84.4	84.3/63.1	74.2/72.3	74.1/74.5	**79.8**	58.9	**65.1**
+Attn, CoVe	63.1	8.3	80.7	71.8/80.0	83.4/60.5	69.8/68.4	68.1/68.6	72.9	56.0	**65.1**
Pre-Trained Sentence Representation Models										
CBoW	58.9	0.0	80.0	73.4/81.5	79.1/51.4	61.2/58.7	56.0/56.4	72.1	54.1	**65.1**
Skip-Thought	61.3	0.0	81.8	71.7/80.8	82.2/56.4	71.8/69.7	62.9/62.8	72.9	53.1	**65.1**
InferSent	63.9	4.5	85.1	74.1/81.2	81.7/59.1	75.9/75.3	66.1/65.7	72.7	58.0	**65.1**
DisSent	62.0	4.9	83.7	74.1/81.7	82.6/59.5	66.1/64.8	58.7/59.1	73.9	56.4	**65.1**
GenSen	66.2	7.7	83.1	76.6/83.0	82.9/59.8	**79.3/79.2**	71.4/71.3	78.6	**59.2**	**65.1**

研究结果：

首先，文章实现了一个通用语言理解评估基准（GLUE 基准），包含 9 个句子或句对的 NLU 任务。所有任务建立在带标注的数据集上，数据集覆盖了各种文本类型、不同数据规模和不同难度系数。其次，建立了一个主要基于私有评测数据的在线模型评估平台。该平台与模型无关，并且可以评估任何能够在所有 9 个任务上产生结果的模型。然后，文章还构建了专门的诊断评价数据集，以用作误差分析、模型的定性比较以及对抗性数据的补充。最后，文章实验了句子表示学习的几种主要现有方法的结果。

论文题目：*Linguistically-Informed Self-Attention for Semantic Role Labeling*

中文题目：用于语义角色标注的基于语言学信息的自我注意力方法

论文作者：Emma Strubell, Patrick Verga, Daniel Andor, David Weissand Andrew McCallum

论文出处：Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing

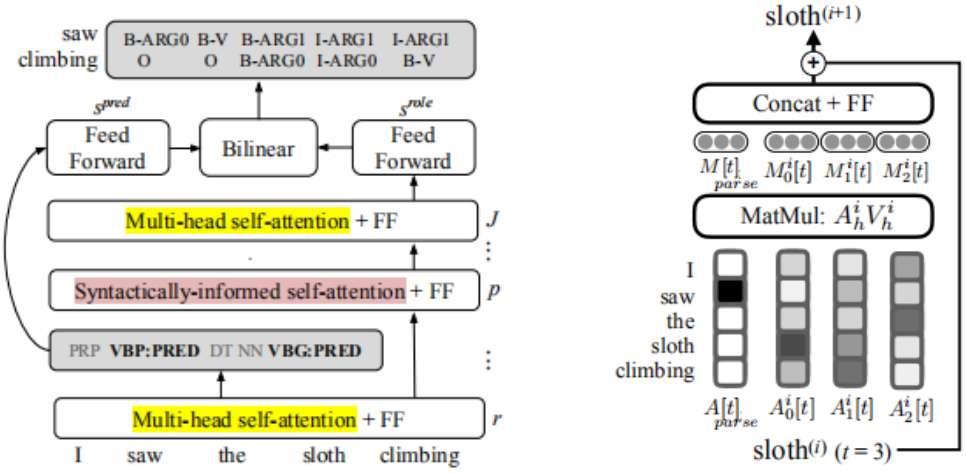
论文地址：https://www.aclweb.org/anthology/D18-1548/

研究问题：

语义角色标注（Semantic Role Labeling, SRL）是一种提取文本高级表示的技术。目前最先进的基于深度神经网络的语义角色标注模型并没有显式使用文本的语言特征。然而，有研究已经指出，语法树可以有效改善 SRL 任务。鉴于此，文章提出了一种用于语义角色标注的基于语言学的自注意力方法（linguistically-informed self-attention，LISA）。该模型将多头自注意力机制（multi-head self-attention）与多任务学习相结合，包括句法依赖解析、词性标注、谓词检测和语义角色标记。与先前需要大量预处理来准备语言特征的模型不同，LISA 可以仅使用原始的 token 对序列进行一次编码，来同时执行多个预测任务。

研究方法：

文章设计了一个高效的利用语言信息有效执行端到端语义角色标注任务的神经网络模型。该模型融合了神经网络的注意力机制预测句法依赖关系，并在 4 个相关任务上进行了多任务学习。下图（左）是模型结构：词向量输入到具有多头注意力机制的 J 层，在 p 层训练自注意力机制去关注父节点的语义信息。P 层的详细操作见下图（右）。



具体地，模型的基础是一个多头自注意力机制的 token 编码，基于 ELMo 预训练得到的 token 表示结合一个由正弦函数确定的位置编码向量学习 H 个注意头，拼接后组成每个 token 的最终自注意力表示。句法信息的自注意力机制将注

意力中的 key、value 和 query 均增加语义解析信息，其中 key、value 分别对应父节点和依赖关系。最后，共享模型中较低层的参数来预测 POS 词性标记和谓词。

文章把 LISA 模型和四个先进的基线方法比较，下表的结果显示，未加入语义信息的模型已经实现了现有的最优模型性能（如红框所示），当 LISA 加入自己的语义解析时模型性能并没有很大的提升（如绿框所示），但当再加入了目前最优的 D&M 语义解析以后，模型性能有了较大的提升。Gold 表示加入了最优的语义解析，以期模型能有更好的表现。

	Dev			WSJ Test			Brown Test		
GloVe	P	R	F1	P	R	F1	P	R	F1
He et al. (2017) PoE	81.8	81.2	81.5	82.0	83.4	82.7	69.7	70.5	70.1
He et al. (2018)	81.3	81.9	81.6	81.2	83.9	82.5	69.7	71.9	70.8
SA	83.52	81.28	82.39	84.17	83.28	83.72	72.98	70.1	71.51
LISA	83.1	81.39	82.24	84.07	83.16	83.61	73.32	70.56	71.91
+D&M	**84.59**	**82.59**	**83.58**	**85.53**	**84.45**	**84.99**	**75.8**	**73.54**	**74.66**
+Gold	*87.91*	*85.73*	*86.81*	—	—	—	—	—	—
ELMo									
He et al. (2018)	84.9	**85.7**	85.3	84.8	**87.2**	86.0	73.9	**78.4**	76.1
SA	85.78	84.74	85.26	86.21	85.98	86.09	77.1	75.61	76.35
LISA	**86.07**	84.64	**85.35**	86.69	86.42	86.55	78.95	77.17	78.05
+D&M	85.83	84.51	85.17	**87.13**	86.67	**86.90**	**79.02**	77.49	**78.25**
+Gold	*88.51*	*86.77*	*87.63*	—	—	—	—	—	—

Table 1: Precision, recall and F1 on the CoNLL-2005 development and test sets.

研究结果：

文章提出了一种多任务神经网络模型，该模型有效地融合了丰富的语言信息用于语义角色标注。通过一系列实验证明了 LISA 的性能优于最先进的现有模型。具体实验结果：在 CoNLL-2005SRL 数据集上，LISA 模型在谓词预测、词嵌入任务上比当前最好的算法在 F1 值上高出了 2.5（新闻专线数据）和 3.5 以上（其他领域数据），减少了约 10%的错误。在 ConLL-2012 英文角色标记任务上，该方法也获得了 2.5F1 值的提升。LISA 同时也比当前最好的基于上下文的词表示学习方法（ELMo）高出了 1.0 的 F1（新闻专线数据）和多于 2.0 的 F1 值（其他领域数据）。

论文题目：*OpenKiwi: An Open Source Framework for Quality Estimation*

中文题目: OpenKiwi: 一个用于质量评估的开源框架

论文作者：Fabio Kepler、Jonay Trenous、Marcos Treviso、Miguel Vera、Andre F. T. Martins

论文出处：Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019

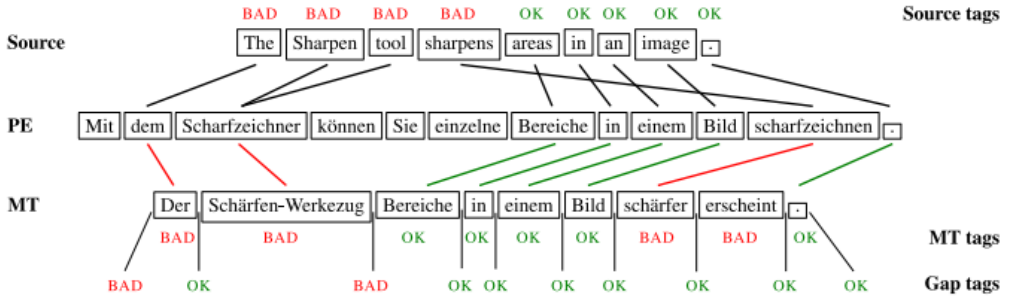
论文地址：https://arxiv.org/abs/1902.08646

研究问题：

文章介绍了一个基于 PyTorch 的用于翻译质量评估的开源框架——OpenKiwi。该框架支持单词级和句子级的质量评估系统的训练和测试，实现和集成了 WMT 2015-18 质量评估比赛的获奖系统。文章在 WMT2018 的两个数据集（English-German SMT and NMT））上对 OpenKiwi 进行基准测试。实验结果证明了，该框架在单词级任务上实现了最先进的性能，在句子级任务上实现了几乎最先进的性能。

研究方法：

质量评估(Quality Estimation，QE)提供了机器翻译和人工翻译之间缺失的一环，目标是在没有参考译文的情况下评估翻译系统的质量。句子级的质量评估旨在预测整个翻译句子的质量，如基于人后期编辑所花费的时间，或者修改它需要的编辑操作数。单词级别质量评估的目标是给机器翻译的每个词、单词之间的间隙（根据上下文需要插入的）和源语言单词(原句中被错译或省略的单词)分配质量标签。下图是一个单词级标注示例。



文章研发的 OpenKiwi（https://github. com/Unbabel/OpenKiwi）实现和集成了 WMT 2015-18 质量评估比赛的获奖系统，且允许轻松地添加和运行新模型，而不需要过多地关注输入数据处理、输出生成和评估。OpenKiwi 基于 PyTorch 深度学习框架实现，可以单独运行或通过 API 集成到其他项目。此外，OpenKiwi 提供了根据 WMT2018 数据的预训练模型，并支持根据新数据训练新 QE 模型。

文章对该框架进行了基准测试，使用了 WMT 2018 质量评估比赛的数据集，结果显示，这些系统的集成版本表现最好，堆叠的架构在预测单词级标签方面非常有效。文章还比较了另一个现有的开源工具 deepQuest，在单词级和句子级均获得更优的表现。

研究结果：

文章介绍了一个新的机器翻译质量评估（QE）开源框架—OpenKiwi。OpenKiwi 是在 PyTorch 中实现的，并支持在新数据上训练单词级和句子级的 QE 系统。它在单词级和句子级上都优于其他开源工具包，并产生了新的最先进的单词级 QE 结果。OpenKiwi 一经发布就作为 WMT 2019 QE 的基线系统。此外，所有 WMT 2019 QE 的单词、句子和文档级任务的获奖系统都使用 OpenKiwi 作为其构建基础。

论文题目：*Bridging the Gap between Training and Inference for Neural Machine Translation*

中文题目：架起一座在基于神经元的机器翻译训练和推理之间的桥梁

论文作者：Wen Zhang, Yang Feng, Fandong Meng, Di You, Qun Liu

论文出处：Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics.（2019）.

论文地址：https://arxiv.org/abs/1906.02448

研究问题：

神经机器翻译(Neural Machine Translation, NMT)是根据上下文预测下一个词，从而依次生成目标词。训练时用真值词作为上下文进行预测，而推理时必须从头开始生成整个序列，这会导致误差的积累。此外，单词级训练要求生成的序列与真值序列严格匹配会导致对不同但合理的翻译的过度矫正。针对这一问题，文章提出了一种过矫正恢复的方法。该方法不仅从真值序列中提取上下文，而且通过训练模型从预测序列中提取上下文，即翻译过程中模型不需要再逐词对比标准来确定损失函数。在中文→英文和英语→德语的翻译任务的实验结果表明，该方法可以在多个数据集上实现显著的改进。

研究内容：

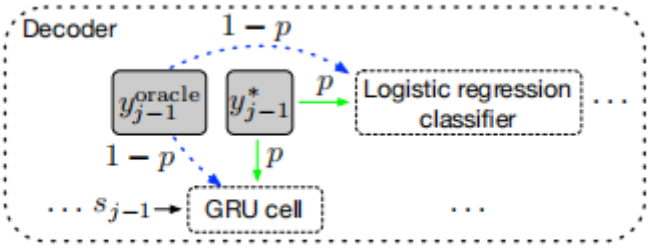


Figure 1: The architecture of our method.

模型主要结构如上图所示，核心思想是：不仅使用真值序列进行约束，在训练过程中，也利用训练模型预测出的上一个词语作为其中的备选词语来约束模型。对于 oracle 词的选择有两种方法，一种是用贪心搜索算法在词级上选择，另一种是在句级上选择最优的 oracle 序列。

在词语级的选择上，在时间步为 j 时，获取 j-1 时间步模型预测出的每个词语的预测分数。为了提高模型的鲁棒性，在预测分数的基础上加上 Gumbel noise，取分数最高的词语作为此时的 Oracle Word。

在句子级的选择上，使用集束搜索，选择集束宽为 k 的句子，然后计算每个句子的 BLEU 分数，选择分数最高的句子。对于生成的实际句子超出或短于这一长度的情况，文章采用强制解码的方式进行干预。

最终选择的 Oracle Word 也会和真值序列的词语混合，然后使用衰减式采样（Decay Sampling）的方法从中挑选出作为约束模型训练的词。

Systems	Architecture	MT03	MT04	MT05	MT06	Average
	Existing end-to-end NMT systems					
Tu et al. (2016)	Coverage	33.69	38.05	35.01	34.83	35.40
Shen et al. (2016)	MRT	37.41	39.87	37.45	36.80	37.88
Zhang et al. (2017)	Distortion	37.93	40.40	36.81	35.77	37.73
	Our end-to-end NMT systems					
this work	RNNsearch	37.93	40.53	36.65	35.80	37.73
	+ SS-NMT	38.82	41.68	37.28	37.98	38.94
	+ MIXER	38.70	40.81	37.59	38.38	38.87
	+ OR-NMT	**40.40**‡†*	**42.63**‡†*	**38.87**‡†*	**38.44**‡	**40.09**
	Transformer	46.89	47.88	47.40	46.66	47.21
	+ word oracle	47.42	48.34	47.89	47.34	47.75
	+ sentence oracle	**48.31***	**49.40***	**48.72***	**48.45***	**48.72**

Table 1: Case-insensitive BLEU scores (%) on Zh→En translation task. "‡", "†", "⋆" and "∗" indicate statistically significant difference ($p<0.01$) from RNNsearch, SS-NMT, MIXER and Transformer, respectively.

文章对 NIST 中文→英文（Zh→En）和 WMT14 英语→德语（En→De）的翻译任务进行了实验。结果表明，文章提出的方法可以在多个数据集上实现提升。

同时在 RNNsearch 模型和 Transformer 模型上也验证了该方法。结果表明，新方法可以显著提高两种模型的性能。

研究结果：

端到端的 NMT 模型在训练时逐字逐句地生成翻译，将真实单词作为上下文，而不是将模型生成的前一个单词作为上下文进行推理。为了减少训练和推理之间的差异，在预测一个单词时，文章使用抽样方法将真实单词或先前预测的单词作为上下文输入。被预测的单词（称为 oracle 单词）可以通过单词级或句子级优化生成。与词级 oracle 相比，句子级 oracle 进一步赋予了该模型过度矫正恢复的能力。通过两个基线模型和实际翻译任务的相关工作验证了该方法的有效性，并对所有数据集进行了显著的改进。文章还指出，句子级的 oracle 优于单词级别的 oracle。

论文题目：*Do you know that Florence is packed with visitors?Evaluating state-of-the-art models of speaker commitment*

中文题目：你知道佛罗伦萨到处都是游客吗?评估说话者结论确定性的最新模型

论文作者：Nanjiang Jiang, Marie-Catherine de Marneffe

论文出处：Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019.

论文地址：https://www.aclweb.org/anthology/P19-1412/

研究问题：

当一个演讲者问“你知道佛罗伦萨挤满了游客吗”时，我们可能会相信，但如果她问的是“你认为佛罗伦萨挤满了游客吗”，我们可能就不会相信了。推断说话者承诺（或称事件真实性）对于信息抽取和问答均至关重要。对说话者承诺的预测，是判断说话者在句子中对某一事件承诺到何种程度，是实际的、非实际的还是不确定的。本篇文章通过在数据集上分析模型误差的语言相关性，来探讨语言缺陷会导致说话者承诺模型出现错误模式的假设。

研究内容：

评估数据集选取的 CommitmentBank 包含 1200 条数据，涉及四种包含时态嵌入动词的取消蕴涵环境(否定句、情态动词、疑问句、条件句)。对于每条数据，从 Mechanical Turk 寻找至少 8 个母语为英语的人标注其说话者承诺。

(1)	**Context**	The answer is no, no no. Not now, not ever.
	Target	I *never* **believed** I could wish anyone dead but last night changed all that. Gold:1.56, Rule-based:3.0, Hybrid: 0.50
(2)	**Context**	Revenue is estimated at $18.6 million. The maker of document image processing equipment said the state procurement division had declared FileNet in default on its contract with the secretary of state uniform commercial code division.
	Target	FileNet said it *doesn't* **believe** the state has a valid basis of default and is reviewing its legal rights under the contract , Gold: -0.47, Rule-based: 3.0, Hybrid: 1.08 but said it can't predict the outcome of the dispute.
(3)	**Context**	**A**: Yeah, that's crazy. **B**: and then you come here in the Dallas area, um,
	Target	I *don't* **believe** that people should be allowed to carry guns in their vehicled . Gold: -2.64, Rule-based: 3.0, Hybrid: 1.40

Table 1: Examples from the CommitmentBank, with gold scores and predictions from rule-based and hybrid models. Embedding verbs in bold, entailment-canceling environments italicized. The gold score is the mean annotators' speaker commitment judgments towards the content of the complement.

文章评估了两种最先进的说话者承诺模型：Stanovsky 等人提出的基于规则的方法和 Rudinger 等人提出的神经网络方法，结果显示基于规则的模型表现得更好，但整体表现均不是很好，因为 CommitmentBank 与其他任何数据集相比，其相关性更低，绝对错误率更高。

为了更好地解释模型的输出，文章在分类实验中对它们进行了评估。使用高斯混合模型来获得三个平均得分的聚类，用两个模型进行预测。将均值最高的聚类标记为真（+），均值最低的聚类标记为假（-），剩余真值不确定（o）。结果显示两类模型对 o 类都没有预测能力。

	Precision		Recall		F1		Count
	Rule	Hybr.	Rule	Hybr.	Rule	Hybr.	
+	0.58	0.64	0.91	0.51	0.71	0.56	251
–	0.99	0.67	0.55	0.20	0.70	0.31	268
o	0.00	0.06	0.00	0.46	0	0.11	37
Total	0.74	0.61	0.67	0.35	0.66	0.41	556

Table 4: Classification performance of the models.

研究结果：

文章在 CommitmentBank 上评估了两种最先进的说话者承诺模型。研究发现，带有语言学知识的模型比基于 LSTM 的模型表现更好，这表明如果想要在这样的有挑战性的自然语言数据中捕捉说话者承诺信息的话，语言学知识是必不可少

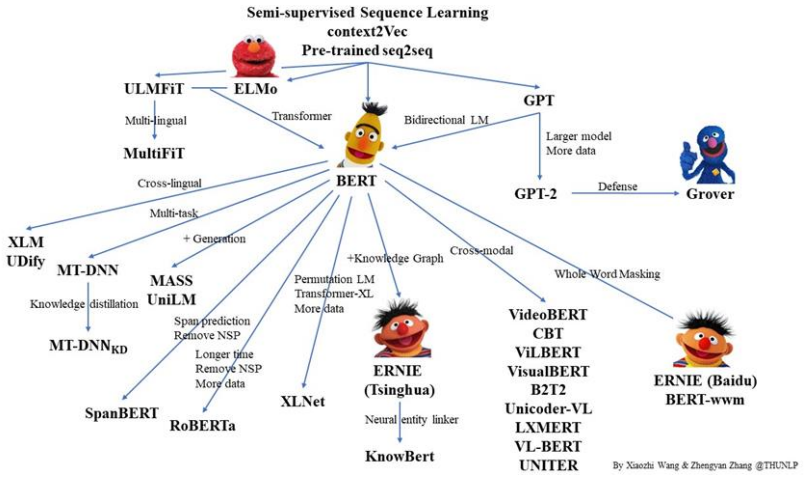
的。根据语言特征对标注数据进行分类可以发现不对称的错误模式。虽然模型在某些情况下（如否定）取得了良好的性能，但很难泛化到其他更丰富的自然语言的语言学结构（如条件句），从而指出了研究的改进方向。

5.5 自然语言处理最新进展

近年来，预训练语言模型在自然语言处理领域有了重要进展。预训练模型指的是首先在大规模无监督的语料上进行长时间的无监督或者是自监督的预先训练（pre-training），获得通用的语言建模和表示能力。之后在应用到实际任务上时对模型不需要做大的改动，只需要在原有语言表示模型上增加针对特定任务获得输出结果的输出层，并使用任务语料对模型进行少许训练即可，这一步骤被称作微调（fine tuning）。

自 ELMo、GPT、BERT 等一系列预训练语言表示模型（Pre-trained Language Representation Model）出现以来，预训练模型在绝大多数自然语言处理任务上都展现出了远远超过传统模型的效果，受到越来越多的关注，是 NLP 领域近年来最大的突破之一，是自然语言处理领域的最重要进展。

BERT（Bidirectional Encoder Representation from Transformer）是 Google AI 于 NAACL2019 提出的一个预训练语言模型。BERT 的创新点是提出了有效的无监督预训练任务，从而使得模型能够从无标注语料中获得通用的语言建模能力。模型的部分细节在前文的论文解读中已经给出，不再赘述。



BERT 之后涌现了许多对其进行扩展的模型（如上图所示），包括：跨语言预训练的 XLM 和 UDify，跨模态预训练的模型，融合知识图谱的 ERNIE，将 seq2seq 等语言生成任务整合入 BERT 类模型的 MASS, UniLM 等。其中几个重要的进展包括：

（1）XLNet 使用 Transformer-XL 替代了 Transformer 作为基础模型，拥有编码超长序列的能力。XLNet 提出了一个新的预训练语言任务：Permutation Language Modeling（排列语言模型），模型将句子内的词语打乱顺序，从而使得预测当前词语时可以利用双向信息。XLNet 相对 BERT 也使用了更多的语料。

（2）RoBERTa 采用了与 BERT 具有相同的模型结构，同样采用了屏蔽语言模型任务进行预训练，但舍弃了 BERT 中下句预测模型。此外，RoBERTa 采用了更大规模的数据和更鲁棒的优化方法，从而取得了更好的表现。

（3）ALBERT 模型针对 BERT 参数量过大难以训练的问题做了优化，一是对词向量矩阵做分解，二是在层与层之间共享参数。此外，ALBERT 将下句预测模型替换为句序预测任务，即给定一些句子预测它们的排列顺序。

6 语音识别

6.1 语音识别概念

语音识别是让机器识别和理解说话人语音信号内容的新兴学科，目的是将语音信号转变为文本字符或者命令的智能技术，利用计算机理解讲话人的语义内容，使其听懂人类的语音，从而判断说话人的意图，是一种非常自然和有效的人机交流方式。它是一门综合学科，与很多学科紧密相连，比如语言学、信号处理、计算机科学、心理和生理学等[8]。

语音识别首先要对采集的语音信号进行预处理，然后利用相关的语音信号处理方法计算语音的声学参数，提取相应的特征参数，最后根据提取的特征参数进行语音识别。总体上，语音识别包含两个阶段：第一个阶段是学习和训练，即提取语音库中语音样本的特征参数作为训练数据，合理设置模型参数的初始值，对模型各个参数进行重估，使识别系统具有最佳的识别效果；第二个阶段就是识别，将待识别语音信号的特征根据一定的准则与训练好的模板库进行比较，最后通过一定的识别算法得出识别结果。显然识别结果的好坏与模板库是否准确、模型参数的好坏以及特征参数的选择都有直接的关系。

实际上，语音识别也是一种模式识别，其基本结构如下图所示。和一般模式识别过程相同，语音识别包括如图所示 3 个基本部分。实际上，由于语音信息的复杂性以及语音内容的丰富性，语音识别系统要比模式识别系统复杂的多。

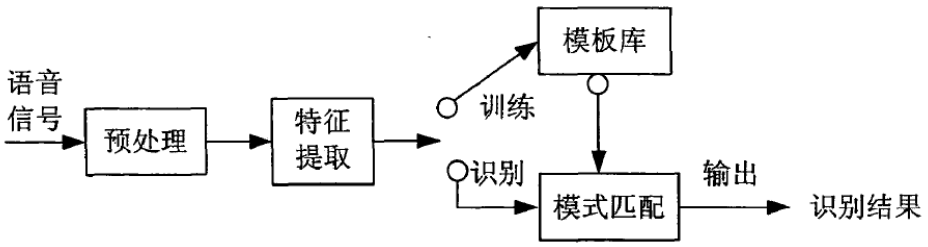


图 6-1 语音识别系统框架

其中，预处理主要是对输入语音信号进行预加重和分段加窗等处理，并滤除其中的不重要信息及背景噪声等，然后进行端点检测，以确定有效的语音段。特征参数提取是将反映信号特征的关键信息提取出来，以此降低维数减小计算量，

用于后续处理，这相当于一种信息压缩。之后进行特征参数提取，用于语音训练和识别。常用的特征参数有基于时域的幅度、过零率、能量以及基于频域的线性预测倒谱系数、Mel 倒谱系数等。

6.2 语音识别发展历史

语音识别的研究工作可以追溯到 20 世纪 50 年代。在 1952 年，AT＆T 贝尔研究所的 Davis，Biddulph 和 Balashek 研究成功了世界上第一个语音识别系统 Audry 系统，可以识别 10 个英文数字发音。这个系统识别的是一个人说出的孤立数字，并且很大程度上依赖于每个数字中的元音的共振峰的测量。1956 年，在 RCA 实验室，Olson 和 Belar 研制了可以识别一个说话人的 10 个单音节的系统，它同样依赖于元音带的谱的测量。1959 年，英国的 Fry 和 Denes 研制了一个能够识别 4 个元音和 9 个辅音的识别器，他们采用了谱分析仪和模式匹配器。所不同的是他们对音素的序列做了限制（相当于现在的语法规则），以此来增加字识别的准确率。但当时存在的问题是的理论水平不够，都没有取得非常明显的成功。

60 年代，计算机的应用推动了语音识别技术的发展，使用了电子计算机进行语音识别，提出了一系列语音识别技术的新理论—动态规划线性预测分析技术，较好的解决了语音信号产生的模型问题。该理论主要有三项研究成果。首先是美国新泽西州普林斯顿 RCA 实验室的 Martin 和他的同事提出一种基本的时间归一化方法，这种方法有效的解决了语音事件时间尺度的非均匀性，能可靠的检测到语音的起始点和终止点，有效地解决了识别结果的可变性。其次，苏联的 Vintsyuk 提出了用动态规划的方法将两段语音的时间对齐的方法，这实际上是动态时间规整（Dynamic Time Warping）方法的最早版本，尽管到了 80 年代才为外界知晓。第三个是卡耐基梅隆大学的 Reddy 采用的是音素的动态跟踪的方法，开始了连续语音识别的研究工作，为后来的获得巨大成功的连续语音识别奠定了基础。

70 年代，语音识别研究取得了重大的具有里程碑意义的成果，伴随着自然语言理解的研究以及微电子技术的发展，语音识别领域取得了突破性进展。这一时期的语音识别方法基本上是采用传统的模式识别策略。其中苏联的 Velichko 和 Zagoruyko 的研究为模式识别应用于语音识别这一领域奠定了基础；日本的迫江和千叶的研究则展示了如何利用动态规划技术在待识语音模式与标准语音模式

之间进行非线性时间匹配的方法；日本的板仓的研究则提出了如何将线性预测分析技术加以扩展，使之用于语音信号的特征抽取的方法。同时，这个时期还提出了矢量量化和隐马尔可夫模型理论。

80 年代，语音识别研究进一步走向深入。这一时期所取得的重大进展有：（1）隐马尔科夫模型（HMM）技术的成熟和不断完善，并最终成为语音识别的主流方法。（2）以知识为基础的语音识别的研究日益受到重视。在进行连续语音识别的时候，除了识别声学信息外，更多地利用各种语言知识，诸如构词、句法、语义、对话背景等方面的知识来帮助进一步对语音识别和理解。同时在语音识别研究领域，还产生了基于统计概率的语言模型。（3）人工神经网络（ANN）在语音识别中的应用研究的兴起。ANN 具有较好的区分复杂分类边界的能力，显然它十分有助于模式识别。在这些研究中，大部分采用基于反向传播算法（BP 算法）的多层感知网络[9]。

20 世纪 90 年代，语音识别技术逐渐走向实用化，在建立模型、提取和优化特征参数方面取得了突破性的进展，使系统具有更好的自适应性。许多发达国家和著名公司都投入大量资金用以开发和研究实用化的语音识别产品，从而许多具有代表性的产品问世。比如 IBM 公司研发的汉语 ViaVoice 系统，以及 Dragon 工司研发的 DragonDictate 系统，都具有说话人自适应能力，能在用户使用过程中不断提高识别率。

21 世纪之后，深度学习技术极大的促进了语音识别技术的进步，识别精度大大提高，应用得到广泛发展。2009 年，Hinton 将深度神经网络（DNN）应用于语音的声学建模，在 TIMIT 上获得了当时最好的结果。2011 年底，微软研究院的俞栋、邓力又把 DNN 技术应用在了大词汇量连续语音识别任务上，大大降低了语音识别错误率。从此语音识别进入 DNN-HMM 时代。DNN 带来的好处是不再需要对语音数据分布进行假设，将相邻的语音帧拼接又包含了语音的时序结构信息，使得对于状态的分类概率有了明显提升。同时 DNN 还具有强大环境学习能力，可以提升对噪声和口音的鲁棒性。

目前，语音识别技术已逐渐被应用于工业、通信、商务、家电、医疗、汽车电子以及家庭服务等各个领域。例如，现今流行的手机语音助手，就是将语音识

别技术应用到智能手机中，能够实现人与手机的智能对话功能。其中包括美国苹果公司的 Siri 语音助手，智能 360 语音助手，百度语音助手等[10]。

6.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为语音识别领域全球学者分布情况：



图 6-2 语音识别领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布在其东西海岸；亚洲也有较多的人才分布，主要在我国东部及日韩地区；欧洲的人才主要集中在欧洲中西部；其他诸如非洲、南美洲等地区的学者非常稀少；语音识别领域的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，语音识别领域中男性学者占比 87.3%，女性学者占比 12.7%，男性学者占比远高于女性学者。

语音识别领域学者的 h-index 分布如下图所示，大部分学者的 h-index 分布在中间区域，其中 h-index 在 30-40 区间的人数最多，有 752 人，占比 37.3%，小于 20 区间的人数最少，只有 6 人。

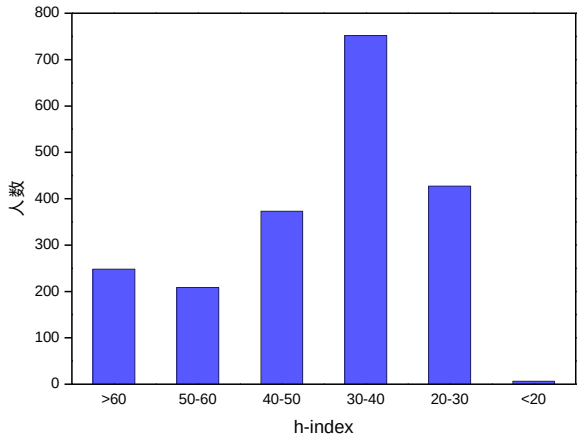


图 6-3 语音识别领域学者 h-index 分布

- 中国人才分布

我国专家学者在语音识别领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等亚洲国家相比，中国在语音识别领域学者数量较多且有一定的优势。



图 6-4 语音识别领域中国学者分布

中国与其他国家在语音识别领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 6-1 语音识别领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	922	14529	16	1548
中国-英国	207	3088	15	358
中国-新加坡	131	1788	14	221
中国-澳大利亚	92	577	6	194
中国-加拿大	84	921	11	165
中国-法国	76	1318	17	132
中国-日本	75	921	12	151
中国-德国	68	1099	16	110
中国-丹麦	32	501	16	31
中国-巴基斯坦	25	772	31	40

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在语音识别领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席；中国与巴基斯坦合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中巴合作达到了较高的水平。

6.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

IEEE International Conference on Acoustics, Speech and Signal Processing

IEEE Transactions on Audio, Speech, and Language Processing

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，噪声（noise）、语言模型（language modeling）、音频（audio）是本领域中最热的关键词。



论文题目：*X-Vectors: Robust DNN Embeddings for Speaker Recognition*

中文题目：X 向量：用于说话人识别的鲁棒 DNN 嵌入

论文作者：David Snyder, Daniel Garcia-Romero, Gregory Sell, Daniel Povey and Sanjeev Khudanpur. X-Vectors: Robust DNN Embeddings for Speaker Recognition.

论文出处：2018 IEEE International Conference on Acoustics, Speech and Signal Processing（ICASSP）

论文地址：https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8461375

研究问题：

捕捉说话者特征是语音识别领域具有重大意义的研究内容。大多数说话人识别系统都是基于 i-vectors 来实现的。标准的基于 i-vectors 的方法由通用背景模型（UBM）和大型投影矩阵 T 组成，该模型以无监督方式来学习。在早期的系统中，神经网络经训练后，被用来分离说话者，从网络中提取帧级表示，并将其用作高斯说话者模型的特征。近年来，使用深度神经网络（DNN）捕获说话者特征是当前非常活跃的研究领域。DNN 嵌入性能也随着训练数据量的增加而高度扩展。

研究方法：

在本文中，研究者们使用数据增强来提高用于说话人识别的深度神经网络（DNN）嵌入的性能。经过训练后，用于区分说话者的 DNN 将可变长度话语映射到固定维嵌入，并将其称为 x 矢量。本文使用一系列数据增强的方法，包括增

加噪声、增加混响等，用以增加训练数据的数量并提高其鲁棒性。训练后，DNN 可对训练数据中的 N 个说话者进行分类。一个训练示例包括大量语音特征（平均约 3 秒）和相应的说话者标签。该模型从图层 segment6 的仿射成分中提取嵌入。不包括 softmax 输出层和 segment7（因为训练后不需要它们），总共有 420 万个参数。基于 i-vector 和 x-vector 的系统均使用了 PLDA 分类器。x-vector 和 i-vector 需先中心化，其后再使用 LDA 投影。在 SITW 开发中将 LDA 尺寸调整为 i-vector 为 200，x-vector 为 150。降维后，再使用自适应 s 范数进行长度归一化和建模。

研究结果：

没有数据扩充的系统分别在 SWBD 和 SRE 数据集上对提取器进行了训练。不使用增强，SITW 上的最佳结果是通过 i-vector（BNF）获得的结果，比 DCF10-2 处的 x-vector 系统好 12%。与 SITW 上的 x-vector 系统相比，声学 i-vector 系统还实现了稍低的错误率。但是，即使不进行扩展，也可以通过 x-vector 获得 SRE16 粤语的最佳结果。就 DCF10-2 而言，这些嵌入比任意 i-vector 系统效果好约 14%。使用了数据增强的对比实验结果表明，PLDA 增强对所有系统都有明显的改进。x-vector 可以从 PLDA 增强中获得比 baseline 系统更高的改进效果。在 SITW 上，x-vector 系统的误码率略低于 i-vector（声学），但在大多数工作点上仍落后于 i-vector（BNF）。在 SRE16 上，在 DCF10-2 中，x-vector 比 i-vector 保持约 14%的优势。

论文题目：*Boosting Noise Robustness of Acoustic Model via Deep Adversarial Training*

中文题目：通过深度对抗训练提高声学模型的噪声鲁棒性

论文作者：Bin Liu, Shuai Nie, Yaping Zhang, Dengfeng Ke, Shan Liang, Wenju Liu Boosting Noise Robustness of Acoustic Model via Deep Adversarial Training

论文出处：2018 IEEE International Conference on Acoustics, Speech and Signal Processing（ICASSP）

论文地址：https://ieeexplore.ieee.org/document/8462093

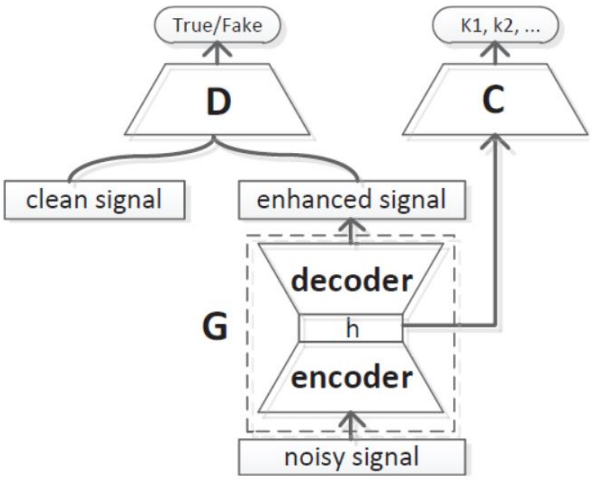
研究问题：

在真实环境中，语音信号很容易受到噪声和混响的干扰，尤其在远场条件下，由于声波在传播过程中其能量随传播距离呈指数衰减，语音信号受到噪声和混响的干扰更加严重，因此自动语音识别系统性能很难得到有效提升。针对语音识别系统在噪声环境下识别性能下降的问题，作者提出了深度对抗声学模型训练框架，减小了噪声环境语音数据和真实训练数据的分布差异，从而提升声学模型的鲁棒性。

研究方法：

语音识别模型的噪声鲁棒性问题主要来源于纯净训练数据和带噪测试数据的分布差异。生成式对抗网络（Generative Adversarial Networks, GAN）可以通过对抗训练的方式，连续逼近指定的数据分布。GAN 由生成器和判别器组成，生成器用来生成样本，判别器用来判断样本是否来自真实训练集。二者进行对抗训练，使得生成器生成的样本尽可能逼近真实训练数据。

针对语音识别系统在噪声环境下识别性能下降的问题，作者提出深度对抗和声学模型联合训练的框架，如下图所示，框架由生成器（G）、判别器（D）以及分类器（C）组成。生成器用来把带噪语音数据分布变成纯净语音；判别器用来判定语音信号是否来自真实纯净训练集；声学模型作为分类器，指导生成器提取区分性特征。生成器、判别器和声学模型进行联合对抗训练，三者相互配合相互促进。



通过深度对抗的联合训练策略，可以减小噪声环境语音数据和真实训练数据的分布差异，提升声学模型的鲁棒性。相对于语音增强方法，该框架没有增加计算的流程和复杂度，而且不需要一一对应的带噪数据和纯净数据，可作为通用训练框架提升已有声学模型的噪声鲁棒性。

研究结果：

作者使用了 CHiME-4 数据及对提出的方法进行测试，结果表明该方法能够有效提升声学模型的鲁棒性，在词错误率（WER）上相比于基线系统有较大的提升。

论文题目：*Modality Attention for End-to-end Audio-visual Speech Recognition*

中文题目： 基于模态注意力的端到端音视觉语音识别

论文作者：Pan Zhou, Wenwen Yang, Wei Chen, Yanfeng Wang, Jia Jia.

论文出处：2019 IEEE International Conference on Acoustics, Speech and Signal Processing （ICASSP）

论文地址：https://ieeexplore.ieee.org/document/8683733

研究问题：

随着语音识别的快速发展，纯粹靠声音的识别技术越来越成熟，识别准确率达到 95%以上，但是在嘈杂环境中语音识别的准确率会明显下降。而人在嘈杂环境中不仅靠声音信息，还结合讲话者的嘴唇和面部视觉信息来理解讲话者的意思。解决嘈杂环境下的语音识别问题可以通过在语音基础上加入视觉信息，通过视、听模态信息的融合来增强语音识别的效果（Automatic Visual Speech Recognition）。

利用视觉信息来增强语音识别的效果需要解决两个难题：一是两者帧率不同，如何将两种模态信息融合在一起，二是如何选择音频和视频的权重。

研究方法：

作者提出一种基于模态重要程度的注意力机制，可以根据模态的信息含量自适应调整模态的权重来融合音视觉特征。

具体方法是分别使用两个神经网络编码器对输入的不同模态的序列进行逐层特征抽取，得到高层特征表达。然后，由解码器分别对不同模态的特征表达进行注意力计算，得到声音和视觉模态信息中对应于当前解码时刻的上下文向量（context vector）。不同模态的上下文向量经由模态间的注意力自动对不同模态的信息进行融合，最后输入到输出层得到解码输出预测。

在编解码框架下，由于融合的是不同模态的上下文向量，而不是对原始特征或者编码器输出的高层特征进行直接融合，解决了不同模态的特征长度不同的问题。

同时，这种模态注意力（Modality Attention）依据不同模态各自的重要程度计算出相应的融合系数，反应了不同模态在当前解码时刻的不同贡献度，可以随着不同解码时刻的不同信噪比等得到不同的模态融合权重，得到更加鲁棒的融合信息。

研究结果：

作者在 150 小时的电视新闻类音视觉数据上进行了测试，在信噪比为 0dB（信号与噪声大小相当）时，多模态识别将准确率有很大程度的提高。而且模型在不同噪声下，体现出了对语音和视频两种不同模态间的依赖。随着噪声的提升，模型在融合音视觉时，对视觉信息的依赖比例在逐渐提升。

论文题目：*State-of-the-Art Speech Recognition with Sequence-to-Sequence Models*

中文题目：先进的序列识别语音识别模型

论文作者：Chung-Cheng Chiu, Tara N. Sainath, Yonghui Wu, Rohit Prabhavalkar, Patrick Nguyen, Zhifeng Chen, Anjuli Kannan, Ron J. Weiss, Kanishka Rao, Ekaterina Gonina, Navdeep Jaitly, Bo Li, Jan Chorowski, Michiel Bacchiani.

论文出处：2018 IEEE International Conference on Acoustics, Speech and Signal Processing（ICASSP）

论文地址：https://ieeexplore.ieee.org/abstract/document/8462105

研究问题：

序列到序列模型已经在自动语音识别（ASR）社区中获得了普及，这是一种将常规 ASR 系统的分离的声学，发音和语言模型（AM，PM，LM）折叠到单个神经网络中的方法。但到目前为止，我们还不清楚这样的方法是否可以取代当前基于 HMM 的最新技术的神经网络声学模型。尽管序列到序列模型是完全神经网络化的，无需有限的状态转换器、词典或文本规范化模块。训练这种模型比传统的 ASR 系统更简单：它们不需要决策树进行引导，也不需要从单独的系统生成的时间对齐。但是，迄今为止，这些模型都无法在大型词汇连续语音识别（LVCSR）任务上胜过最先进的 ASR 系统。

研究方法：

本文的目的是探索各种结构和优化方面的改进，以允许序列到序列模型在语音搜索任务上明显优于传统的 ASR 系统。在此工作中我们将重点放在对 LAS 模型的改进上。LAS 模型是一个单一的神经网络，其中包括类似于常规声学模型的编码器。我们既考虑对模型结构的修改，也考虑优化过程。在结构方面，首先，我们探索单词模型（WPM），我们比较了 LAS 的字素和 WPM，并发现 WPM 有适度的改进。接下来，我们探索合并多头注意力，它使模型能够学习到编码特征的多个位置。

研究结果：

实验结果显示，结构改进（WPM，MHA）后，在 WER 方面提高了 11%，而优化改进（MWER，SS，LS 和同步训练）后又提高了 27.5%，而语言模型记录的改进又提高了 3.4%。应用于 Google 语音搜索任务后，我们的 WER 为 5.6%，而混合 HMM-LSTM 系统的 WER 为 6.7%。在命令任务中测试了相同的模型，在 WER 指标方面，我们的模型达到 4.1%，混合系统达到 5%。

论文题目：*Deep Audio-visual Speech Recognition*

中文题目：深度视听语音识别

论文论文作者：Triantafyllos Afouras; Joon Son Chung; Andrew Senior; Oriol Vinyals; Andrew Zisserman.

论文出处：IEEE Transactions on Pattern Analysis and Machine Intelligence

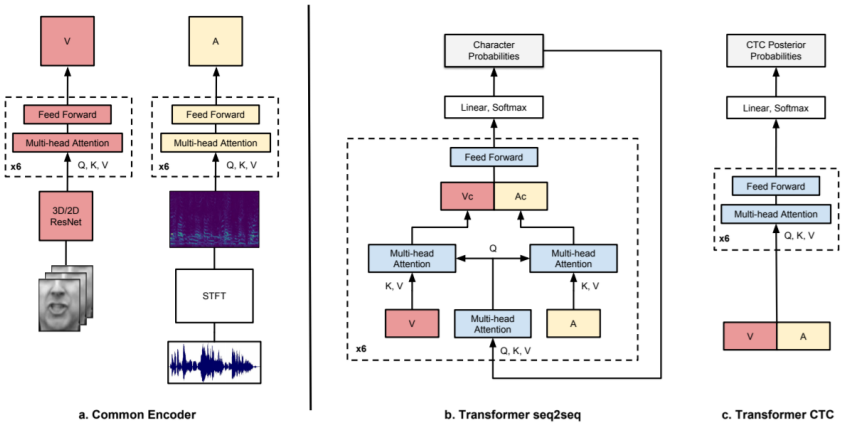
论文地址：https://ieeexplore.ieee.org/abstract/document/8585066

研究问题：

唇读，作为一种仅凭视觉信息就能识别所说内容的能力，是一项令人印象深刻的技能。由于同音字的存在，它在字面上本质上是模棱两可的-不同的字符会产生完全相同的口音序列（例如“p”和“b”）。合理的使用句子中相邻单词的上下文和/或语言模型在一定程度上解决此类歧义。唇读技术可以应用于许多场景：例如，在嘈杂的环境中“命令”向手机发送指令或消息；转录和重新复制档案无声电影；解决多人同时语音并且总体上改善了自动语音识别的性能。由于在计算机视觉任务中众所周知的两个发展，使得上述这些应用成为可能。基于为语音识别和机器翻译而开发的最新编码器-解码器体系结构——唇读模型变得尤为重要。

研究方法：

与以前的工作着重于识别有限数量的单词或短语不同，我们将唇读作为一个开放世界的问题来解决-无限制的自然语言句子和野外视频。首先，我们比较了两种唇读模型，一种使用 CTC 损失，另一种使用序列间损失。两种模型都建立在变压器自我关注架构的基础上。其次，我们研究了唇读在多大程度上与音频语音识别相辅相成，特别是当音频信号有噪声时；再次，我们引入并公开发布了两个用于视听语音识别的新数据集：LRS2-BBC，由英国电视台的数千个自然句子组成；和 LRS3-TED，其中包括从 YouTube 获得的数百小时的 TED 和 TEDx 演讲。我们训练的模型在唇读基准数据集上大大超越了所有先前的工作。



研究结果：

实验结果显示，效果最佳的网络是 TM-seq2seq，使用语言模型进行解码时，LRS-BBC 的 WER 达到 50%，与之前的 70.4%的最新水平相比，提高了 20%以上。在 LRS2-BBC 上进行评估时，TM-seq2seq 模型展示出增加波束宽度的效果。对比实验表明，当音频信号有噪声时，嘴巴的运动为语音识别提供了重要线索。甚至在音频信号干净的情况下也可以提高性能。例如，使用视听 TM-CTC 模型时，单词错误率从仅音频的 10.1%降低到 LRS2-BBC 的 8.2%，从 LRS3-TED 的 6.0%降低到 5.0%。与仅音频模型相比，使用视听 TM-seq2seq 时获得的收益相似。当在原始话语中添加噪声来合成的嘈杂音频与两个数据集的纯音频情况相比时，性能下降了 60%以上。这表明在仅限于音频模型的性能上，该模型对单词错误率的评分与仅使用嘴唇获得的错误率相似。但是，将这两种方式组合起来可带来显着的改进，所有模型和数据集的字错误率均下降 20%-30%。因此，在存在较大背景噪音的情况下，视听模型的性能要比仅视频或仅音频的模型好得多。

论文题目：*Parameter Uncertainty for End-To-End Speech Recognition*

中文题目：端到端语音识别中的参数不确定性研究

论文作者：Stefan Braun and Shih-Chii Liu.

论文出处：2019 IEEE International Conference on Acoustics, Speech and Signal Processing.

论文地址：https://ieeexplore.ieee.org/abstract/document/8683066

研究问题：

近期端到端（End-to-End）的自动语音识别（Automatic Speech Recognition, ASR）研究相比于 DNN-HMM 混合系统，在模型结构和训练过程方面有了明显简化。传统的端到端模型通常使用确定性参数（Deterministic Parameters），即每个参数对应一个确定的实数值。然而在对相关任务进行训练过程中，该类方法仅对参数大小进行编码，没有直接对参数的不确定性（Uncertainty）或重要性（Importance）进行直接编码，但是这些内容也包含重要的信息。因此学者们开始研究将参数以概率的形式进行编码，来探索神经网络中的参数不确定性。相关

研究工作表明，在自动语音识别之外的其他多个任务的剪枝实验中，参数信噪比（Signal-to-Noise Ratio, SNR）与参数重要性展现出很高的相关性。目前在语音识别领域相关的研究还很少，已知仅有的一项研究从贝叶斯模型角度使用变分推断框架（variational inference framework）导出概率网络图。本文从参数角度提出另一种概率网络，避免了对贝叶斯模型解释的需求。

研究方法：

本文研究使用不确定性参数的端到端方法将自动语音识别任务的领域适用性，包括纯净语音和带噪语音；提出基于信噪比（SNR-based）的正则化方案来控制参数根据其重要性来更新；使用不同的信噪比水平的概率网络来评估；对比了在领域适用过程中不同信噪比水平的网络如何容忍（tolerate）参数剪枝及灾难性遗忘（catastrophic forgetting）程度在网络中是如何变化的。

基础端到端的模型包括 5 层的双向 LSTM 网络（每个方向包含 320 个单元）和最终 640×59 的网络映射至输出标签。

确定性模型（deterministic models）使用默认 LSTM 单元，参数集合θ_D包含 LSTM 权重w^{LSTM}，偏差b^{LSTM}及映射权重w^{PROJ}

$$\theta_D = \{w^{LSTM}, b^{LSTM}, w^{PROJ}\}$$

概率模型（probabilistic models）使用具有高斯权重的 LSTM 模型，参数集合θ_P包含 LSTM 权重的均值μ^{LSTM}，参数化权重标准差β^{LSTM}，偏差b^{LSTM}及映射权重w^{PROJ}

$$\theta_P = \{\mu^{LSTM}, \beta^{LSTM}, b^{LSTM}, w^{PROJ}\}$$

使用 Xavier uniform initialization（下式）初始化w^{LSTM}，μ^{LSTM}和w^{PROJ}。

$$A_{ij} \sim \mathcal{U}\left(-\frac{\sqrt{6}}{i+j}, \frac{\sqrt{6}}{i+j}\right), A \in \mathbb{R}^{i \times j}$$

参数化标准差β^{LSTM}使用下式进行初始化

$$B_{ij} = \log\left(\exp\left(\frac{1}{2}\frac{\sqrt{6}}{i+j}\right) - 1\right), B \in \mathbb{R}^{i \times j}$$

在概率模型中，对参数化标准差β^{LSTM}通过使用权值衰减（衰减项$\mathcal{L}_\beta = ||\beta||_2^2$）来增强低信噪比参数。

研究结果：

本文对使用具有概率权重参数的 LSTM 对端到端的语音识别模型进行了评估。测试集使用 Wall Street Journal（在纯净条件下的数据）和 CHiME-4 的语音识别任务（含有噪音数据）。实验结果表明在参数剪枝和领域适用性方面概率模型获得了比确定性模型更好的结果。概率模型的关键优势是对特定参数信噪比的可用性，在训练时与参数的重要程度相关性较高。

论文题目：*Stochastic Adaptive Neural Architecture Search for Keyword Spotting*

中文题目：面向关键词检出的随机自适应神经网络结构搜索

论文作者：Tom Véniat, Olivier Schwander and Ludovic Denoyer

论文出处：2019 IEEE International Conference on Acoustics, Speech and Signal Processing.

论文地址：https://ieeexplore.ieee.org/document/8683305

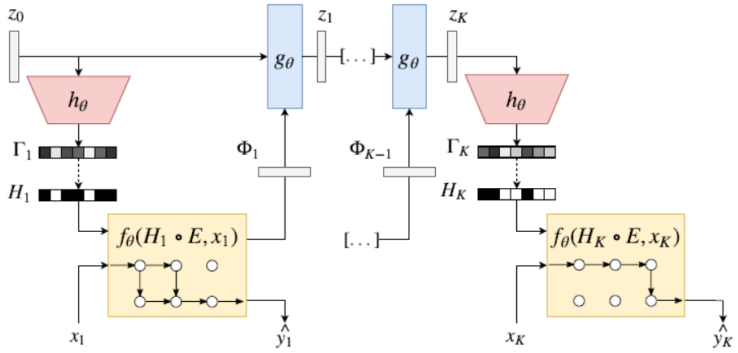
研究问题：

目前关键词定位（Keyword Spotting）问题（如在实时音频流中确定关键词）的主要方法是在连续的滑动窗口中使用神经网络模型进行识别。在目前神经网络搜索（Neural Architecture Search）的研究中发现的网络结构都是静态的（相同的神经网络结构在预测时重用）。由于任务的复杂性，目前基准模型通常很大，导致预测阶段计算资源及能源消耗水平较高。

研究方法：

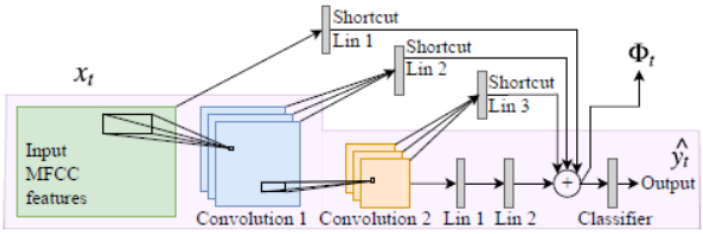
文章提出了随机自适应神经网络结构搜索（Stochastic Adaptive Neural Architecture Search, SANAS）模型，能够在模型推断阶段自适应地在线调整神经网络的结构（当任务简单时使用较小的结构，当任务复杂时使用较大的结构）。关键词定位（Keyword Spotting）可以抽象为一个音频流序列标注问题，在每个时间步长（timestep），系统接收一个数据点x_t，生成一个输出标签y_t（在音频流中x_t通常为一个时频特征图，y_t为给定关键字是否存在的判断）。

文章定义了一种可以根据上下文的隐含表示预测在每个时步（timestep）进行网络结构变化的设置。



在时步 t 中，从前一隐藏状态生成的结构分布Γ_t，$\Gamma_t =$（z_t, θ），其中z_t为上下文$x_1, y_1, \ldots, x_{t-1}, y_{t-1}$的编码隐含表示，在每一步$z_t$的更新根据神经网络结构$\mathcal{A}$，参数$\theta$确定$z_{t+1} = g$（$z_t, x_t, \theta, \mathcal{A}$）。然后从$\Gamma_t$中抽取出离散结构$\mathrm{H}_t$，并通过输入$x_t$进行评估。评估过程给出特征向量$\Phi_t$（$\mathrm{x}_t, \theta, E \circ \mathrm{H}_t$）来计算下一个隐含状态，并根据$f$（$\mathrm{z}_t, \mathrm{x}_t, \theta, E \circ \mathrm{H}_t$）来预测模型$\hat{y}_t$。虚线代表 sampling 操作，在推理阶段，每个 timestep 中选出具有最高概率的结构。

下图为基于卷积神经网络（cnn-trad-fpool3）的 SANAS 结构：



网络层之间的连接通过上述模型采样生成，高亮的网络结构是增加了快捷连接（shortcut connections）的基准模型。

研究结果：

实验评估数据使用了 Speech Commands 数据集。实验对比了传统的静态模型和本文提出的方法，结果表明 SANAS 方法能够很大程度上降低 FLOPs（每秒浮点运算次数），同时相对于基准方法识别出了更多的关键字，准确率也更高。

论文题目：Adversarially Trained End-to-end Korean Singing Voice Synthesis System

中文题目：基于对抗训练的端到端韩语歌声语音合成系统

论文论文作者：Juheon Lee, Hyeong-Seok Choi, Chang-Bin Jeon, Junghyun Koo, Kyogu Lee.

论文出处：20th Annual Conference of the International Speech Communication Association INTERSPEECH 2019.

论文地址：https://www.isca-speech.org/archive/Interspeech_2019/pdfs/1722.pdf

研究问题：

深度学习技术的发展催生了一批基于神经网络的歌声语音合成系统，系统基于歌声、歌词及乐谱平行数据来训练声学模型（acoustic model），用来预测声码器（vocoder）的输入特征。尽管这些基于神经网络的歌声合成系统能够有较好的表现，但是预测声码器特征的网络仍然具有局限：不能超过声码器的性能上限。因此，本文研究能够直接生成线性频谱图来代替声码器特征的端到端模型，其挑战在于端到端模型增加了模型的复杂性，而且需要大量的训练数据。

研究方法：

本文提出的歌声合成系统能够基于适量的训练数据使用端到端的方式进行训练。系统结构如下图：

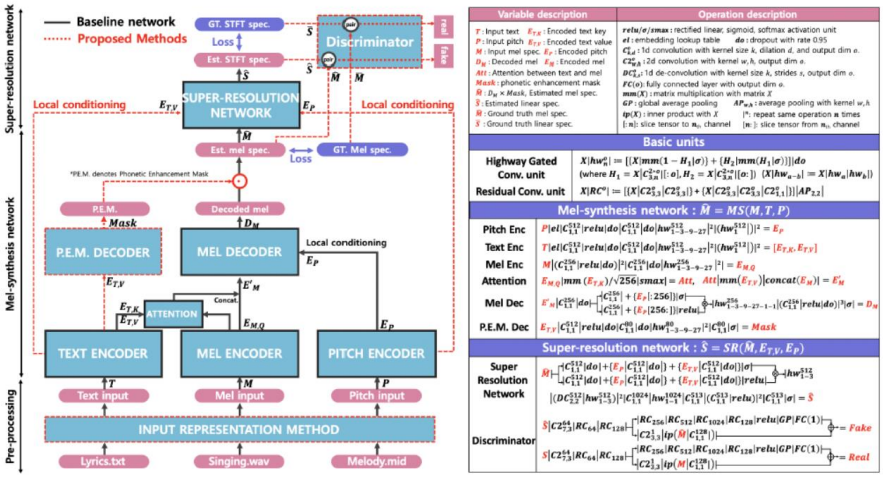


Figure 1: *Proposed system overview (left), detailed structure of each sub-module (right). $X|F|$ denotes $F(X)$.*

系统主要由两个模块构成：旋律合成网络 mel-synthesis 和超分辨率网络 super-resolution。mel-synthesis 网络根据前面的旋律输入 $M_{0:L-1}$，时序对齐的文本$T_{1:L}$及音调输入 $P_{1:L}$训练生成旋律谱图；super-resolution 网络根据文本和音调信息作为条件输入，将生成的旋律谱图 M进行上采样（upsample）；最后判别器（discriminator）将上采样结果和生成的旋律谱图以对抗的方式训练网络。

在测试阶段，从给定文本及音调输入中以自回归的方式生成旋律谱图的帧序列，然后通过 super-resolution 网络上采样为线性谱图，最后通过 Griffin-Lim 算法转换为声波形式（waveform）。

研究结果：

实验使用手工收集整理的歌声数据集，包含了 60 首流行歌曲。实验表明使用文本信息对 phonetic enhancement mask 进行建模是有效的，能够生成更为准确的发音。同时在 super-resolution 阶段使用条件对抗（conditional adversarial）训练方法能够获得更高的声音质量。

6.5 语音识别进展

随着人工智能的迅速发展，语音识别的技术越来越成为国内外研究机构的焦点。人们致力于使机器能够听懂人类的话语指令，并希望通过语音实现对机器的控制。作为一项人机交互的关键技术，语音识别在过去的几十年里取得了飞速的发展，在研究和探索过程中针对语音识别的各部流程进行了各种各样的尝试和改造，以期发现更好的方法来完成语音识别流程中的各个步骤，以此来促进在不同环境下语音识别的效率和准确率。研究人员从最简单的非常小词汇量的阅读式的语音识别问题开始，逐渐转向越来越复杂的问题。

近年来智能语音进入了快速增长期，语音识别作为语音领域的重要分支获得了广泛的关注，如何提高声学建模能力和如何进行端到端的联合优化是语音识别领域中的重要课题。

语音识别经历了从 2012 年最开始的 DNN 的引入时的 Hybrid HMM 结构，再到 2015 年开始吸引大家研究兴趣的 CTC 算法，而后到 2018 年的 Attention 相关结构的研究热点。Attention 相关算法在语音识别或者说话人识别研究的文章中

出现频率极高。从最开始 Attention，到 Listen-Attend-Spell，再到 Self-Attention（或者 Transformer），在不同的文章被作者多次介绍和分析，频繁出现在了相关文章的 Introduction 环节中。在 Attention 结构下，依然还有很多内容需要研究者们进一步地探索：例如在一些情况下 Hybrid 结构依然能够得到 State-of-the-art 的结果，以及语音数据库规模和 Attention 模型性能之间的关系。

在近两年的研究中，端到端语音识别仍然是 ASR（Automatic Speech Recognition）研究的一大热点，正如上文提到的，基于 Attention 机制的识别系统已经成为了语音技术研究主流。同时，随着端到端语音识别框架日益完善，研究者们对端到端模型的训练和设计更加的关注。远场语音识别（far-field ASR），模型结构（ASR network architecture），模型训练（model training for ASR），跨语种或者多语种语音识别（cross-lingual and multi-lingual ASR）以及一些端到端语音识别（end-to-end ASR）成为研究热点。

在语音合成方面，高音质语音生成算法及 Voice conversion 是近两年研究者关注的两大热点，Voice Conversion 方向的研究重点主要集中在基于 GAN 的方法上。在语言模型方面（Language Model）的研究热点主要包括 NLP 模型的迁移，低频单词的表示，以及深层 Transformer 等。

在说话人识别方面，说话人信息，特别是说话人识别及切分，正被越来越多的研究者所重视。目前 Attention 在说话人方面更类似一种 Time Pooling，比 Average Pooling 及 Stats Pooling 更能捕捉对说话人信息更重要的信息，从而带来性能提升。说话人识别技术经历深度学习带来的性能飞跃后，在模型结构、损失函数等方面的探讨已经较为成熟，以 TDNN、ResNet 加上 LMCL、ArcFace 的主流模型开始不断刷新各数据集的性能上限。模型以外的因素逐渐成为制约说话人系统的瓶颈。说话人技术目前也逐渐暴露出与人脸识别同样的易受攻击的问题。因此，ASVspoof 这样的 Challenge 从 2015 年起就开始关注声纹反作弊问题。相信随着此类研究的不断深入，结合声纹系统的性能提升，声纹将有望变成我们的“声音身份证”。

7 计算机图形学

7.1 计算机图形学概念

国际标准化组织 ISO 将计算机图形学定义为：计算机图形学是一门研究通过计算机将数据转换成图形，并在专门显示设备上显示的原理方法和技术的学科。它是建立在传统的图形学理论、应用数学及计算机科学基础上的一门边缘学科。这里的图形是指三维图形的处理。简单来讲，它的主要研究内容是研究如何在计算机中表示图形，以及利用计算机进行图形的计算处理和显示的相关原理和算法。

在计算机图形学的开创之初，他主要解决的问题是在计算机中表示三维结合图形以及如何利用计算机进行图形的生成处理和显示的相关原理和算法，目的是产生令人赏心悦目的真实感图像，这仅仅是狭义的计算机图形学。随着近些年的发展，计算机图形学的内容已经远远不止这些，广义的计算机图形学研究内容非常广泛，包括图形硬件、图形标准、图形交互技术、栅格图形生成算法、曲线曲面造型、实体造型、真实版图形的计算、显示算法、科学计算可视化、计算机动画、虚拟现实、自然景物仿真等等。

计算机图形学的总体框架可以包括以下几个部分：数学和算法基础、建模、渲染以及人机交互技术。计算机图形学需要一些基本的数学算法，例如向量和几何的变化、几何建模式的三维空间变化、三维到二维的图形变换等等。建模是进行图形描述和计算，由于在多维空间中有各种组合模型，有一些是解析式表达的简单形体，也有一些隐函数表达的复杂曲线，因此需要进行复杂的建模工作。渲染也叫绘制，指的是模型的视觉实现过程，例如对光照纹理等理论和算法进行处理，其中也需要大量的计算。交互技术可以说是图形学交互的重要工具，是计算机图形学的重要应用。

7.2 计算机图形学发展历史

20 世纪 50 年代：1950 年，美国 MIT 的旋风一号（whirlwind I）计算机配备了世界上第一台显示器——阴极射线管（CRT）来显示一些简单的图形，使得计算机摆脱了纯数值计算的单一用途，能够进行简单的图形显示，从此计算机具有

了图像显示功能，但是还不能对图形进行交互操作，这时的计算机图形学处于准备和酝酿时期，并称之为“被动式”图形学。

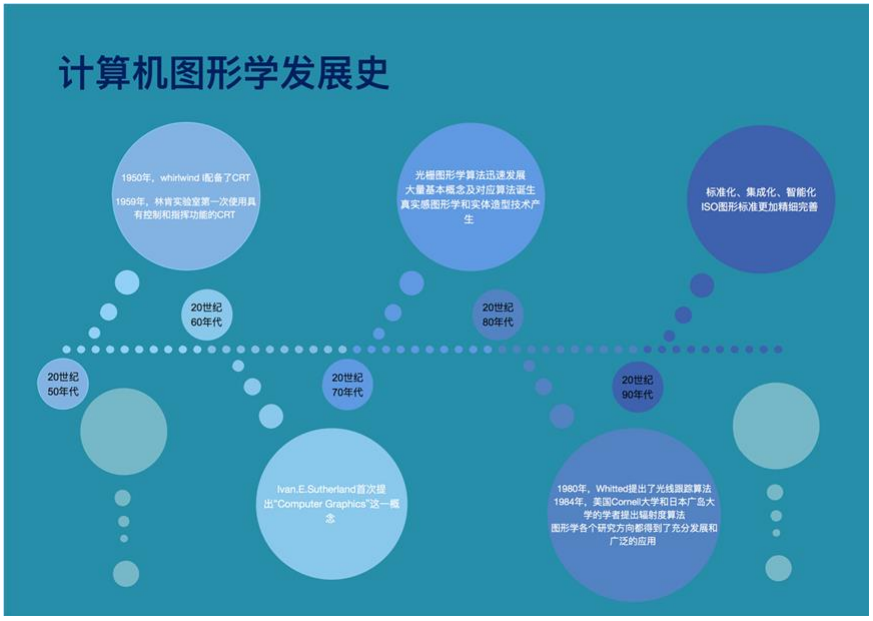


图 7-1 计算机图形学发展历史

50 年代末期，MIT 的林肯实验室在“旋风”计算机上开发 SAGE（Semi-Automatic Ground Environment System）空中防御体系。SAGE 于 1957 年投入试运行，已经能够将雷达信号转换为显示器上的图形并具有简单的人机交互功能，操作者使用光笔点击屏幕上的目标即可获得敌机的飞行信息，这是人类第一次使用光笔在屏幕上选取图形。1959 年，麻省理工学院林肯实验室第一次使用了具有指挥和控制功能的 CRT，“被动式”图形学开始迈向交互式计算机图形学。

20 世纪 60 年代：1962 年美国 MIT 林肯实验室的 IvanE.Sutherland 发表了一篇题为“sketchpad：一个人机交互通信的图形系统”的博士论文，首次使用了“Computer Graphics”这一概念，证明了交互式计算机图形学是一个可行的、有应用价值的研究领域，从而确立了计算机图形学正式成为一个独立学科的分支。1968 年 IvanE.Sutherland 又发表了《头戴式三维显示器》的论文，在头盔的封闭环境下，利用计算机成像的左右视图匹配，生成立体场景，使人置身于虚拟现实中。IvanE.Sutherland 为计算机图形学技术做出了巨大的贡献，被称作计算机图形学的开山鼻祖，1988 年 IvanE.Sutherland 被授予 A.M 图灵奖。并且这一时期，光栅图形学算法开始萌芽。

20 世纪 70 年代：图形学在这一时期进入了兴盛期，光栅图形学算法迅速发展，区域填充、裁剪、消隐等概念及其相应算法纷纷被提出，实用的 CAD 图形系统也开始出现。除此之外，真实感图形学和实体造型技术的产生也是 70 年代计算机图形学的两个重要进展。1970 年 J.Bouknight 在 ACM 上发表论文，提出了第一个光反射模型，指出物体表面朝向是确定物体表面上一点光强的主要因素，并用 Lambert 漫反射定律计算物体表面上各多边形的光强，对光照射不到的地方用环境光代替。1971 年 Henri Gouraud 在 IEEE Trans.Computer 上提出被称为 Gouraud 明暗处理的“漫反射模型＋插值”思想，对多面体模型，用漫反射模型计算多边形顶点的光亮度，再用增量法插值计算多边形的其他内部点。1975 年 Phong 在 ACM 上发表论文提出了著名的简单光照模型“Phong 模型”，Phone 模型虽然只是一个经验模型，但其真实度已经达到了较好的显示效果。这些都是真实感图形学最早的开创性工作。从 1973 年开始，相继出现了英国剑桥大学 CAD 小组的 Build 系统、美国罗彻斯特大学的 PADL-1 系统等实体造型系统，这些都为 CAD 领域的发展做出了重要贡献。

70 年代图形软件标准化程度提高，1974 年，ACMSIGGRAPH“与机器无关的图形技术”的工作会议的召开，提出了图形软件标准化问题，ACM 成立图形标准化委员会，制定“核心图形系统”（core graphics system），ISO 发布 CGI、CGM、GKS、PHIGS 一系列的图形标准，其中 19777 年的 CKS 是 ISO 批准的第一个图形软件标准软件，是一个二维图形软件标准，1986 年，ISO 公布了程序员级的分层结构交互图形系统 PHIGS，这是一些非官方的图形软件，广泛应用于工业界并成为事实上的标准，PHIGS 是对 CKS 的扩充，增加的功能有对象建模、彩色设定、表面绘制和图形管理等。伺候 PHIGS 的扩充成为 PHIGS+；1988 年的 CKS3D，是 ISO 批准的第二个图形软件标准软件，是一个三维图形软件标准。

20 世纪 80 年代以后：出现了带有光栅扫描显示器的微型计算机和图形工作站，极大的推动了计算机图形学的发展，如 Machintosh、IBM 公司的 PC 及其兼容机，Apollp、Sun 工作站等。随着奔腾 III 和奔腾 IV 系列 CPU 的出现，计算机图形软件功能开始部分地由硬件实现。高性能显卡和液晶显示屏的使用，高传输率大容量硬盘的出现，特别是 Internet 的普及使得微型计算机和图形工作站在运算速度、图形显示细节上的差距越来越小，这些都为图形学的飞速发展奠定了物

质基础。1980 年 Turner Whitted 提出了光透视模型，并第一次给出光线跟踪算法的范例，实现了 Whitted 模型；1984 年美国 Cornell 大学和日本广岛大学的学者分别将热辐射工程中的辐射度算法引入到计算机图形学中，用辐射度的方法成功地模拟了理想漫反射表面间的多重漫反射效果。以上二者的提出，标志着真实感图形的显示算法已逐渐成熟。80 年代中期以后，超大规模集成电路的发展，计算机运算能力的提高，图形处理速度的加快，促使了图形学各个研究方向都得到了充分发展和广泛的应用。

20 世纪 90 年代以后：微机和软件系统的普及使得图形学的应用领域日益广泛，计算机图形学朝着标准化、集成化和智能化的方向发展，多媒体、人工智能、计算机可视化、虚拟现实等分支蓬勃发展，三维造型也获得了长足发展。ISO 公布的图形标准越来越精细，更加成熟。这时存在着一些事实上的标准，如 SGI 公司开发的 OpenGL 开放式三维图形标准，微软公司为 PC 游戏开发的应用程序接口标准 DirectX 等，Adobe 公司 Postscript 等，均朝着开放、高效的方向发展[11]。

7.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为计算机图形学全球学者分布情况：



图 7-2 计算机图形学领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显；欧洲也有较多的人才分布，主要在欧洲中西部；亚洲的人才主要集中在我国东部及日韩地区；其他诸如非洲、南美洲等地区的学者非常稀少；计算机图形学的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，计算机图形学领域中男性学者占比 93.7%，女性学者占比 6.3%，男性学者占比远高于女性学者。

计算机图形学领域学者的 h-index 分布如下图所示，分布情况大体呈阶梯状，大部分学者的 h-index 分布在中低区域，其中 h-index 在小于 20 区间的人数最多，有 1240 人，占比 60.1%，50-60 区间的人数最少，有 50 人。

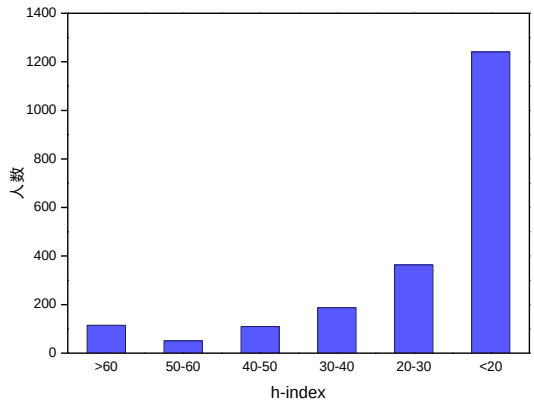


图 7-3 计算机图形学领域学者 h-index 分布

- 中国人才分布

我国专家学者在计算机图形领域的分布如上图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩等地相比，中国在计算机图形领域学者数量略多但差距较小。



图 7-4 计算机图形学领域中国学者分布

中国与其他国家在计算机图形学领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 7-1 计算机图形学领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	237	8729	37	407
中国-加拿大	69	3550	51	85
中国-以色列	59	3203	54	58
中国-英国	34	2299	68	72
中国-新加坡	28	1080	39	37
中国-德国	24	625	26	42
中国-瑞士	21	779	37	37
中国-印度	16	784	49	28
中国-沙特阿拉伯	16	468	29	22
中国-法国	15	485	32	36

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在计算机图形领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席；中国与英国合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中英合作达到了较高的水平。

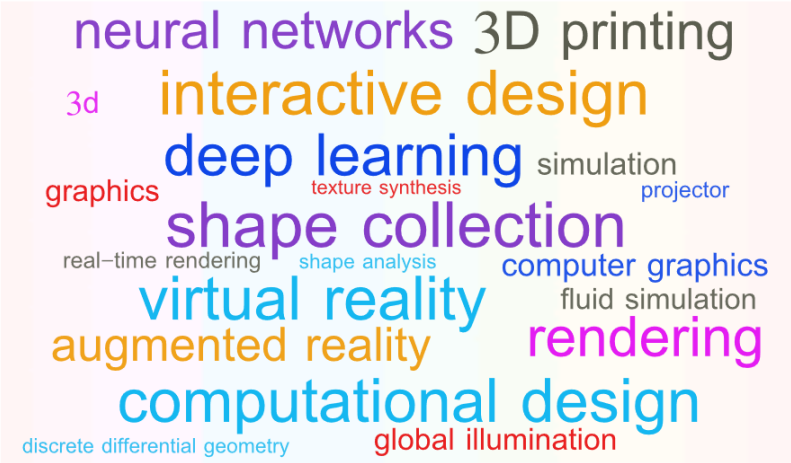
7.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

ACM SIGGRAPH Conference

ACM Transactions on Graphics

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，形状集合（shape collection）、交互设计（interactive design）、计算设计（computational design）是本领域中最热的关键词。



论文题目：*A Style-based Generator Architecture for Generative Adversarial Networks*

中文题目：基于样式的生成式对抗网络生成器架构

论文作者：Tero Karras, Samuli Laine, Timo Aila.

论文出处：The IEEE Conference on Computer Vision and Pattern Recognition- CVPR 2019

论文地址：

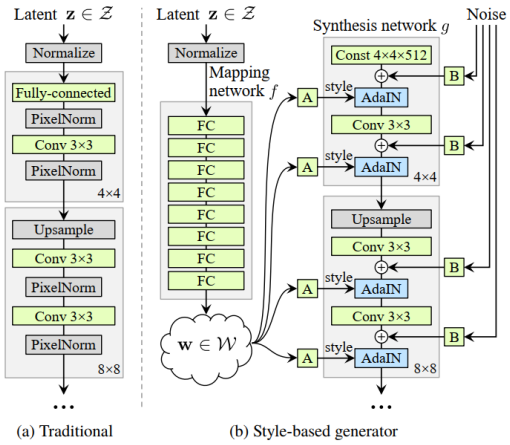
http://openaccess.thecvf.com/content_CVPR_2019/papers/Karras_A_Style-Based_Generator_Architecture_for_Generative_Adversarial_Networks_CVPR_2019_paper.pdf

研究问题：

本文针对自动的无监督的习得图像的高层属性（譬如人脸对应的身份信息以及拍摄姿态）以及对于生成的每幅图像产生一些特定的随机化的变换（譬如脸部瑕疵以及头发的细节），生成较为直观且可控的合成结果进行了研究。通过借鉴风格迁移的思想，提出了一种新的对抗网络中的生成器架构。该架构不仅在传统的分布距离的度量上优势明显，并且较好地将控制图像变化的隐变量分离出来进行独立建模。

研究方法：

下图为提出的方案与传统的生成器的对比：



传统的潜在编码表示由前向神经网络的第一层生成，而作者提出不再使用网络编码，而是如图(b)的右侧子图所示，由一个可通过反向传播学习的常数作为网络输入。在这一新的架构中，由映射网络 f 将采样自隐空间 Z 的隐变量 z，经由非线性映射网络 f，投射到潜在的中间表达空间 W 中，该空间控制着不同的视觉特征的表达。通过可学习的仿射变换，又将潜在空间中的编码 w 解码为样式信息 y。 而 y 通过对自适应样本归一化（adaptive instance normalization, AdaIN）后的特征分布进行放缩和偏置操作：

$$\mathrm{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i},$$

影响生成网络 g 的不同分辨率的卷积层输出通道的重要性，进而决定了输出图片的视觉特征。相较于风格迁移中使用图片作为迁移目标，这里使用解码自向量 w 的具有空间无关特性的样式编码 y，结合高斯噪声来输入生成网络 g 创建新的图像。

研究结果：

作者提出的生成器设计将不同的样式信息引入到生成网络 g 的不同层，可以实现在不同分辨率上对 w 施加影响，进而控制高级属性以及细小特征上的改变。如图（a-c）所示在不同分辨率上将源分布生成的样式编码引入目标图像，产生不同的由高层（a）到细节（c）的改变。作者展示在经典的名人脸部数据集以及新提出的更加多样化的人脸数据集上的当前最佳性能。并且，作者将其应用于车辆以及卧室场景数据集上，也取得了较好的视觉效果。



(a) (b) (c)

论文题目：*TempoGAN: A Temporally Coherent, Volumetric GAN for Super-resolution Fluid Flow*

中文题目：用于超分流体的时间一致性生成对抗网络

论文作者：You Xie, Erik Franz, Mengyu Chu, Nils Thuerey

论文出处：ACM Transaction on Graphics – SIGGRAPH-2018

论文地址：https://arxiv.org/pdf/1801.09710.pdf

研究问题：

生成对抗网络（GAN）过去几年在表示和生成复杂的自然图像方面取得了很大的进展。这些工作表明，深度卷积神经网络（CNNs）能够有效提取输入图片的特征分布，并用于生成全新的，不同于原始输入的图片。同时，相关工作如 PG-GAN 在从粗糙输入生成高分辨率的自然图像方面也很成功。 但是，这些生成模

型在其推导和改进过程中并没有将数据的时间信息考虑进来，而这些时间信息在现实的物理系统中是十分重要的。因此本文考虑数据的时间信息，扩展现有的 GAN 方法，以生成高分辨率数据集，并确保其时间连续性。

研究方法：

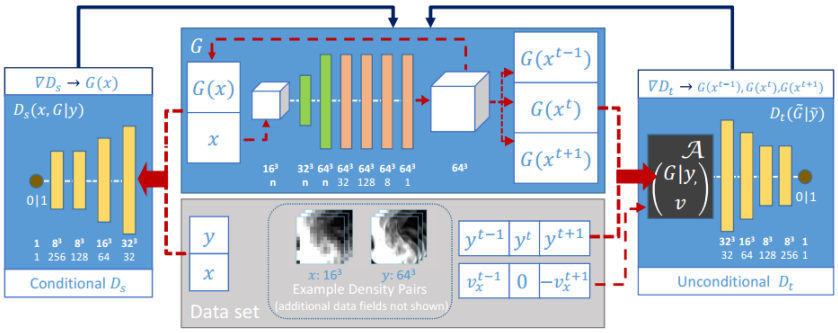
本文在之前特征空间损失相关工作的基础上，进一步改善生成自然图像的逼真度和解决模式崩溃问题。为了实现上述目标，本文在生成器中引入神经网络的部分特征空间的 L2 损失。具体来讲，生成网络的中间结果将被神经网络的间接参考数据所约束，为此，本文引入新的损失函数形式：

$$\mathcal{L}_f = \mathbb{E}_{n,j}\lambda_f^j \left\| F^j(G(x)) - F^j(y) \right\|_2^2$$

其中j是判别网络中的某一层，F^j表示相应层的激活响应，系数λ_f^j是权重项，用于平衡各个层之间的不平衡性。进一步地，为了使生成的数据具有时间一致性，文章还设计了下面的 L2 损失函数：

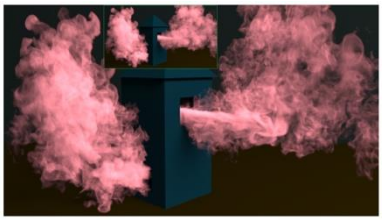
$$\mathcal{L}_{2,t} = \|G(x^t) - \mathcal{A}(G(x^{t-1}), v_x^{t-1})\|_2^2$$

下图是文章设计的网络结构，使用全部由卷积层构成的网络作为生成器，使得模型能够应用于生成任意尺寸的数据。网阔结构借鉴了 U-Net 和 ResNet，通过 skip 连接综合 low-level 信息和 high-level 信息。实验证明这种结构生成的自然图像质量最高。



研究结果：

文章设计的生成网络能够学会生成高精度的，逼真的且时间上具有一致性的流体图形。所设计的方法仅使用单个时间步长的低分辨率流体数据即可完成。在二维和三维数据上都取得了不错的效果。


论文题目：*Temporal Segment Networks for Action Recognition in Videos*

中文题目：用于视频行为识别的时间分段网络

论文作者：Wang, Limin and Xiong, Yuanjun and Wang, Zhe and Qiao, Yu and Lin, Dahua and Tang, Xiaoou and Van Gool, Luc.

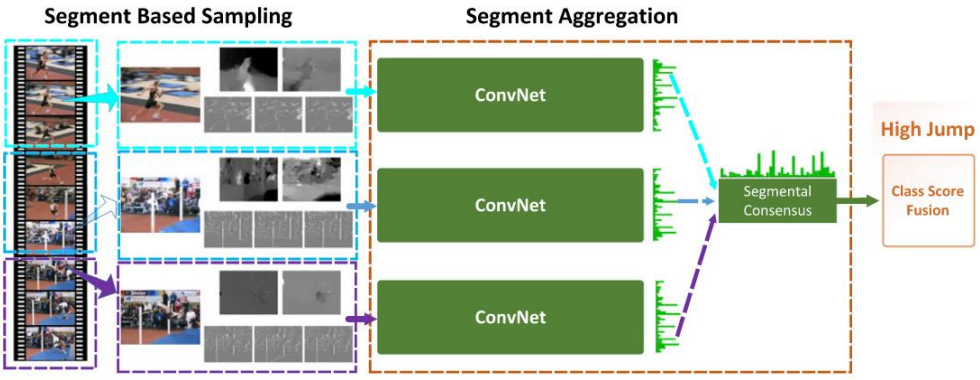
论文出处：IEEE transactions on pattern analysis and machine intelligence (2018)

论文地址：https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8454294

研究问题：

行为识别是计算机图像处理的重要应用领域，基于视频的行为识别是行为识别问题中的重要方法。传统的方法往往对单帧图像进行分析，这种方法缺乏前后帧的信息建模。有一种前后帧建模的方式是提取前后帧的光流信息，利用光流信息来完成运动的建模。然而光流只能描述短暂的顺序关系，无法对较长的时序进行建模。

研究方法：



这篇文章对整个视频进行分段，每一段作为一个时序模块。具体来讲每个时序模块中都进行两种建模，一个是单帧图像独立的卷积神经网络（CNN）模型，而另一个则是采用光流信息进行输入的卷积神经网络（CNN）模型。与之前方法

只进行较短时间的建模不同的是，这篇文章将视频分成三个较长的片段，对三个长片段进行分析。从而完成较长时序上的建模。

研究结果：

这篇文章所提出的 TSN 方法在 HMDB51, UCF101, THUMOS14, ActivityNet v1.2 以及 Kinetics400 数据库上进行了广泛的实验，取得了很高的精度。当采用 RGB 的前后帧差分图来代替光流图时，这篇文章的方法能够取得很快的实测速度。

论文题目：*Non-local Neural Networks*

中文题目：非局部神经网络

论文作者：Xiaolong Wang, Ross Girshick, Abhinav Gupta, Kaiming He

论文出处：The IEEE Conference on Computer Vision and Pattern Recognition- CVPR 2018

文章连接：

http://openaccess.thecvf.com/content_cvpr_2018/papers/Wang_Non-Local_Neural_Networks_CVPR_2018_paper.pdf

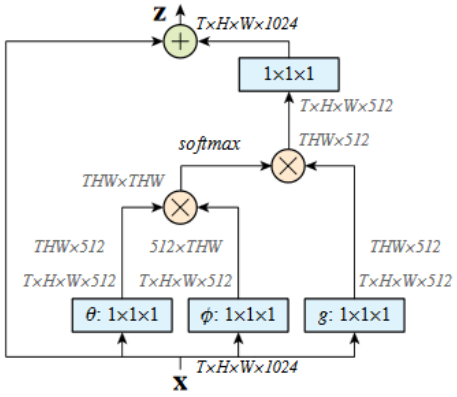
研究问题：

在图像目标检测、视频分类等任务中，捕捉长距离、大范围内的数据依赖关系是一个很重要的问题。在传统卷积神经网络中，通常使用较大的卷积核来达到捕捉远距离像素之间关系的目的。在序列化数据如视频中，通常使用循环神经网络建立时序上距离远的像素之间的联系。但是卷积和循环操作本身都是局部操作，只能在有限的空间范围内捕捉，难以获得更远甚至全局的像素间依赖关系。

研究方法：

受到 non-local mean 方法的启发，作者提出非局部模块，它可以很好地捕捉时间或空间维度上较远距离的像素间的依赖关系。非局部模块的计算过程如上图所示。

以视频分类为例，输入到非局部模块的特征图是一个形状为(T, H, W, C)的四维张量，首先将这个经过一个卷积的特征图与另一个经过卷积的自身的转置在 C 维度上做矩阵乘，之后经过 softmax 得到一个形状为(THW, THW)的张量，该张量与另一个经过卷积层的特征图做矩阵乘，结果再经过一个卷积层后与输入的特征图相加得到输出。



研究结果：

作者在 Kinetics、Charades 和 COCO 数据集上针对视频分类、目标检测等任务进行了实验。视频分类任务中，在 Kinetics 数据集上，带有非局部模块的网络的 top1 精度会比其基准网络提升大约 1.6 个百分点，top5 精度提升约 1 个百分点；在 Charedes 数据集上测试集精度大约有 2.3 个百分点的提升。目标检测任务中，在 COCO 数据集上，作者提出的网络相比其基准网络提升约 1.4 点 mAP。实验结果证明了非局部模块对于建立图像中或图像间长距离依赖关系的有效性。

论文题目：*Squeeze-and-Excitation Networks*

中文题目：压缩与激励网络

论文作者：Jie Hu, Li Shen, Gang Sun

论文出处：The IEEE Conference on Computer Vision and Pattern Recognition- CVPR 2018

文章连接：

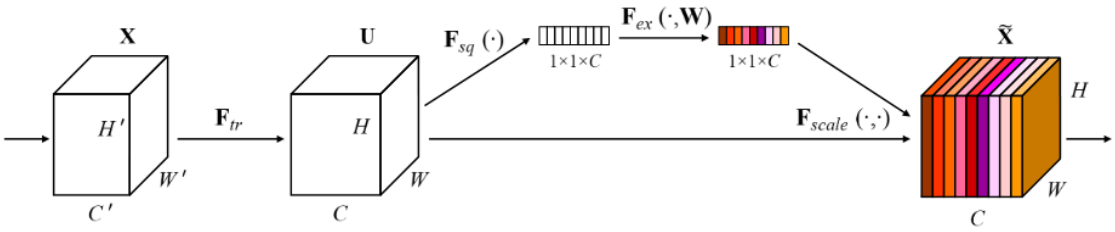
http://openaccess.thecvf.com/content_cvpr_2018/papers/Hu_Squeeze-and-Excitation_Networks_CVPR_2018_paper.pdf

研究问题：

如何通过建模特征通道之间的依赖关系提升网络性能，通过学习，使得有效特征图权重变大，而削弱无效图权重，从而使网络关注到重要部分。

研究方法：

提出的 SE 模块（Squeeze-and-Excitation Block）包括压缩和激励两个关键操作，利用它对特征通道间的信息进行融合。融合时不引入新的空间维度，而是采用了一种全新的特征重标定策略。通过学习的方式来自动获取到每个特征通道的重要程度，然后依照这个重要程度去提升有用的特征并抑制对当前任务用处不大的特征。如下图所示，对于一张形状为(H, W, C)的特征图，SE 模块首先进行压缩操作，在空间维度将特征图进行全局平均得到一个表示全局信息的特征向量。其次是激励操作，作者设计这一步的思想类似于 RNN 中的门控机制，通过一个全连接和激活层为每个特征通道生成权重，全连接层中的参数用来显式地建模各个通道间的相关性。最后进行重加权操作，通过乘法逐通道地将特征权重乘到输入的特征图上，完成特征重标定操作。



研究结果：

作者分别在 ImageNet 和 COCO 数据集上进行了图像分类和目标检测任务的实验。分类任务中，分别以 ResNet50、ResNet101、ResNet152 为基准，加入了 SE 模块的网络与基准相比，top1 分类误差分别减小了 1.51%、0.79%、0.85%，而相应的 GFLOPs 只增加了 0.01、0.02、0.02。在目标检测任务上，使用 Faster RCNN 模型，在主干网络上加入 SE 模块的模型所得 AP 值比基准网络高出大约 2.2~2.4 个百分点。从实验结果可以看出 SE 模块对于帮助网络建立特征通道间的依赖关系有很大帮助。

论文题目：*Neural Ordinary Differential Equations*

中文题目：神经网络常微分方程

论文作者：Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, David Duvenaud .

论文出处：The 32nd Conference on Neural Information Processing Systems - NeurIPS 2018

论文地址：https://arxiv.org/abs/1806.07366

研究问题：

已有的神经网络模型由大量隐层堆叠而成，是一种离散的序列化的表示形式。作者考虑将神经网络的隐状态参数化为连续时间下的隐变量模型，实现消除离散的层的概念的连续的新型神经网络模型。该种方案将神经网络视为带求解的隐状态序列，通过已有的常微分方程求解方案(ODE-solver)来求解网络的输出。为避免常规的数值解法的较高计算代价及累积误差，作者利用伴随敏感度方法(adjoint sensitivity method)将 ODE-solver 视为黑箱，利用两次对于增广 ODE 反向传播算法求解对于参数的梯度，而无需真正关注内部的具体操作。此种求解方案广泛的适用于各类 ODE-solver，可以在求解精度与速度间进行平衡。

研究方法：

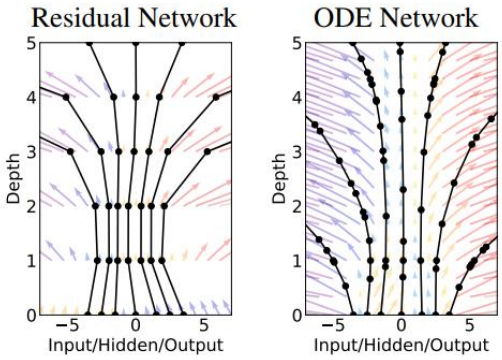
下式可以视为 ResNet 中的残差连接，表示某卷积层的输出与该层的输入的加和：

$$\mathbf{h}_{t+1} = \mathbf{h}_t + f(\mathbf{h}_t, \theta_t)$$

而这种残差连接又可以被视为 ODE-solver 中欧拉法的对连续系统的一种离散表示。在传统的神经网络中，离散的时间步体现为层的概念。 通过不断增加层，将得到尽可能小的时间步，进而得到参数化的神经网络隐状态的连续表示，也即上式的微分形式 ：

$$\frac{d\mathbf{h}(t)}{dt} = f(\mathbf{h}(t), t, \theta)$$

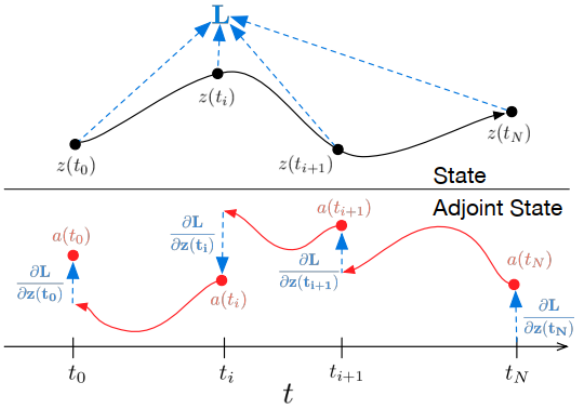
从初始层 h(0)触发，输出层 h(T)可以被定义为关于该初始值的在时间步 T 时的上述 ODE 的解。下图给出了残差网络所对应的隐变量的离散化表示形式，以及神经常微分方程网络的连续化状态表示：



若通过反向传播算法来求解连续状态的网络，因其前向过程往往对应于密集的数值积分方法，对积分求解微分将会带来巨大的内存开销和数值误差。通过将 ODE-solver 视为黑箱模型，采用伴随敏感度方法将使得计算随问题规模线性增加，同时伴随着较低的内存开销。考虑如下损失：

$$L(\mathbf{z}(t_1)) = L\left(\mathbf{z}(t_0) + \int_{t_0}^{t_1} f(\mathbf{z}(t), t, \theta)dt\right) = L\left(\text{ODESolve}(\mathbf{z}(t_0), f, t_0, t_1, \theta)\right)$$

伴随敏感度方法求解方案如下图：



考虑如图所示的增广的连续时间系统，其包含原始状态和损失函数关于状态的敏感性两部分。如果损失直接依赖于在多个观测点的状态，那么伴随状态也应相应的在损失于观测点处的偏导数方向进行更新。对于该初值问题的参数θ的反向传播算法如下：

Algorithm 1 Reverse-mode derivative of an ODE initial value problem

Input: dynamics parameters θ, start time t_0, stop time t_1, final state $\mathbf{z}(t_1)$, loss gradient $\partial L / \partial \mathbf{z}(t_1)$

$s_0 = [\mathbf{z}(t_1), \frac{\partial L}{\partial \mathbf{z}(t_1)}, \mathbf{0}_{|\theta|}]$ ▷ Define initial augmented state

def aug_dynamics($[\mathbf{z}(t), \mathbf{a}(t), \cdot], t, \theta$): ▷ Define dynamics on augmented state

  return $[f(\mathbf{z}(t), t, \theta), -\mathbf{a}(t)^\intercal \frac{\partial f}{\partial \mathbf{z}}, -\mathbf{a}(t)^\intercal \frac{\partial f}{\partial \theta}]$ ▷ Compute vector-Jacobian products

$[\mathbf{z}(t_0), \frac{\partial L}{\partial \mathbf{z}(t_0)}, \frac{\partial L}{\partial \theta}] = \text{ODESolve}(s_0, \text{aug_dynamics}, t_1, t_0, \theta)$ ▷ Solve reverse-time ODE

return $\frac{\partial L}{\partial \mathbf{z}(t_0)}, \frac{\partial L}{\partial \theta}$ ▷ Return gradients

研究结果：

作者在图像相关的监督学习任务，MNIST 书写数字分类上验证了 ODE 网络的性能。实验结果显示，使用相较于对应离散表示的残差网络，连续化状态表示的 ODE 网络具有与之相当的性能，但参数量减小为原来的 1/3，并且具有恒定的内存复杂度以及可以随求解精度变化的计算复杂度。

论文题目：*The Lottery Ticket Hypothesis: Finding Sparse, Trainable Neural Networks*

中文题目：彩票假设：寻找稀疏可训练的神经网络

论文出处：The 7th International Conference on Learning Representations – ICLR2019

论文地址：https://arxiv.org/pdf/1803.03635.pdf

研究问题：

网络稀疏是神经网络加速压缩的常用策略之一。目前主流的稀疏方法由三个阶段组成，即预训练、裁剪和重训练。在重训练过程中已有的方法都是将预训练得到的权重用于初始化，这里隐含的假设是裁剪后的子网络难以从头训练，因此需要借助预训练的结果辅助优化。本文则指出在预训练前的原始大网络中，存在特定的子网络结构和其初始化权重的组合，作者称其为“中奖彩票”，可以在与原大网络相同迭代次数的训练设置下获得更好的测试精度和收敛表现，也就是本文提出的“彩票假设”。以此为动机，作者一方面给出通过非结构化稀疏找到“中奖彩票”的方法，另一方面通过详尽的对比实验验证了“中奖彩票”的存在性以及方法的有效性。

研究方法：

为了找到给定初始化权重下原始网络中的“中奖彩票”，本文提出基于非结构化稀疏的搜索和训练策略，具体两种策略如下所示。

Strategy 1: Iterative pruning with resetting.

1. Randomly initialize a neural network $f(x; m \odot \theta)$ where $\theta = \theta_0$ and $m = 1^{|\theta|}$ is a mask.
2. Train the network for j iterations, reaching parameters $m \odot \theta_j$.
3. Prune $s\%$ of the parameters, creating an updated mask m' where $P_{m'} = (P_m - s)\%$.
4. Reset the weights of the remaining portion of the network to their values in θ_0. That is, let $\theta = \theta_0$.
5. Let $m = m'$ and repeat steps 2 through 4 until a sufficiently pruned network has been obtained.

Strategy 2: Iterative pruning with continued training.

1. Randomly initialize a neural network $f(x; m \odot \theta)$ where $\theta = \theta_0$ and $m = 1^{|\theta|}$ is a mask.
2. Train the network for j iterations.
3. Prune $s\%$ of the parameters, creating an updated mask m' where $P_{m'} = (P_m - s)\%$.
4. Let $m = m'$ and repeat steps 2 and 3 until a sufficiently pruned network has been obtained.
5. Reset the weights of the remaining portion of the network to their values in θ_0. That is, let $\theta = \theta_0$.

对比这两种策略，不同之处在于策略 1 在每次迭代裁剪后仍用原初始权重开始训练，策略 2 则是用上一次迭代得到的权重继续训练。而在确定子网络结构后，两种策略都是用原初始权重进行训练。实验验证策略 1 下子网络最终收敛的性能会明显好于策略 2，因此文中之后的对比实验都是基于前者进行。

研究结果：

文章先在 MNIST 和 CIFAR10 上用浅层的全连接和卷积神经网络进行对比实验，结果表明采用上述策略得到的稀疏子网络用原网络的初始权重初始化相比于重新初始化在相同训练设置下最终模型收敛性能有显著性提升，并且在一定稀疏率条件下甚至超过原大网络。更进一步地，同样在 CIFAR10 上，当采用更深的网络如 VGG19 和 ResNet18 时，需要对子网络采用更小的学习速率并结合 Warmup 等训练策略才能找到“中奖彩票”。从这些实验结果出发，作者希望能启发更多沟通权重初始化、网络结构与归纳偏置、神经网络泛化性及优化的研究和工作。

论文题目：*Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations*

中文题目：挑战无监督表征学习中的解耦合假设

论文出处：The 36th International Conference on Machine Learning – ICML2019

论文地址：https://arxiv.org/pdf/1811.12359.pdf

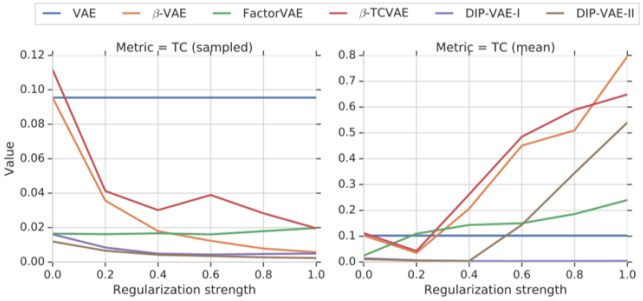
研究问题：

在使用生成模型进行表征学习时通常假设数据的生成过程由式$P(x) = P(z)P(x|z)$表示，其中x表示原始数据如图片，z则对应于语义明确的可解释变量并假设不同分量间相互独立。解耦合表征学习的目的即学习到某种特征变换$r(x)$能与这些解释变量间构成对应关系，使得学到的特征更加高效鲁棒便于下游任务提取任务相关的信息。本文通过搭配已有解耦合表征学习方法和不同数据集及评价指标构建不同实验设置，批判性指出现有工作的不足和基本假设上的缺陷。

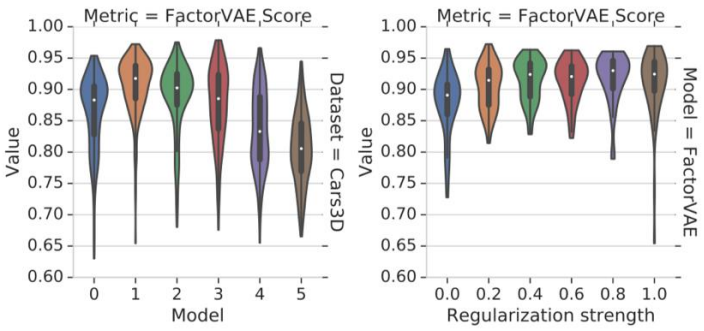
研究方法：

首先文章从理论上给出一个直觉性的结论，即任何有效的解耦合表征学习必须基于某种对模型和数据分布的归纳偏置。接着为了进一步验证其他基本假设，文章考虑了在 6 种不同已有方法与评价指标，以及 7 个不同数据集下进行对比实验。这里我们简要说明几个主要实验及其结果：

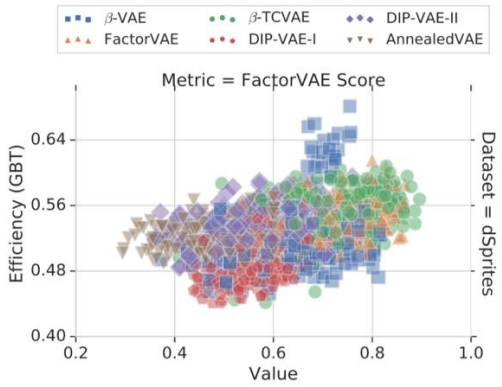
1）已有工作中为了使学得的满足正态分布先验的隐层特征z解相关，大都在训练过程中对编码器的输出加入相关性的正则项。然而如下右图所示，这种方式下训练最后得到的表征不同维度间的相关性反而不断上升。



2）相比于不同方法的选择，训练超参甚至随机数种子对模型最终的性能表现更为重要，这从下图每个方法在不同超参和随机数种子下性能表现的方差上可以看出。



3）直觉上我们认为解耦合的表征能更好辅助下游任务的训练，下图衡量了不同评分表征学习模型的样本利用效率，在实验给定评价指标下并未明显观察到表征解耦合带来的样本利用效率的提升。



研究结果：

总的来说，本文首先强调了归纳偏置在无监督解耦合表征学习中的重要性并给出理论上的证明。然后实验部分，文章主要归纳出以下几点结论：1、对隐层特征z后验分布加入的可分解约束并不能保证学得解耦合的特征表示；2、训练超参和随机数种子对不同方法性能的影响甚至超过方法本身带来的性能差异；3、在当前解耦合表征性能的衡量指标下，更好的解耦合表示并未使得下游任务样本利用效率上的明显提升。

7.5 计算机图形学进展

随着数字化技术和互联网的发展，计算机图形学在许多领域都已经得到了广泛的应用，如遥感图像分析、多媒体通信、医疗诊断、机器人视觉等。当前计算机图形学的研究逐渐向多学科交叉融合方向发展，即有与认知计算、计算器学习、人机交互的融合，也有与大数据分析、可视化的融合；不仅针对三维数字模型，

而且涵盖了图像视频，与计算机视觉深度交叉。计算机图形学的快速发展，一个潜在的趋势是不再有明确清晰的主题，更多的体现出方法和技术的创新。

针对近两年计算机图形学重要期刊会议的相关论文，对该领域内容热点研究内容和前沿技术方法进行了综合分析。目前，热点研究内容主要集中在自监督学习（Self-Supervised Learning）、全景分割（Panoptic Segmentation）、网络结构搜索（Neural Architecture Search）和生成式对抗网络（Generative Adversarial Networks）等方面。

自监督学习研究早期主要集中在代理任务的设计和选取上，怎样的代理任务才能最好地提取出有益于下游任务的特征以及为何这些代理任务能够有效，这些是理论层面上自监督学习仍需要解决的问题。随着大量围绕着实例判别代理任务的相关工作的提出，有一些工作将其中的核心思想进行展开提出了所谓对比学习的概念。通过将原来两个图片实例特征间的对比延伸到任意两个模态间特征的对比，使得模型学习不同模态间一致的特征表达并用最大化互信息作为新的衡量准则。

在已有的工作中，比较典型的代理任务有将图片分块然后预测不同分块间的相对位置或者将分块打乱后重排得到原图，以及基于图片的上下文信息进行补全和图片不同颜色通道间的相互预测等。目前在图像与图形学领域，取得性能突破的方法主要仍局限在监督学习的框架之下，随着无标记数据的不断爆增和模型性能进一步提升的需求，无监督学习将会越来越受到学术界和工业界的重视。而作为目前无监督学习中的一支，自监督学习因其良好的特征判别能力和对大规模数据扩展能力，也将受到更广泛的关注[12-23]。

全景分割作为一个统一的任务在 2018 年被提出，它的目标是为图像中的所有像素点都分配一个语义类别和一个实例编号，从另一个角度来说，全景分割算法需要预测出图像中每一个像素点的所属类别和所属实例。在全景分割任务的基础上，近期的进展主要体现在三个方面：（1）从图像整体的角度考虑全景分割，共享网络主干（backbone）形成设计整体网络结构；（2）考虑图像中不同元素之间的交互，建模物体与语义概念之间的关系；（3）提出可学习模块，解决预测结果层面的冲突。接下来，我们将分别介绍有代表性的工作。全景分割作为一个

最近被提出的视觉任务，受到了很大的关注，目前方法也在探讨的过程中，具有很大的发展潜力[24-38][24]。

目前深度学习的方法在各类图像与图形分析任务中取得了非常大的成功，伴随这一成功而来的是对网络结构设计要求的不断提高。自动化网络设计自然而然地成为了自动化机器学习的下一个目标。早期的相关工作证明了使用强化学习算法可以发现好的网络架构，但是这些方法在计算过程中需要消耗大量计算资源，因此后续的工作都集中在如何减少计算负担上。搜索空间的设计也是一项重要研究热点，同时，研究人员又拓宽了神经结构搜索的视野，将多种优化目标考虑在内，而不仅仅是减少搜索时间和提高网络精度。具有代表性的工作如尝试限制模型参数的数量或类似的方法，以有效地部署在移动设备上。在此基础上，还有一些工作将网络结构搜索技术扩展到搜索深度网络相关组件上[39-48]。

在图像合成方面，近期最引人关注的工作就是生成对抗网络，生成对抗网络由一个生成网络 G 和一个判别网络 D 组成。生成网络 G 和判别网络 D 在训练阶段使用对抗的方式进行学习，生成网络 G 的目标是生成尽可能真实的图片使得判别网络认为这是一张真实的图片；而判别网络 D 的任务则是判别合成的图像是真实的还是生成的。在这种两者对抗的学习过程中，生成 G 学会如何生成真实的图片。目前在生成对抗网络研究中，条件生成对抗网络、损失函数的改进、模型结构的改进及训练方法的改进是主要研究方向[49-55]。

8 多媒体技术

8.1 多媒体概念

“多媒体”一词译自英文“Multimedia”，而该词又是由 multiple 和 media 复合而成，核心词是媒体。媒体（medium）在计算机领域有两种含义:一是指存储信息的实体，如磁盘、光盘、磁带、半导体存储器等，中文常译为媒质；二是指传递信息的载体，如数字、文字、声音、图形和图像等，中文译作媒介，多媒体技术中的媒体是指后者。其实，“媒体”的概念范围是相当广泛的。“媒体”有下列五大类：（1）感觉媒体（Perception medium）指的是能使人产生直接感觉的媒体。如声音、动画、文本等；（2）表示媒体（Representation medium）指的是为了传送感觉媒体而人为研究出来的媒体。诸如语言编码、电报码、条形码等等；（3）显示媒体（Presentation medium）指的是用于通信中使电信号和感觉媒体之间产生转换用的媒体。如键盘、鼠标器、打印机等；（4）存储媒体（Storage medium）指的是于存放某种媒体的媒体。如纸张、磁带、磁盘、光盘等；（5）传输媒体（Transmission medium）指的是用于传输某些媒体的媒体。常用的有如电话线、电缆、光纤等[56]。

多媒体技术就是融计算机、声音、文本、图像、动画、视频和通信等多种功能于一体的技术，它借助日益普及的高速信息网，可实现计算机的全球联网和信息资源共享，并且它给传统的计算机系统、音频和视频设备带来了方向性的变革，将对大众传媒产生深远的影响。因此多媒体将加速计算机进人家庭和社会各个方面的进程，给人们的工作、生活和娱乐带来深刻的革命。多媒体技术涉及的内容包括：

- 多媒体数据压缩：多模态转换、压缩编码；
- 多媒体处理：音频信息处理，如音乐合成、语音识别、文字与语音相互转换；图像处理，虚拟现实；
- 多媒体数据存储：多媒体数据库；
- 多媒体数据检索：基于内容的图像检索，视频检索；

- 多媒体著作工具：多媒体同步、超媒体和超文本；
- 多媒体通信与分布式多媒体：CSCW、会议系统、VOD 和系统设计；
- 多媒体专用设备技术：多媒体专用芯片技术，多媒体专用输入输出技术；
- 多媒体应用技术：CAI 与远程教学，GIS 与数字地球、多媒体远程监控等[57]。

8.2 多媒体技术发展历史

- 启蒙发展阶段

多媒体技术的一些概念和方法，起源于 20 世纪 60 年代。1965 年，纳尔逊（Ted Nelson）为计算机处理文本文件提出了一种把文本中遇到的相关文本组织在一起的方法，并为这种方法杜撰了一个词，称为“hypertext（超文本）”。与传统的方式不同，超文本以非线性方式组织文本，使计算机能够响应人的思维以及能够方便地获取所需要的信息。万维网（WWW）上的多媒体信息正是采用了超文本思想与技术，组成了全球范围的超媒体空间。

多媒体技术实现于 20 世纪 80 年代中期。1984 年美国 Apple 公司在研制 Macintosh 计算机时，为了增加图形处理功能，改善人机交互界面，创造性地使用了位映射（bitmap）、窗口（window）、图符（icon）等技术。这一系列改进所带来的图形用户界面（GUI）深受用户的欢迎，加上引入鼠标作为交互设备，配合 GUI 使用，大大方便了用户的操作。Apple 公司在 1987 年又引入了“超级卡”（Hypercard），使 Macintosh 机成为更容易使用、易学习并且能处理多媒体信息的机器，受到计算机用户的一致赞誉。

- 标准化阶段

自 20 世纪 90 年代以来，多媒体技术逐渐成熟。多媒体技术从以研究开发为重心转移到以应用为重心。

由于多媒体技术是一种综合性技术，它的实用化涉及到计算机、电子、通信、影视等多个行业技术协作，其产品的应用目标，既涉及研究人员也面向普通消费者，涉及各个用户层次，因此标准化问题是多媒体技术实用化的关键。在标准化

阶段，研究部门和开发部门首先各自提出自己的方案，然后经分析、测试、比较、综合，总结出最优、最便于应用推广的标准，指导多媒体产品的研制。

静态图像的一个标准，是国际电信联盟（ITU）的 T.81。静态图像的主要标准称为 JPEG 标准（ISO/IEC10918）。它是 ISO 和 IEC 联合成立的专家组 JPEG（Joint Photographic Experts Group）建立的适用于单色和彩色、多灰度连续色调静态图像国际标准。该标准在 1991 年通过，成为 ISO/IEC10918 标准，全称为“多灰度静态图像的数字压缩编码”。

视频/运动图像的主要标准是国际标准化组织（ISO）下属的一个专家组 MPEG（Moving Picture Experts Group）制定的 MPEG-1（ISO/IEC11172）、MPEG-2（ISO/IEC13818）和 MPEG-4（ISO/IEC14496）三个标准。与 MPEG-1、4 等效的国际电信联盟（ITU）标准，在运动图像方面有用于视频会议的 H.261（Px64）、用于可视电话的 H.263。

在多媒体数字通信方面（包括电视会议等）制定了一系列国际标准（表 01-03-2），称为 H 系列标准。这个系列标准分为两代。H.320、H.321 和 H.322 是第一代标准，都以 1990 年通过的 ISDN 网络上的 H.320 为基础。H.323、H.324 和 H.310 是第二代，使用新的 H.245 控制协议并且支持一系列改进的多媒体编、解码器。

更深层次的多媒体技术标准也开始推出或列入开发中。一个典型的标准是称作“多媒体内容描述接口”的 MPEG-7 标准（ISO/IEC15938）。与已经推出的几个 MPEG 标准不同，MPEG-7 是一个关于表示音/视信息的标准。它的七个组成部件中，系统、描述定义语言（DDL）、视频、音频和多媒体描述方案等已经成为正式标准，参考软件和一致性测试则计划在 2002 年 9 月成为标准。

- 蓬勃发展时期

随着多媒体各种标准的制定和应用，极大地推动了多媒体产业的发展。很多多媒体标准和实现方法（如 JPEG、MPEG 等）已被做到芯片级，并作为成熟的商品投入市场。与此同时，涉及到多媒体领域的各种软件系统及工具，也如雨后春笋，层出不穷。这些既解决了多媒体发展过程必须解决的难题，又对多媒体的普及和应用提供了可靠的技术保障，并促使多媒体成为一个产业而迅猛发展。

代表之一是进一步发展多媒体芯片和处理器。1997 年 1 月美国 Intel 公司推出了具有 MMX 技术的奔腾处理器（Pentium processor with MMX），使它成为多媒体计算机的一个标准。奔腾处理器在体系结构上有三个主要的特点：（1）增加了新的指令，使计算机硬件本身就具有多媒体的处理功能（新添 57 个多媒体指令集），能更有效地处理视频、音频和图形数据。（2）单条指令多数据处理（SIMD, Single Instruction Multiple Dataprocess）减少了视频、音频、图形和动画处理中常有的耗时的多循环。（3）更大的片内高速缓存，减少了处理器不得不访问片外低速存储器的次数。奔腾处理器使多媒体的运行速度成倍增加，并已开始取代一些普通的功能卡板。

随着网络电脑（Internet PC、NC）及新一代消费性电子产品，如电视机顶盒（Set-Top Box）、DVD、视频电话（Video Phone）、视频会议（Video Conference）等观念的崛起，强调应用于影像及通讯处理上最佳的数字信号处理器（DSP），经过另一番的结构包装，可由软件驱动组态的方式，进入咨询及消费性的多媒体处理器市场。

现在多媒体技术及应用正在向更深层次发展。下一代用户界面，基于内容的多媒体信息检索，保证服务质量的多媒体全光通信网，基于高速互联网的新一代分布式多媒体信息系统等等,多媒体技术和它的应用正在迅速发展，新的技术、新的应用、新的系统不断涌现[58]。

8.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为多媒体领域全球学者分布情况。

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布在其东西海岸；亚洲东部也有较多的人才分布；欧洲的人才主要集中在欧洲中西部；其他诸如非洲、南美洲等地区的学者非常稀少；多媒体领域的人才分布与各地区的科技、经济实力情况大体一致。



图 8-1 多媒体领域全球学者分布

此外，在性别比例方面，多媒体领域中男性学者占比 91.7%，女性学者占比 8.3%，男性学者占比远高于女性学者。

多媒体领域学者的 h-index 分布如下图所示，大部分学者的 h-index 分布在中低区域，其中 h-index 在 20-30 区间的人数最多，有 691 人，占比 34.2%，50-56 区间的人数最少，有 124 人。

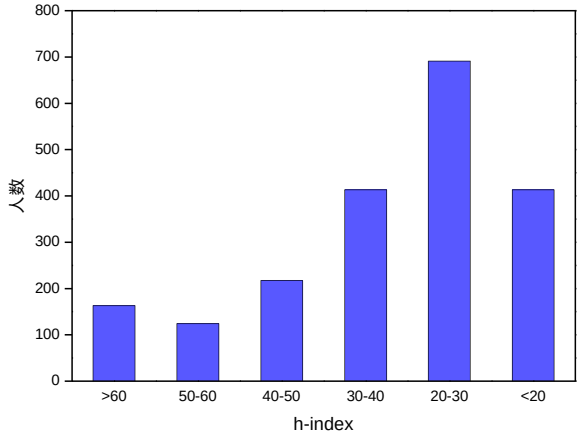


图 8-2 多媒体领域学者 h-index 分布

- 中国人才分布

我国专家学者在多媒体领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通

过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等亚洲国家相比，中国在多媒体领域学者数量较多且有一定的优势。



图 8-3 多媒体领域中国学者分布

中国与其他国家在多媒体领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 8-1 多媒体领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	676	18348	27	1107
中国-新加坡	231	6827	30	364
中国-澳大利亚	101	2919	29	166
中国-英国	71	1315	19	143
中国-加拿大	63	1261	20	117
中国-意大利	24	636	27	34
中国-巴基斯坦	17	412	24	23
中国-荷兰	17	252	15	34
中国-德国	16	817	51	39
中国-法国	12	271	23	31

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在多媒体领域合作之密切；同时，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 5 席；中国与德国合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中德合作达到了较高的水平。

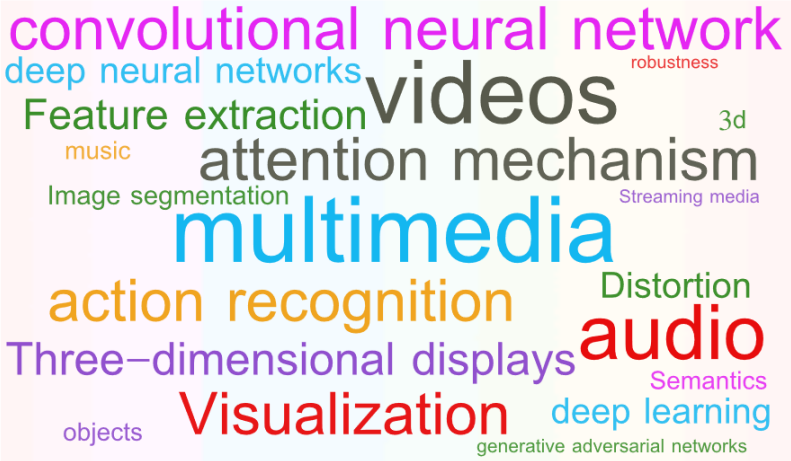
8.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

ACM International Conference on Multimedia

IEEE Transactions on Multimedia

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，多媒体（multimedia）、视频（videos）、音频（audio）是本领域中最热的关键词。



论文题目：*Beyond Narrative Description: Generating Poetry from Images by Multi-Adversarial Training*

中文题目：超越叙事描述:通过多重对抗训练，从意象中生成诗歌

论文作者：Bei Liu，Jianlong Fu，Makoto P. Kato，Masatoshi Yoshikawa

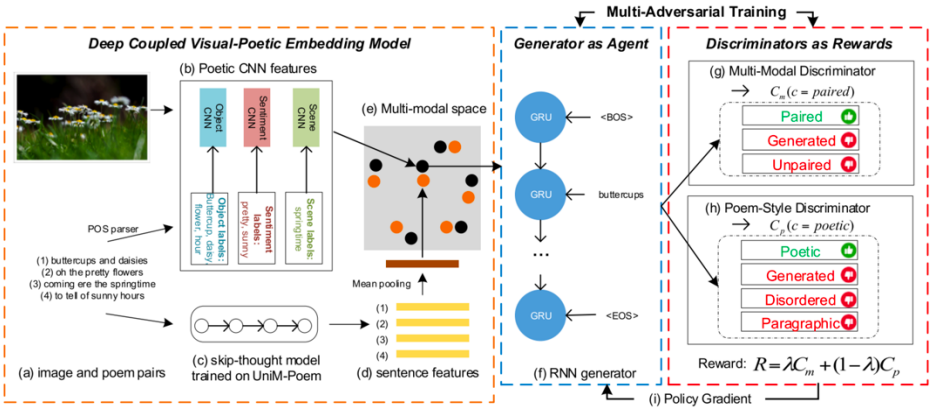
论文出处：26th ACM International Conference on Multimedia – ACMMM'18

论文地址：https://arxiv.org/pdf/1804.08473v4.pdf

研究问题：

本文主要研究了从图像自动生成诗歌的方法。这项任务涉及多个挑战，包括从图像中发现诗意线索（例如，从绿色中获得希望），以及生成满足图像相关性和语言水平的诗意的诗歌。

研究方法：



本文通过使用策略梯度的多对抗训练将诗歌生成的任务划分为两个相关的子任务，从而确保跨模式的关联性和诗歌语言风格。为了从图像中提取诗意线索，提出深度耦合的视觉诗意嵌入，可以同时学习图像中的对象，情感和场景的诗歌表达。进一步引入两个判别网络来指导诗歌的产生，包括多模式判别器和诗歌风格判别器。

如图，为了更好地从图像中学习诗歌线索以产生诗歌，首先从多模式诗歌数据集中学习具有图像 CNN 特征和具有 skip-thought 向量特征的深度耦合视觉诗词嵌入模型。然后，使用该嵌入模型从诗歌语料库 UniM-Poem 中检索相关且多样的诗歌。这些检索到的诗歌和 MultiM-Poem 的图像一起构成了一个放大的图像-诗对数据集 MultiM-Poem（Ex）。

进一步利用最新的顺序学习技术来训练端到端的图像到诗歌模型。为避免因为长序列的长度过长（所有诗行在一起）而导致的曝光偏差问题，以及没有可用于评分生成的诗的特定损失的问题，使用递归神经网络（RNN）来生成多首诗，并采用对抗训练，通过政策梯度进一步优化。

研究结果：

本文发布了人类注释的具有两个不同属性的两个诗歌数据集：1）第一个人类注释的图像-诗歌对数据集（共有 8 对 292 对）；2）迄今为止最大的公共英语诗歌语料库数据集（共包含 92,265 种不同的诗歌）。在 MultiM-Poem，UniM-Poem 和 MultiM-Poem（Ex）上进行实验，以生成图像诗歌。生成的诗歌以客观和主观两种方式进行评估。其中定义了有关相关性、新颖性和翻译一致性的自动评估指标，并就生成的诗歌的相关性、连贯性和想象力进行了用户研究，本文提出的模型在客观评估和主观评估中都显示出优于从图像生成诗歌的最新方法的优异性能。对 500 多个人类受试者进行的图灵测试，其中 30 名评估者是诗歌专家，证明了方法的有效性。

论文题目：*Audiovisual Zooming: What You See Is What You Hear*

中文题目：视听缩放：你所看到的就是你所听到的

论文作者： Arun Asokan Nair， Austin Reiter，Changxi Zheng，Shree Nayar

论文出处：27th ACM International Conference on Multimedia – ACMMM’19

论文地址：https://dl.acm.org/citation.cfm?id=3351010

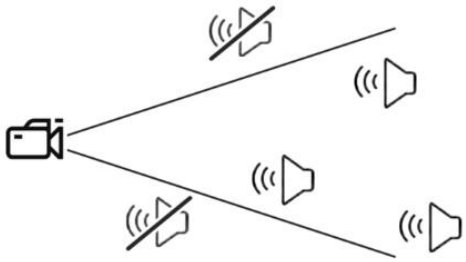
研究问题：

在移动平台上捕获视频时，感兴趣的目标通常会被周围的环境污染。为了消除视觉上的不相关性，摄像机平移和缩放提供了隔离所需视场的方法。但是，捕获的音频仍然被 FOV 外部的信号污染。这种效果是不自然的。本文提出了视听缩放的概念，从而形成听觉场以匹配视觉。框架建立在经典的波束成形概念之上，波束成形是一种使用麦克风阵列从单个方向增强声音的计算方法。但是，波束成形本身不能包含听觉 FOV，因为 FOV 可能包括任意数量的定向源。本文将视听缩放公式化为广义特征值问题，并提出了一种在移动平台上进行有效计算的算法。

研究方法：

在这项工作中，增加了麦克风阵列和波束成形方法，以实现视听缩放，而无需从训练数据中学习。受麦克风阵列波束成形的激励，可将单个麦克风采样的信号视为某些潜在随机过程的随机变量。从这个角度出发，本文在频域中估计了两个复值矩阵，称为频谱矩阵：一个描述来自 FOV 内部的麦克风信号的自相关和

互相关，另一个描述来自 FOV 外部的信号。使用这两个矩阵，可以将增强向 FOV 的问题公式化为可以在移动设备上轻松解决的广义特征值问题。为了分析本方法，本文得出了频谱矩阵估计的理论误差范围，并揭示了误差残差与经典最小方差无失真响应（MVDR）波束形成器性能的联系。根据经验，通过仿真，了解各种设计参数如何影响麦克风阵列。这些推论为实验提供了信息。最终算法很简单，可以轻松地部署在移动设备上。



研究结果：

本文提供了算法组件的理论分析以及用于理解麦克风阵列各种设计选择的数值研究。最后，在两个不同的移动平台上演示了视听缩放：移动智能手机和用于视频会议设置的 360°球形成像系统。

论文题目：*Emotion Recognition in Speech using Cross-Modal Transfer in the Wild*

中文题目：在语音识别中应用跨模态传输进行情感识别

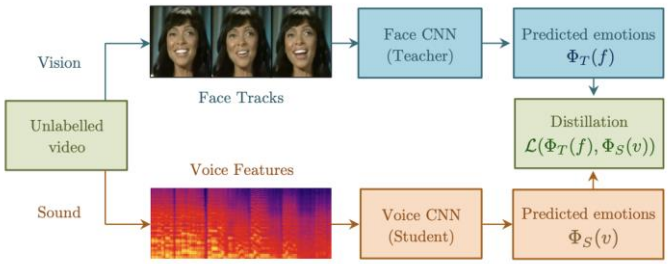
论文作者：Samuel Albanie, Arsha Nagrani, Andrea Vedaldi, Andrew Zisserman

论文出处: MM2018

论文地址：https://arxiv.org/pdf/1808.05561.pdf

研究问题：

传统的语音情感识别往往需要大量的标注数据，与此同时，数据的标注往往是模糊不清的，有大量噪声。本文基于这样一个简单的假设: 人在说话时脸像的情感和语音情感是一致的，利用这种一致性，本文提出了一种 teacher-student 模型，将视觉的情感信息迁移到语音，取得了很好的实验结果。



研究方法：

本文提出了一种基于 teacher-student 的 cross-modal transfer 模型，将视觉上的脸像情感信息迁移到语音。对于一段无标注的视频，视觉上脸像的预测结果为音频的预测结果提供了 soft-label，相比于人手工标注的 hard label，teacher-student 的监督模式极大地提升了语音情感识别的性能。

研究结果：

本文已经证明了使用大数据集的情绪未标记的视频的价值，情绪的跨模态转移从脸到语言。这种方法的好处是显而易见的，语音情感模型在标准的基准上获得了合理的分类性能，结果远远高于随机。作者还在 FERPlus 基准（supervised）上实现了面部情绪识别的最新性能，并在两个标准数据集（RML 和 eNTERFACE）上为语音情绪识别的交叉模式蒸馏方法设置了基准。这种方法的最大优点是视频数据几乎是无限的，可以从 YouTube 和其他来源免费获得。

论文题目：*Multi-View Image Generation from a Single-View*

中文题目：从单视图生成多视图图像

论文作者：Bo Zhao, Xiao Wu, Zhi-Qi Cheng, Hao Liu, Zequn Jie, Jiashi Feng

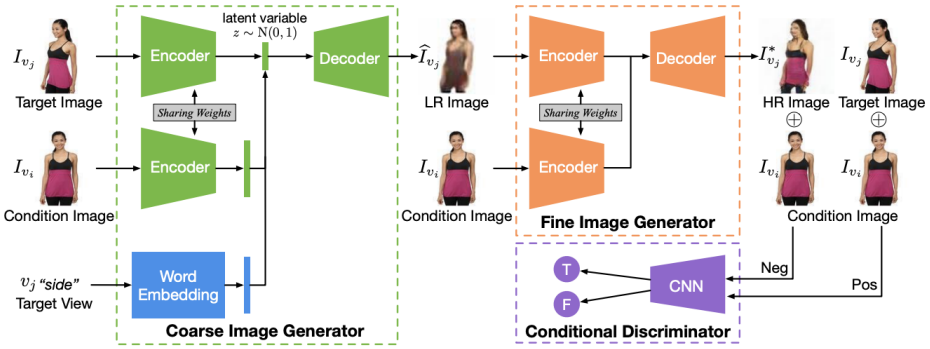
论文出处：MM2018

论文地址：https://arxiv.org/pdf/1704.04886.pdf

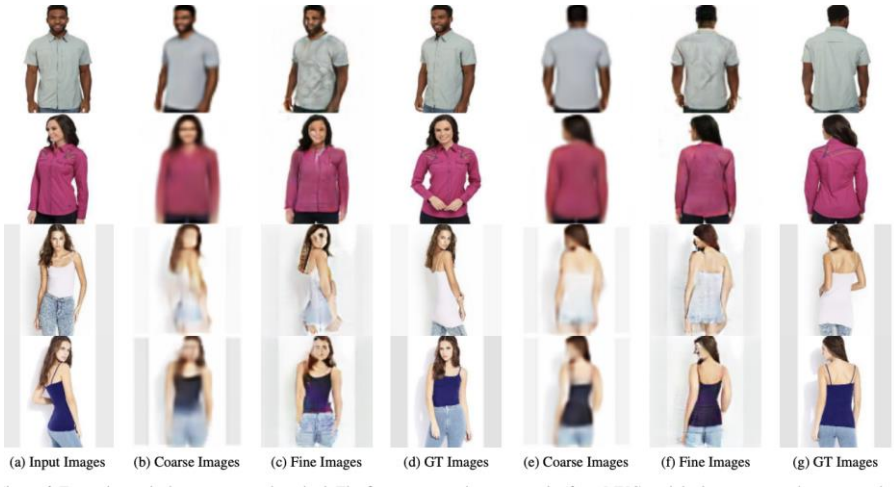
研究问题：

本文旨在解决利用一个视角的图像合成多视角图像的问题，这一问题极具挑战性。作者提出了将变分推断和对抗生成网络结合的 VariGAN，使用了一种从粗到细的方式，首先生成低分辨率的对应视角图片，再利用对抗训练的方法提升图

像分辨率。该方法在多个数据集上取得了很好的结果。



研究方法：



为了解决单视角图片合成多视角图片的问题，作者提出了 VariGAN，具体如下：（1）粗粒度生成器首先用变分推断的方式生成对应视角的图像，变分自编码器以原始图像，目标视角为输入，输出低分辨率图像，并用变分自编码器对应的损失监督训练。（2）细粒度的模型将低分辨率的图像精细化，用对抗训练的损失作为监督，生成了更加细粒度的图像。在具体实现中，作者设计了双路径 U-Net 来实现细粒度生成器，最终在 MVC 和 Deepfashion 数据集上取得了较好的结果。

研究结果：

本文提出了一种变分生成对抗网络（VariGANs）来合成以不同视角为输入图像的现实服装图像。该方法利用变分推理的方法，实现了由粗到精的图像生成。具体地说，提供具有一定视图的输入图像，粗图像生成器首先生成具有目标视图

的对象的基本形状。然后用精细图像生成器将细节信息填充到粗图像中，并对缺陷进行修正。通过大量的实验，该模型可以得到比现有方法更合理的结果。消融研究也证实了所提出的静脉曲张各组成部分的重要性。

论文题目：*Understanding the Teaching Styles by an Attention based Multi-task Cross-media Dimensional modelling*

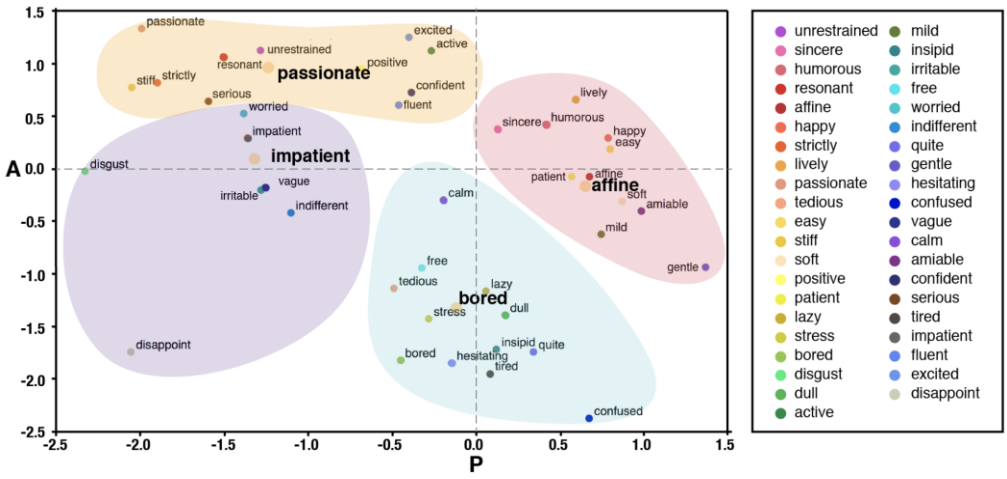
中文题目：通过一个基于注意力的多任务跨媒体维度建模来理解教学风格

论文作者：Suping Zhou，Jia Jia，Yufeng Yin，Xiang Li，Yang Yao，Ying Zhang，Zeyang Ye，Kehua Lei，Yan Huang，Jialie Shen

论文出处：27th ACM International Conference on Multimedia – ACMMM'19

论文地址：https://dl.acm.org/citation.cfm?id=3351059

研究问题：



教师的授课风格在帮助学生更好学习中起着重要的作用。文章探讨了如何有效的分析理解教师的授课风格。具体来说，文章研究了 1）如何定量描述不同教师的教学风格；2）如何对教师的跨模态教学数据（言语、面部表情和身体动作、内容等）与教学风格之间的关系进行建模。

研究方法：

首先，基于愉悦度激活度维度理论，文章构建了一个二维教学风格语义空间（TSSS），对教学风格进行定量、全面的描述。从好未来教育集团提供的 10000 多份家长对教师评价的反馈问卷中，筛选出了最常用的 41 个授课风格形容词，

并在 TSSS 上手工标注授课风格形容词的坐标，即愉悦度激活度值。

同时，文章提出了一种基于注意力机制的多路径多任务深度神经网络（AMMDNN），该神经网络能够准确、可靠地捕捉跨模态特征与 TSSS 之间的内在联系。文章利用从好未来教育集团收集的 4541 句跨模态教学数据集，设计了大量的测试实验来评估 AMMDNN 对跨模态特征和 TSSS 上坐标值之间的映射效果。

研究结果：

结果表明，所提出的 AMMDNN 模型性能均优于基线模型（对于 CCC 一致性相关系数指标，平均提升 0.0842）。同时，文章将二维坐标与教学风格形容词联系起来，可以更合理、更生动地描述教学风格。最后，文章还进行了一些有趣的案例研究，包括不同教师和课程之间的教学风格比较，以及利用所提出的方法进行教学质量分析。

论文题目：*Dance with Melody: An LSTM-autoencoder Approach to Music-oriented Dance Synthesis*

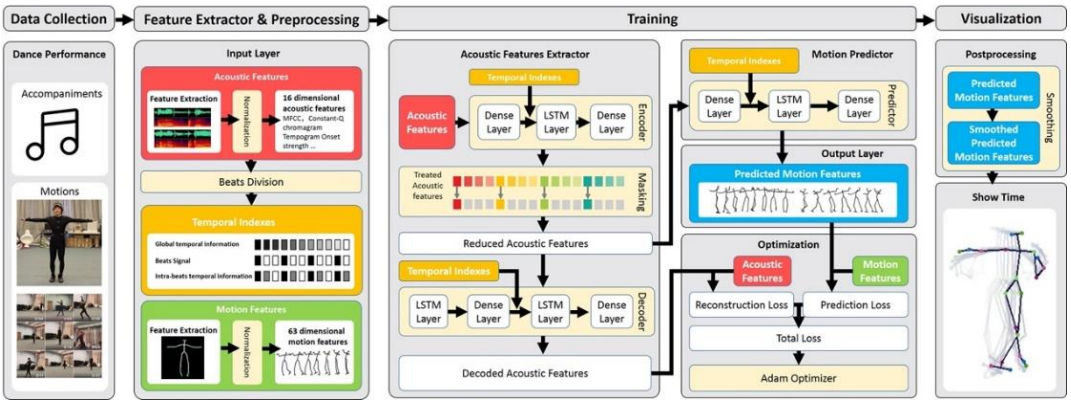
中文题目：与旋律共舞：一种面向音乐的舞蹈合成的 LSTM-autoencoder 方法

论文作者： Taoran Tang，Jia Jia，Hanyang Mao

论文出处：26th ACM International Conference on Multimedia – ACMMM'18

论文地址：https://dl.acm.org/citation.cfm?id=3240526

研究问题：



舞蹈深受音乐的影响。对音乐舞蹈编舞的综合研究可以促进舞蹈教学和人类行为研究等多个领域的研究。尽管在研究音乐和舞蹈之间的关系方面已经付出了相当大的努力，但是基于音乐的舞蹈编排的综合仍然是一个开放的问题。目前该问题主要有两个挑战：1）如何根据音乐选择合适的舞姿，即技术舞蹈手册中命名和指定的舞步组；2）如何根据音乐艺术地加强舞蹈编排。

研究方法：

为了解决这些问题，文章提出了一种基于长短时记忆网络（LSTM）-自动编码器的模型来解决面向音乐的舞蹈自动编排方法。具体地说，该模型旨在提取声音和运动特征之间的映射，以便音乐的情感将被合成的舞蹈所反映。因此，该模型可以学习舞者如何调整自己的局部关节姿势和动作节奏，以表达音乐情感的变化，以及舞蹈编排中的动作选择规则。

由于缺乏模型训练所需的数据，文章构建了一个音乐舞蹈数据集，包含 40 个四种舞蹈类型的完整舞蹈编排，共 907200 帧，用光学运动捕捉设备（Vicon）采集。该数据集还记录了舞蹈编排所用的音乐，使收集到的数据特别有助于面向音乐的舞蹈合成。提取了包括 63 维运动特征、16 维声学特征和 3 维时间指标在内的综合特征。从而为神经网络的训练提供了有针对性、精确性和完整性的特征。目前这是最大的音乐舞蹈数据集，且该文章公开了数据集。

研究结果：

文章进行了一些定性和定量实验来量化模型性能。由于舞蹈是一种艺术创作，除了考虑常用的欧几里德损失函数外，还考虑了用户评价来评价模型的表现。实验结果表明，与多个基线模型相比，文章所提出的模型成功地提取了与舞姿选择相关的声学特征，在选择与一段动画片的长度、节奏和情感相匹配的舞姿时表现良好。

论文题目：*MMGCN：Multi-modal Graph Convolution Network for Personalized Recommendation of Micro-video*

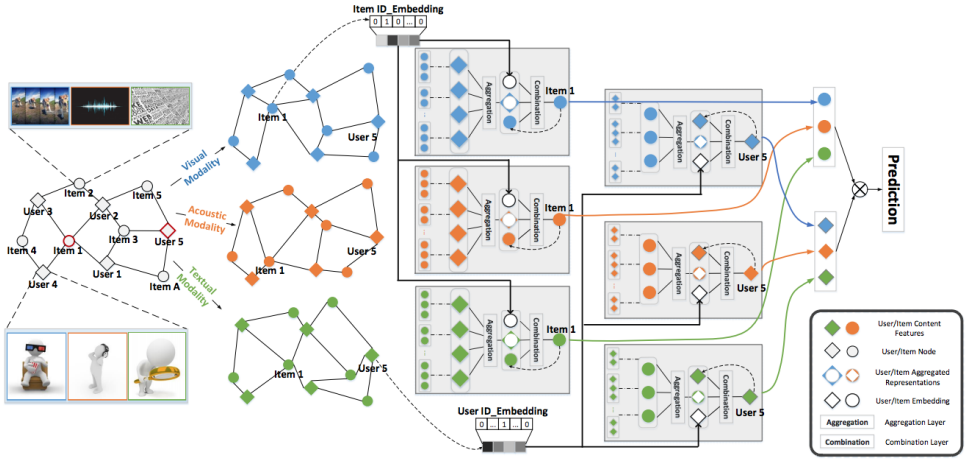
中文题目：MMGCN：用于微视频个性化推荐的多模态图卷积网络

论文作者：Yinwei Wei，Xiang Wang，Liqiang Nie，Xiangnan He，Richang Hong，Tat-Seng Chua

论文出处：Proceedings of the 27th ACM International Conference on Multimedia

论文地址：https://dl.acm.org/citation.cfm?id=3351034

研究问题：



个性化推荐在许多在线内容共享平台中起着核心作用。为了提供优质的微视频推荐服务，重要的是考虑用户与项目（即短视频）之间的交互以及来自各种模态（例如视觉，听觉和文本）的项目内容。现有的多媒体推荐作品在很大程度上利用多模态内容来丰富项目表示，而为利用用户和项目之间的信息交换来增强用户表示并进一步捕获用户对不同模式的细粒度偏好所做的工作却较少。本文中利用用户-项目交互来指导每种模式中的表示学习，并进一步个性化微视频推荐。基于图神经网络的消息传递思想设计了一个多模态图卷积网络（MMGCN）框架，该框架可以生成用户和微视频的特定模态表示，以更好地捕获用户的偏好，如上图所示。

研究方法：

在每个模态中构造一个 user-item 二部图，并用其邻居的拓扑结构和特征丰富每个节点的表示。用户的历史记录反应了个人的兴趣，用户群组也可以对项目进行建模。多模态图卷积网络（MMGCN）包括三个部分：聚合层、组合层、预测层。聚合层对于每个模态（如视频），将交互过项目的相应内容（如帧）通过聚合函数来衡量邻居的影响，将结构信息和邻居分布编码得到用户的表示；组合

层融合了结构信息、启发式信息、模态关联，通过用户群组来增强项目的表示。通过递归地进行这种聚合和组合，用户和项目的表示可以体现多跳邻居的信息，使得用户关于特定模态的偏好可以被很好的表示。最后，对于未知交互的预测可以通过用户和微视频表示的相似度来得到。

研究结果：

在三个公开可用的数据集 Tiktok，Kwai 和 MovieLens 上进行的大量实验，证明了论文提出的模型能够明显优于目前最新的多模态推荐方法。在本文中，作者明确地建模特定于模式的用户偏好来增强微视频推荐。作者设计了一个新的基于 gcn 的框架，称为 MMGCN，以利用用户和微视频之间在多种模式下的信息交换，细化它们的特定模式表示，并进一步模拟用户对微视频的细粒度偏好。在三个公开的微视频数据集上的实验结果很好地验证了该的模型。此外，作者还可视化了一些示例来演示特定于模式的用户首选项

论文题目：*Routing Micro-videos via A Temporal Graph-guided Recommendation System*

作文题目：通过临时图形引导推荐系统路由微视频

论文作者：Yongqi Li, Meng Liu, Jianhua Yin, Chaoran Cui, Xin-Shun Xu, Liqiang Nie

论文出处：Proceedings of the 27th ACM International Conference on Multimedia

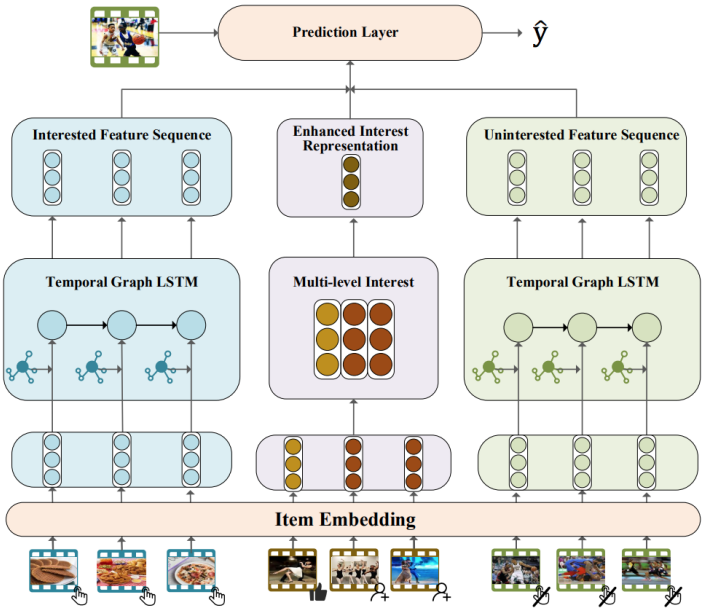
论文地址：https://dl.acm.org/citation.cfm?id=3350950

研究问题：

在过去的几年中，短视频已成为社交媒体时代的主流趋势。同时，随着短视频数量的增加，用户经常被他们不感兴趣的视频所淹没。尽管现有的针对各种社区的推荐系统已经取得了成功，但由于短视频平台中的用户具有其独特的特征：多样化的动态兴趣，多层次的兴趣以及负样本，因此它们无法以一种好的方式应用于短视频。为了解决这些问题，论文提出了一个时间图指导的推荐系统（ALPINE）。首先设计了一个新颖的基于图的顺序网络，能够同时对用户的动态兴趣和多样化兴趣进行建模。同样，可以从用户的真实负样本中捕获不感兴趣

的信息。除此之外，通过用户矩阵将用户的多层次兴趣引入推荐模型，该矩阵能够学习用户兴趣的增强表示。最后，系统可以通过考虑上述特征做出准确的推荐。

研究方法：



如上图所示，ALPINE 模型由三个部分组成：基于图的顺序网络、多层次兴趣建模层、预测层。基于图的顺序网络将兴趣图与长短时记忆网络结合，利用用户的点击和未点击历史行为，分别提取动态的、多样的感兴趣和不感兴趣的特征序列。多层次兴趣建模层利用用户的关注、点赞等行为，对用户兴趣的表示进行增强。最后，基于用户的感兴趣特征序列、不感兴趣特征序列、增强的兴趣表示，分别将未知短视频的表示与相应特征通过 Vanilla Attention 得到增强的表示，再通过多层感知机得到相应的评分，将三个评分加权得到最终的预测输出。

研究结果：

在两个公共数据集上进行的实验，证明了提出的 ALPINE 模型较其他推荐算法取得了更好的性能，也证明了模型各组成部分的有效性。在这项工作中，作者提出了一个基于时间图的 LSTM 模型来智能地将微视频路由到目标用户。为了捕获用户动态的、多样的兴趣，作者将用户的历史交互序列编码到时序图中，设计了一种基于时序图的 LSTM 模型。由于不同的交互反映了不同程度的兴趣，作者构建了多层次的兴趣建模层来增强用户的兴趣表示。此外，该模型从真实的阴

性样本中提取不感兴趣的信息，以提高推荐性能。为了验证该方案，作者在两个公共数据集上进行了大量的实验，实验结果证明了该模型的有效性。

8.5 多媒体技术进展

近年来，随着数字化技术的发展，多媒体技术突飞猛进，音视频技术是当前最活跃、发展最迅速的高新技术领域之一。多媒体分析以文本、图像、声音、视频等多种不同类型媒体的数据为研究对象，主要的研究目的一方面是使计算机具备人类的多媒体（如视、听）理解能力，另一方面是从多媒体数据中挖掘信息和知识、帮助人类更好地理解世界。

多媒体技术研究领域包括多媒体信息处理、多媒体数据压缩编码、多媒体内容分析与检索技术、多媒体交互与集成、多媒体通信与网络、多媒体内容安全、多媒体系统与虚拟现实等。在近几年的研究中，多媒体技术呈现出与计算机体系结构、计算机网络、人机交互、信息安全、社会网络等多学科交叉融合的发展趋势。

近两年多媒体领域研究热点主要集中在大规模图像视频分析、社会媒体研究、多模态人机交互、计算视觉、计算图像、实时视频流化等方面。

由于多媒体数据往往是多种信息的传递媒介（例如一段视频中往往会同时使得文字信息、视觉信息和听觉信息得到传播），多模态学习已逐渐发展为多媒体内容分析与理解的主要手段。

在计算图像方面，大规模数据集的构建仍是一个热点研究方向，尤其语义对象的像素级标注需求越来越强烈，能够人机交互标注的过程中不断学习的协同标注方法得到了广泛关注。

无监督学习是多媒体数据分析的长远目标。目前很多领域拥有大量的数据，但是这些数据都是没有经过标记的。因此除了基本的数据勘探和异常检测场景，这些数据基本无法使用。近期在使用未标记的数据来改进（标记数据）监督学习过程方面已经取得了许多进展。

此外自动机器学习（AutoML）和元学习（Meta Learning）的最新研究成果及其在多媒体上的应用也逐渐增多。

在图像压缩处理方面，也有一些研究工作将深度学习用于图像或视频压缩后处理，并得到了一定的效果。然而，现有工作的一个主要问题是用于后处理的深度网络较为复杂，计算速度慢，不满足实际应用的需求。如何在处理效果和处理速度之间取得一个折中，是压缩后处理的一个主要挑战。

图神经网络（Graph Neural Network, GNN）在多媒体领域的应用是近两年的热点研究方向，应用场景包括：个性化推荐，如基于多模态图卷积网络（MMGCN）的多模态推荐方法；短视频推荐，如使用基于图的顺序网络进行建模；多视频摘要，如采用图卷积网络衡量每个视频的重要性和相关性；基于文本的行人搜索，如使用深度对抗图注意力卷积网络（A-GANet）利用文本和视觉场景图学习联合特征空间；视频关系监测，如使用转移图神经网络（DoT-GNN）解决图像外观变化的问题。

随着 Mask-RCNN 与 RetinaNet 的发展，物体检测研究日趋成熟，但即便如此，就应用而言，当前的技术依然存在诸多缺陷，为此，针对现代目标检测的基本框架（backbone、head、scale、batchsize 与 post-processing），神经网络架构搜索（NAS）以及细粒度图像分析（FGIA）等 3 个方面的潜在难题成为主要研究内容，尤其是后两者，将成为未来视觉物体检测的两个重要研究维度。

9 人机交互技术

9.1 人机交互概念

人机交互（Human-Computer Interaction, HCI），是人与计算机之间为完成某项任务所进行的信息交换过程，是一门研究系统与用户之间的交互关系的学问。系统可以是各种各样的机器，也可以是计算机化的系统和软件。人机交互界面通常是指用户的可见部分，用户通过人机交互界面与系统交流，并进行操作。人机交互技术是计算机用户界面设计中的重要内容之一，与认知学、人机工程学、心理学等学科领域有密切的联系。

目前关于人机交互的定义主要有三种：一是 ACM（Association for Computing Machinery）的观点，它将人机交互定义为：有关交互计算机系统设计、评估、实现以及与之相关现象的学科；二是伯明翰大学教授 AlanDix 的观点：他认为人机交互是研究人、计算机以及他们之间相互作用方式的学科，学习人机交互的目的是使计算机技术更好地为人类服务；三是宾夕法尼亚州立大学 JohnM.Carroll 的观点：他认为人机交互指的是有关可用性的学习和实践，是关于理解和构建用户乐于使用的软件和技术，并能在使用时发现产品有效性的学科。无论是哪一种定义方式，人机交互所关注的首要问题都是人与计算机之间的关系问题。

人机交互技术的发展与国民经济发展有着直接的联系，它是使信息技术融入社会，深入群体，达到广泛应用的技术门槛。任何一种新交互技术的诞生，都会带来其新的应用人群，新的应用领域，带来巨大的社会经济效益，从企业的角度，改善人机交互能够提高员工的生产效率；学习人机交互能够降低产品的后续支持成本。从个人的角度，可以帮助用户有效地降低错误发生的概率，避免由于错误引发的损失。在现代和未来的社会里，只要有人利用通信、计算机等信息处理技术进行社会活动时，人机交互都是永恒的主题，鉴于它对科技发展的重要性，人机交互是现代信息技术、人工智能技术研究的热门方向[59]。

9.2 人机交互发展历史

人机交互的发展历史，是从人适应计算机到计算机不断地适应人的发展史。交互的信息也由精确的输入输出信息变成非精确的输入输出信息。

9.2.1 简单人机交互

由于受到制造技术和成本等原因限制，早期的人机交互在设计上较少考虑人的因素，强调输入输出信息的精确性，使用不够自然和高效[60]。

- 早期的手工作业

这个时期交互的特点是由设计者（或本部门同事）来使用计算机，他们采用手工操作和依赖机器（二进制机器代码）的方法去适应计算机。

- 作业控制语言及交互命令语言

这一阶段特点是计算机的主要使用者—程序员可采用批处理作业语言或交互命令语言的方式和计算机打交道，虽然要记忆许多命令和熟练地敲键盘，但已可用较方便的手段来调试程序、了解计算机执行情况。

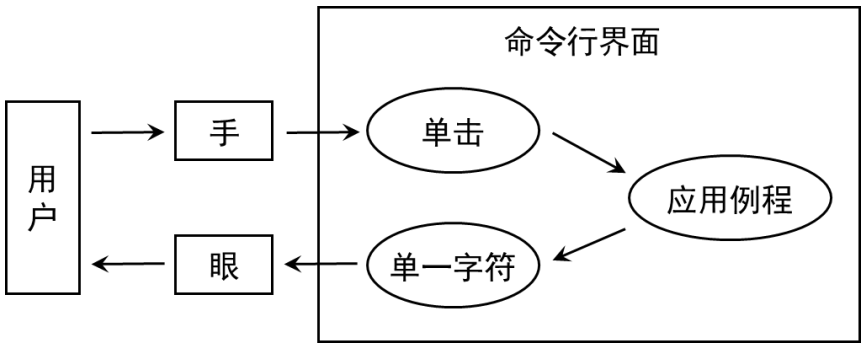


图 9-1 命令行界面概念模型

- 图形用户界面（GUI）

图形用户界面的主要特点是桌面隐喻、WIMP（Window / Icon / Menu / Pointing Device）技术、直接操纵和“所见即所得（WYSIWYG）”。由于 GUI 简明易学、减少了敲键盘、实现了“事实上的标准化”。因而使不懂计算机的普通用户也可以熟练地使用，开拓了用户人群。它的出现使信息产业得到空前的发展。

- 网络用户界面

以超文本标记语言 HTML 及超文本传输协议 HTTP 为主要基础的网络浏览器是网络用户界面的代表。由它形成的万维网（World Wide Web, WWW）已经成为当今 Internet 的支柱。这类人机交互技术的特点是发展快，新的技术不断出现，如搜索引擎、网络加速、多媒体动画、聊天工具等。

9.2.2 自然人机交互

随着网络的普及性发展和无线通讯技术的发展，人机交互领域面临着巨大的挑战和机遇，传统的图形界面交互已经产生了本质的变化，人们的需求不再局限于界面美学形式的创新，用户更多的希望在使用多媒体终端时，有着更便捷、更符合他们使用习惯，同时又有着比较美观的操作界面。利用人的多种感觉通道和动作通道（如语音、手写、姿势、视线、表情等输入），以并行、非精确的方式与（可见或不可见的）计算机环境进行交互，使人们从传统的交互方式的束缚解脱出来，使人们进入自然和谐的人机交互时期。这一时期的主要研究内容包括：多通道交互、情感计算、自然语言理解、虚拟现实、智能用户界面等方面。

- 多通道交互

多通道交互（Multi Modal Interaction，MMI）是近年来迅速发展的一种人机交互技术，它既适应了“以人为中心”的自然交互准则，也推动了互联网时代信息产业（包括移动计算、移动通信、网络服务器等）的快速发展[61]。

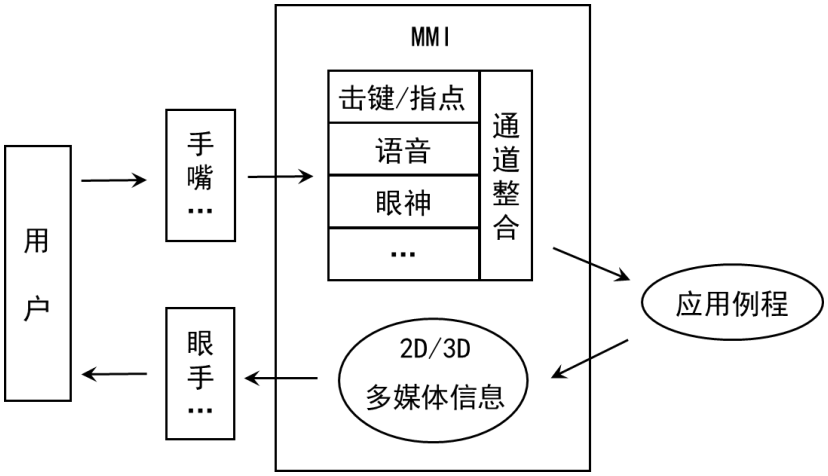


图 9-2 命令行界面概念模型

MMI 是指“使用多种通道与计算机通信的人机交互方式。通道（modality）涵盖了用户表达意图、执行动作或感知反馈信息的各种通信方法，如言语、眼神、脸部表情、唇动、手动、手势、头动、肢体姿势、触觉、嗅觉或味觉等”。采用这种方式的计算机用户界面称为“多通道用户界面”。目前，人类最长使用的多通道交互技术包括手写识别、笔式交互、语音识别、语音合成、数字墨水、视线跟踪技术、触觉通道的力反馈装置、生物特征识别技术和人脸表情识别技术等方面。

- 情感人机交互

让计算机具有情感能力首先是由美国 MIT 大学 Minsky 教授（人工智能创始人之一）提出的。他在 1985 年的专著“The Society of Mind”中指出，问题不在于智能机器能否有任何情感，而在于机器实现智能时怎么能够没有情感[62]。从此，赋予计算机情感能力并让计算机能够理解和表达情感的研究、探讨引起了计算机界许多人士的兴趣。这方面的工作首推美国 MIT 媒体实验室 Picard 教授领导研究小组的工作。情感计算一词也首先由 Picard 教授于 1997 年出版的专著“Affective Computing（情感计算）”中提出并给出了定义，即情感计算是关于情感、情感产生以及影响情感方面的计算。

MIT 对情感计算进行全方位研究，正在开发研究情感机器人，最终有可能人机融合。其媒体实验室与 HP 公司合作进行情感计算的研究。IBM 公司的“蓝眼计划”，可使计算机知道人想干什么，如当人的眼瞄向电视时，它竟然知道人想打开电视机，它便发出指令打开电视机。此外该公司还研究了情感鼠标，可根据手部的血压及温度等传感器感知用户的情感。CMU 主要研究可穿戴计算机。日本在对感性信息处理的研究中，有众多研究单位参与，主要集中在研究所和高校。特别值得一提的是，日本欧姆龙公司研制生产的机器玩具曾风行一时，最高价达 4000 美元。随后其它公司也进行机器狗等玩具的生产。情感计算的研究不仅具有重要的科学和学术价值，也存在着巨大的商机，有很好的经济效益。

- 虚拟现实

虚拟现实（Virtual Reality, VR）是以计算机技术为核心，结合相关科学技术，生成与一定范围真实环境在视、听、触感等方面高度近似的数字化环境，用户借

助必要的装备与数字化环境中的对象进行交互作用、相互影响，可以产生亲临对应真实环境的感受和体验。虚拟现实是人类在探索自然、认识自然过程中创造产生，逐步形成的一种用于认识自然、模拟自然，进而更好地适应和利用自然的科学方法和科学技术。

虚拟现实技术具有很强的应用性。军事方面，将 VR 技术应用于军事演练，带来军事演练观念和方式的变革，推动了军事演练的发展。如美国的 SIMNET、ACTDSTOW、WARSIM2000 和虚拟之旗 2006 等一系列分布式虚拟战场环境。医学方面，VR 技术已初步应用于虚拟手术训练、远程会诊、手术规划及导航、远程协作手术等方面，某些应用已成为医疗过程不可替代的重要手段和环节。工业领域方面，VR 技术多用于产品论证、设计、装配、人机工效和性能评价等。代表性应用，如模拟训练、虚拟样机技术等已受到许多工业部门的重视。教育文化领域方面，VR 已经成为数字博物馆/科学馆、大型活动开闭幕式彩排仿真、沉浸式互动游戏等应用系统的核心支撑技术。纽约大都会博物馆、大英博物馆、俄罗斯冬宫博物馆和法国卢浮宫等都建立了自己的数字博物馆。



图 9-3 VR 参观卢浮宫概念图

- 智能用户界面

智能用户界面（Intelligent User Interface，IUI）是致力于改善人机交互的高效率、有效性和自然性的人机界面。它通过表达、推理，按照用户模型、领域模型、任务模型、谈话模型和媒体模型来实现人机交互。智能用户界面主要使用人工智能技术去实现人机通信，提高了人机交互的可用性：如知识表示技术支持基于模型的用户界面生成，规划识别和生成支持用户界面的对话管理，而语言、手

势和图像理解支持多通道输入的分析，用户建模则实现了对自适应交互的支持等。当然，智能用户界面也离不开认知心理学、人机工程学的支持。

智能体、代理（Agents）在智能技术中的重要性已“不言而喻”了。Agent 是一个能够感知外界环境并具有自主行为能力的以实现其设计目标的自治系统。智能的 Agent 系统可以根据用户的喜好和需要配置具有个性化特点的应用程序。基于此技术，我们可以实现自适应用户系统、用户建模和自适应脑界面。自适应系统方面，如帮助用户获得信息，推荐产品，界面自适应，支持协同，接管例行工作，为用户裁剪信息，提供帮助，支持学习和管理引导对话等。用户建模方面，目前机器学习是主要的用户建模方法，如神经网络、Bayesian 学习以及在推荐系统中常使用协同过滤算法实现对个体用户的推荐。自适应脑界面方面，如神经分类器通过分析用户的脑电波识别出用户想要执行什么任务（该任务既可以是运动相关的任务如移动手臂，也可以是认知活动如做算术题）。

- 自然语言人界交互

在“计算机文化”到来的社会里，语言已不仅是人与人之间的交际工具，而且是人机对话的基础，是最自然的一种人机交互方式。自然语言处理（Natural Language Processing, NLP）是使用自然语言同计算机进行通讯的技术，因为处理自然语言的关键是要让计算机“理解”自然语言，所以自然语言处理又叫做自然语言理解（Natural Language Understanding, NLU）。

近年来自然语言理解技术在搜索技术方面得到了广泛的应用，现在，已经有越来越多的搜索引擎宣布支持自然语言搜索特性，自然语言人机交互界面在智能短信服务、情报检索、人机对话等方面也具有广阔的发展前景和极高的应用价值，并有一些阶段性成果出现在商业运用中。

9.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为人机交互领域全球学者分布情况：



图 9-4 人机交互领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布在其东西海岸；欧洲也有较多的人才分布；亚洲的人才主要集中在日韩地区；其他诸如非洲、南美洲等地区的学者非常稀少；人机交互领域的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，人机交互领域中男性学者占比 84.6%，女性学者占比 15.4%，男性学者占比远高于女性学者。

人机交互领域学者的 h-index 分布如下图所示，大部分学者的 h-index 分布在中低区域，其中 h-index 在 20-30 区间的人数最多，有 842 人，占比 42.1%，50-60 区间的人数最少，有 136 人。

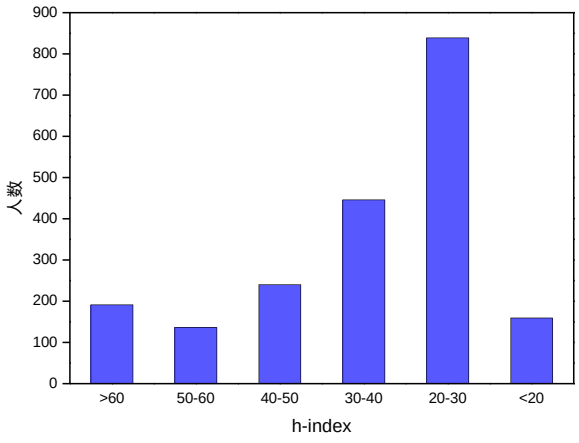


图 9-5 人机交互领域学者 h-index 分布

- 中国人才分布

我国专家学者在人机交互领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩等地相比，中国在人机交互领域学者数量较少。

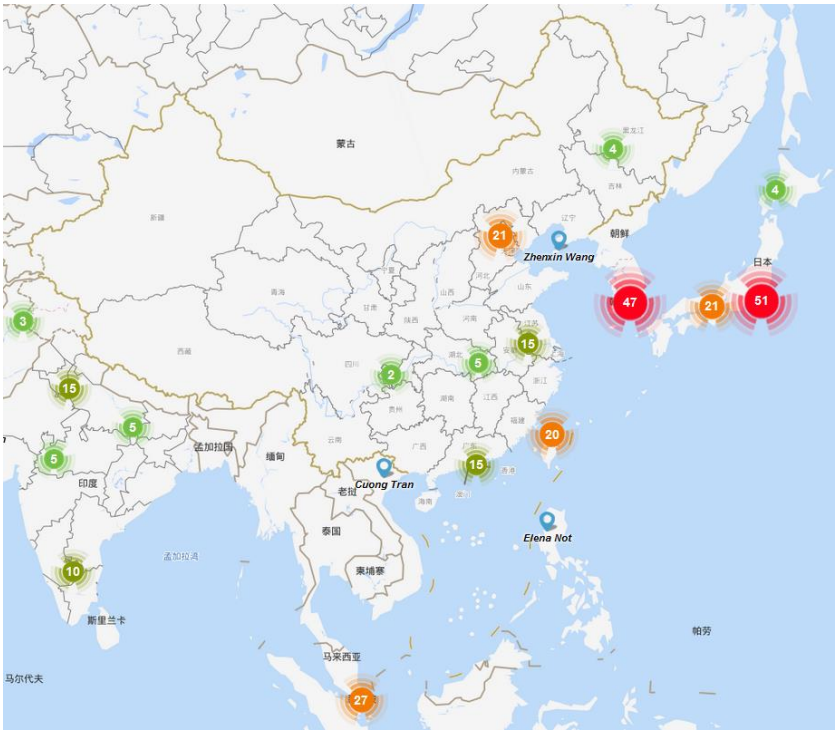


图 9-6 人机交互领域中国学者分布

中国与其他国家在人机交互领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 9-1 人机交互领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	142	3505	25	375
中国-日本	22	371	17	45
中国-英国	20	207	10	59
中国-新加坡	16	279	17	45
中国-德国	11	224	20	39

中国-加拿大	10	35	4	21
中国-芬兰	9	26	3	22
中国-澳大利亚	8	193	24	20
中国-法国	7	174	25	20
中国-荷兰	6	140	23	13

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在人机交互领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 5 席；中国与美国，中国与法国合作的论文平均引用数都达到了最高，说明在合作质量上中美、中法合作达到了较高的水平。

9.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

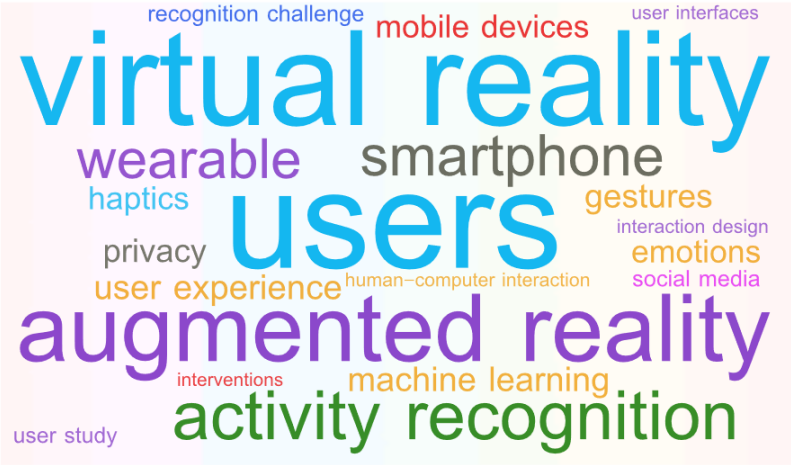
ACM CHI Conference on Human Factors in Computing Systems

ACM Symposium on User Interface Software and Technology

ACM International Conference on Ubiquitous Computing

International Journal of Human Computer Studies

ACM Transactions on Computer-Human Interaction



我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如上图所示。其中，用户（users）、虚拟现实（virtual

reality）、增强现实（augmented reality）是本领域中最热的关键词。

论文题目：*Guidelines for human-AI interaction*

中文题目：人工智能交互指南

论文作者：Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N. Bennett, Kori Inkpen, Jaime Teevan, Ruth Kikin-Gil, and Eric Horvitz

论文出处：ACM CHI Conference on Human Factors in Computing Systems 2019 (CHI 2019)

论文地址：https://doi.org/10.1145/3290605.3300233

研究问题：

人工智能（AI）领域的快速发展给用户界面和交互设计带来了新的机遇和挑战。虽然人机交互届对人和 AI 交互的原则原理已经进行了 20 多年的探讨，我们仍需要更多的研究和创新来解决人工智能新技术及其面向人类的应用不断涌现而带来的新科学及社会问题。作者提出了 18 条具有通用性的、可适用于多种应用场景的、针对人和 AI 交互的设计指导，指出现有知识的空缺及未来的探索方向。这份指南不仅为 AI 设计师提供了具体、可操作的建议，还旨在推动用户体验和工程开发从业者就设计决策的相关问题展开讨论，推动这一领域研究的

研究方法：

文章提出了 18 条人工智能交互设计指导，并进行多轮的实例评估来验证其有效性，包括通过一个用户实验，邀请 48 位设计师以这些设计指导为工具来测试 20 项广泛使用的有 AI 技术支持的用户产品。

指南内容包括：G1.帮助用户清晰的理解 AI 系统的能力所及；G2.帮用户树立对 AI 系统表现（如出错率）的正确期待。：G3.根据上下文设计好服务响应和中断的时机；G4.显示与上下文（用户当前任务及环境）相关的信息；G5.确保用户体验与其社会及文化规范相一致；G6.切实减少 AI 系统的语音及行为中可能存在的刻板印象和偏见。G7.保证需要时能容易且迅速的激活或请求 AI 系统的服务；G8.保证能够方便的取消或忽略不适当的 AI 系统服务；G9.保证 AI 系统出错

时用户可以对其进行编辑、修改、或者恢复设置；G10.AI 系统在不确定用户需求或目的时可进行适当问询或者以得体的方式降级服务；G11.向用户适当解释 AI 系统的行为及背后依据。G12.记住用户最近的交互模式，通过短期记忆提升使用效率；G13.学习用户长期的行为模式，提供个性化的体验；G14.降低 AI 系统升级或进行自适应调整时可能给用户带来的干扰；G15.引导用户在日常使用过程中向 AI 系统反馈其交互体验及偏好；G16.及时向用户传递其行为可能对 AI 系统将来的服务带来的影响；G17.让用户能够在全局层面定制 AI 系统对数据的监测及其行为模式；G18.将 AI 系统的变更适时通知给用户。

本文着重于在用户界面审查过程中可以很容易评估的设计指导意见。因此，没有包含诸如“建立信任”这样的抽象原则，而是侧重可观测的、可能可以帮助建立信任的具体措施。

过去的工作也提出了不少影响人和 AI 交互系统可用性的设计方针，但主要适用于 AI 模型建立阶段。未来可以进一步加强设计师和 AI 工程师的合作，在实际应用中进一步理解如何更好的评定不同指导意见的优先级并平衡各方面的用户体验。

研究结果：

本文通过三轮的评估，从 150 多条 AI 相关的设计建议中提取并验证了 18 条针对人与 AI 交互领域的通用设计指导意见。随着越来越多的科技应用以 AI 作为驱动力，本文希望能通过推广这些设计指导实现真正以人为本的 AI 交互系统。

本文研究整合了人机交互届对人工智能交互设计 20 多年的探索、思考、经验和教训，是一篇非常及时的综述性文章。不同于时下对于如何让 AI 模型更可用的针对 AI 从业人员研究，作者从设计师的角度出发，侧重讨论了在交互设计过程中需要注意的问题。另外，在提出设计指导意见时很好的权衡了适用范围以及可操作性，为其结果真正在实际应用中推广奠定了良好的基础，有着现实的指导意义。

论文题目：*Voice Interfaces in Everyday Life*

中文题目：日常生活中的语音交互界面

论文作者：ACM CHI Conference on Human Factors in Computing Systems 2018(CHI 2018)

论文出处：CHI 2018, April 21–26, 2018, Montreal, QC, Canada

论文地址：https://doi.org/10.1145/3173574.3174214

研究问题：

随着越来越多智能产品内嵌语音交互界面（后简称 VUI），工业界和学界产出了不少计算语言学或自然语言理解方面的研究成果，但却鲜有语音交互设备在生活场景中的实证使用调研。作者以此空白为契机，着重调研了语音交互和其他日常活动并行时的用户反馈，以及多方对话场景下的话论顺序处理。文章也涉及更为宽泛的三个领域：VUI 发展进程，VUI 调研中的会话分析，设计、部署和研究 VUI 时需要的方法论。

研究方法：

该研究沿用 HCI 文献中普遍的常人方法学和会话分析方法，记录了五户家庭一个月内使用 Echo 音箱的数据，收集了用户与机器六小时的对话语料进行解读。唤醒语音助手后，祈使指令和提问是常见的两种触发形式。对语音助手的指令会开启和机器间的话论转换；未收到音时的静默能推动话论前进；用户自发纠正说法也能推动话论延续。语音设备回复后，用户会有后续反馈，本文暂且讨论三种形式：静默也是用户的回复方式；用户会明确指出问题；用户也会纠正会话。

作者不认为语音交互的界面本质是对话的，用“对话界面”来命名基于语音的人机交互方式也不妥当。日常对话中的话论转换和毗邻对是自然产生的，而语音交互界面中的问法其实会刻意迎合会有的答案。因此，作者主张把任务型的 VUI 设计称为请求/回复设计，而非对话设计。

研究结果：

本文通过解读真实场景中和音箱的交互语料，展现了语音设备是如何并行参与到家庭对话中的；也发现除却设备自身的功能可用性，用户在社交场景中的行为也要为最终的 VUI 效果负责；同时本文还探讨了用户使用 VUI 时的触发和接应方式；最后提出三点概念性问题为后续 VUI 的设计和研究提供方向。

本文研究可谓填补了语音设备在真实场景中的调研空白。除此之外，很多易于被忽视的事实也被作者拎了出来。技术人员常常关注语言计算，业界设计师常常掉入逻辑或话术细节，鲜有人高屋建瓴地研究 VUI 使用情况。诸如静默也能推动话论前进、“对话设计”的命名纠正、关照用户发问的易用性等观点都是“跳出了盒子”的崭新思考。

论文题目：*TipText: Eyes-free Text Entry on a Fingertip Keyboard*

中文题目：指尖上的键盘：可盲打的指尖键盘输入法

论文作者：Zheer Xu, Pui Chung Wong, Jun Gong, Te-yen Wu, Aditya Nittala, Xiaojun Bi, Jürgen Steimle, Hongbo Fu, Kening Zhu, Xing-Dong Yang

论文出处：ACM Symposium on User Interface Software and Technology 2019 (UIST 2019)

论文地址：https://doi.org/10.1145/3332165.3347865

研究问题：

随着可穿戴设备以及普适计算的普及，人们越来越需要一种易于携带且跨平台通用的输入设备来进行交互。与此同时，基于拇指和食指指尖的微手势作为一种新型的跨平台交互方式有着得天独厚的优势：快速，简单易学（把食指指尖当作触摸板，用大拇指进行点击）而又隐蔽（有利于保护隐私）。这使得微手势特别适合输入空间十分有限的可穿戴设备。本文的主要目标就是为基于微手势的文字输入设计一种合理而高效的盲打键盘来论证可行性，并通过用户实验证明了这种输入方法可以达到理想的输入效率（在完成 40 个短句后速度可以达到每分钟最高约 13.9 个单词）。

研究方法：

本研究使用用户调研（user study）的数据采集方法和基于用户盲打数据的计算机仿真（computer-based simulation）来为微手势盲打键盘挑选出最好的键盘布局，并使用统计解码（statistical decoding）方法来实现文本输入，最后通过用户评估实验（user evaluation）来衡量整个输入法系统的性能。

本文将字母聚合的概念针对性地引入到指尖键盘上（即每个键对应多个字母，类似于传统的 9 键键盘），并和触控屏上文字输入的统计解码器创新性结合在一起，克服了高混淆度对文字输入效率的影响。

本文选择了基于计算机模拟的计算设计（computational design）方法：通过先采集少量用户数据，再使用计算机进行大规模模拟用户输入的方式来量化评价每个候选布局。这样的流程提高了对用户数据的利用效率，从而在一个极大的设计空间中科学而高效地寻找到了最符合人类行为模型的最优解；

在采集用户数据的时候，本文开创性地采用了“在虚拟世界中重构现实世界”的方法：通过将动作追踪系统和 3D 游戏引擎结合的方式来精确追踪拇指指尖和食指指尖的运动轨迹，从而计算出两者接触时的碰撞点。这种方法将传感器对用户行为的潜在影响降低到最小，最大程度保证了实验结论的严谨性。

研究结果：

本文通过开创性地将新兴的微手势交互技术运用到生活中常见的文字输入任务上，证明了在指尖键盘上进行文字盲打的可行性，不仅进一步打开了应用微手势交互技术的思路，也给进入了可穿戴设备时代的文字输入系统设计带来了更多的可能性。本文遵循计算设计的思想，根据用户的行为数据最大程度为指尖键盘优化了键盘布局，并在此基础上实现了一个概念论证的原型系统，在用户评估中表现出了令人满意的系统性能，从而进一步证明了盲打指尖键盘可以高效地完成文字输入任务。

本文首创了在极小的键盘上进行盲打文字输入任务，为“在可穿戴设备上缺乏足够的文字输入空间”这一问题提供了全新的解决思路。与此同时，本文遵循着计算设计的理念，在一个极其复杂的设计空间中抽丝剥茧，最终找到了一个充分考虑指尖盲打键盘的特点的最优键盘布局。在用户评估的过程中，整个系统展现出了很高的鲁棒性和有效性，以及良好的性能。值得一提的是，本文在人机交互顶级会议 UIST 2019 上获得了最佳论文的荣誉，究其原因在于为一个迫切的现实问题提供了崭新而又极具启发性的想法，并用科学严谨的实验流程向读者展示了如何将一个看似不可行的想法一步步实现的全过程，其中许多实验及设计方法

上的贡献已经超越了文字输入这一特定问题，给整个人机交互领域带来更多的启发和思考。

论文题目：*ElectroDermis: Fully Untethered, Stretchable, and Highly-Customizable Electronic Bandages*

中文题目：电子皮肤：完全不受束缚，可拉伸且高度可定制的电子创可贴

论文作者：Eric Markvicka, Guanyun Wang, Yi-Chin Lee, Gierad Laput, Carmel Majidi, and Lining Yao

论文出处：ACM CHI Conference on Human Factors in Computing Systems 2019(CHI 2019)

论文地址：https://dl.acm.org/citation.cfm?id=3300862

研究问题：

近些年柔性电子材料飞速发展，并在电子产品、医疗、健康检测、柔性交互界面等领域进行着颠覆性的创新。然而，相比于传统的电子设备或者电子器件，柔性电子显示出来的功能和应用还并不强大。很大程度上这是由于缺少对柔性电子工艺的集成和制作流程的设计，使得其从技术到应用还存在不小的差距。本文以此为动机，研究可用于皮肤表面的柔性电子交互界面的完整的设计到制造的流程，并更好地支持柔性电子可拉伸及可定制的特性。

研究方法：

本文采用 HCI 领域常用的原型设计、制作、展示及评估的方法。文章首先对所研究的问题进行了阐述，提出了设计上需考虑的因素，并针对每一个研究因素提出了解决方案。然后文章通过原型及示例应用的展示提出了解决方案。最后文章进行了必要的评测及对未来工作的讨论。

设计工具：文章首先开发了一套设计软件，支持用户通过简单的操作在 3D 扫描的人体模型上标出想要应用的区域，如膝盖、手肘等部位。软件自动完成所选几何平面的铺平，以准备后续的实物切割和制作。软件预存了电路设计图并自动在铺平的表面完成布置。同时，软件还嵌入了图形样式生成器，在不影响功能的同时提高了设备的美观。

多层制造：本文采用了多层制造的工艺步骤。本文首先设计并制备了小型化的电子元器件，包括电池、控制器、若干传感器等，同时采用光刻机制成波浪形的连接铜导线，具有很好的伸缩性和可弯曲。本文考虑市面上常见的、容易买到的、可直接进行激光切割或模切成型、且具有很好的粘附性能的材料：选用氨纶作为基质，可提高设备的弹性和实用性；选用医用薄膜来粘接设备和皮肤，可支持快速的贴合和去掉。在进行多层拼接时，首先将铜包板层压到 PDMS 基底上并进行切割，然后通过热塑性的聚氨酯热敏膜与之进行键合，形成的薄膜背面贴上之前准备的电子元器件，并将整体通过热处理与氨纶基质进行粘合。

功能展示：本文作者通过该工艺展示了多种柔性交互界面的应用场景，包括可以检测体温，心跳的随身贴，检测饮食活动的智能项链，检测伤口愈合情况的创可贴，检测环境中出现的颜色的指示灯，捕捉肢体运动的传感器等。每一个设备从设计到制作完成，都不超过 1 个小时。

性能评测：本文通过拉伸对设备进行了应变测试。通过观察发现该工艺制作的连接线具有很好的拉伸能力，在 171%应变的情况下其导电能力才受到明显的影响。本文也进行了旋扭拉伸测试，并发现在旋转两圈和 70%应变的情况下，设备的性能会收到明显的影响。总体而言，本文展示的设计方案所呈现的柔性界面表现出了优秀的柔韧度。

研究结果：

本文展示了一套完整的柔性交互界面的设计和制作过程，通过多层结构的设计和材料的选取，使得设备不仅能够适用于不同的交互场景，同时具有优秀的拉伸特性。这一制造工艺的设计，有利于新手设计师更便捷的进入到柔性电子设备的开发中。

本文的研究背景是柔性电子产业的飞速发展，带来了电子产品新形态的产生，也推动着可穿戴设备的发展。在关注柔性电子技术的同时，本文作者发现从技术的发展到应用的落地还存在不小的差距，而弥补这一差距的途径之一就是通过设计和制造，使得柔性电子的优秀技术能够快速的得到应用。本文的优势是借荐了如柔性导线结构、柔性多层制造工艺等先进材料领域的成果，并带入了如使用氨纶作为基质，医用薄膜作为粘附层等创新的设计思路，完整地考虑了从设计到制

造到交互应用的各个步骤。本文展示了近几年人机交互与材料学科的交叉融合、设计及创新，已成为人机交互技术发展的一个重要领域。

论文题目：*ReconViguRation: Reconfiguring Physical Keyboards in Virtual Reality*

中文题目：ReconViguRation: 在虚拟现实中重新配置物理键盘

论文作者：Daniel Schneider, Alexander Otte, Travis Gesslein, Philipp Gagel, Bastian Kuth, Mohamad Shahm Damlakhi, Oliver Dietz, Eyal Ofek, Michel Pahud, Per Ola Kristensson, Jo ¨ rg Mu ¨ ller, Jens Grubert

论文出处：IEEE International Symposium on Mixed and Augmented Reality 2019 (ISMAR 2019)

论文地址：https://ieeexplore_ieee.xilesou.top/abstract/document/8794572

研究问题：

迄今为止，键盘的物理布局通常被移植到虚拟现实（VR）中，以便在虚拟的标准办公环境中复制打字体验。本文探讨了如何充分利用 VR 的沉浸性，改变 VR 环境下物理键盘交互的输入输出特性。作者探索了一组输入和输出映射，用于重新配置物理键盘的虚拟模型，并通过具体设计、实施和评估 9 个与 VR 相关的应用程序来探索最终的设计空间：表情符号、语言和特殊字符、应用程序快捷方式、虚拟文本处理宏、窗口管理器、照片浏览器、打地鼠游戏、安全密码输入和虚拟触控条。作者在 20 名用户参与的研究中评估了这些应用程序的可行性，发现它们在 VR 中是可用的。在实证研究和分析的基础上，讨论了 VR 中物理键盘输入输出特性重新映射的局限性和可能性，并指出了该领域未来的研究方向。

研究方法：

本研究在 VR 中重新配置物理键盘的按键，并招募了 20 名用户来评估这种做法的可行性。首先，让用户使用 9 个 VR 应用程序来评估在 VR 环境中重新配置物理键盘的基本用户体验。其次，通过安全密码输入应用深入探讨了客观安全感和感知安全感之间的关系以及感知安全感和文本输入性能之间的权衡。最后，通过虚拟触控条应用分析了改变物理键盘的视觉表现形式对用户体验和性能的影响。应用程序细节如下：

语言和特殊字符：对于多语言输入或特殊字符输入的场景，可以把传统键盘映射成为相应语言或特殊字符的键盘（类似于智能手机中的多语种键盘）。

应用程序快捷方式：将传统键盘中的部分按键映射成浏览器的后退、前进、刷新、主页等快捷按键。

虚拟文本处理宏：将按键映射成为 Microsoft Word 中的宏命令（插入签名/发件人地址/图片）。

窗口管理器：将键盘重新配置为窗口管理器，按下按键可以切换到对应的窗口。

照片浏览器：在键盘的一个按键或几个按键上方显示相应的照片缩略图，按下相应按键可以浏览对应的照片。

打地鼠游戏：将物理键盘重新配置为打地鼠游戏，按下地鼠相应位置的按键可以打地鼠。

安全密码输入：将物理键盘原始的按键顺序打乱，实现虚拟环境中安全输入。

虚拟触控条：将键盘上的一行 10 个按键虚拟为控制视频播放进度的触控条（类似于 MacBook Pro 上的 Touch Bar），通过按下相应按键控制播放进度。

研究结果：

本文通过重新配置单个按键或整个键盘的输入输出，设计了 9 个与 VR 相关的应用程序：表情符号、语言和特殊字符、应用程序快捷方式、虚拟文本处理宏、窗口管理器、照片浏览器、打地鼠游戏、安全密码入口和虚拟触控条，并通过招募 20 名参与者进行用户研究，评估了在 VR 中重新配置物理键盘的可行性，发现这些应用程序在 VR 中是可用的。研究结果表明物理键盘可以在 VR 中以多种灵活的方式集成，作为 VR 中多种不同任务的输入设备，并且可以基于当前任务实时进行重新配置，在未来的 VR 应用中有着光明的前景。

本文利用 VR 的沉浸性特点，通过在 VR 环境中重新配置键盘的视觉显示和功能，评估了重新配置键盘的可行性，提出了一种在 VR 环境中集成物理键盘的新思路。本文设计了 9 个 VR 应用，覆盖了 VR 中办公和游戏的大部分应用场景，

展现了传统键盘输入在 VR 中的应用潜力。值得一提的是，本文在混合现实和增强现实顶级会议 ISMAR 2019 上获得了最佳论文的荣誉，究其原因在于将最传统的输入设备以一种崭新的应用形式带入了 VR 中，并通过合理的用户研究分析了其局限性和可行性，给 VR 中的人机交互带来了更多的启发和思考。

论文题目：*VIPBoard: Improving Screen-Reader Keyboard for Visually Impaired People with Character-Level Auto Correction*

中文题目：VIPBoard: 通过字符级别的自动纠错为视障用户优化读屏键盘

论文作者：Weinan Shi, Chun Yu, Shuyi Fan, Feng Wang, Tong Wang, Xin Yi, Xiaojun Bi, Yuanchun Shi

论文出处：ACM CHI Conference on Human Factors in Computing Systems 2019(CHI 2019)

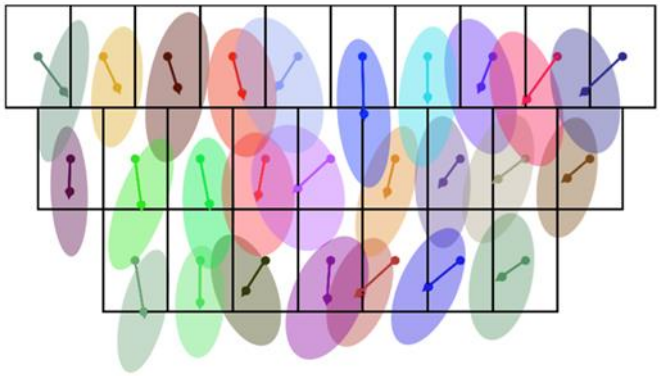
论文地址：https://dl.acm.org/citation.cfm?doid=3290605.3300747

研究问题：

在使用软键盘进行文本输入时，视障用户会根据读屏软件读出的字符来确认输入。因此视障用户只能逐字符地、确保正确地输入目标文本，输入效率较低且无法使用当代先进的自动纠错功能。本文提出了 VIPBoard，一种在不改变原有使用方式的前提下，专为视障用户设计的智能键盘。该键盘可以使用字符级别的自动纠错算法预测出用户最有可能输入的字符，并提供该预测字符的语音反馈，从而减少用户输入错误时所需的调整时间，提升输入效率。用户实验证明 VIPBoard 可以显著减少用户的输入错误（63.0%），提升文本输入速度（12.6%）。

研究方法：

本研究通过用户调研（user study）的数据采集方法得到了视障用户在文本输入时点击的位置，拟合得出反映用户行为的触摸模型。算法上使用基于贝叶斯原理的概率计算的方法预测用户的输入，通过多轮迭代的方法设计出了最小化用户学习成本的交互设计。最后通过用户评估实验（user evaluation）来衡量整个输入法的性能。视障用户进行文本输入时的触摸模型如下图：



本文首次将自动纠错应用到视障人士使用的读屏软键盘算法中。由于通常地的词级别纠错算法并不符合视障用户逐字正确输入的特征，本文创新性地使用了字符级的自动纠错算法，提高了用户输入的准确性，从而减少了输入过程中的调整次数，提高输入速度。

本文结合视障用户的使用行为特征，设计了一套布局自适应策略。该策略首先保证了键盘使用时的鲁棒性，即用户可以在键盘上输入任意想输入的字符，无论预测结果是否正确。同时，策略允许用户在不改变原有读屏键盘使用习惯的前提下使用，从而大大减小了用户的学习成本。

本文的用户实验评估进行地较为完备。文中实现了中文和英文两种语言以及 VIPBoard 和传统键盘两种算法下的系统原型，并让用户分别在两种语言和两种键盘下进行文本输入。实验结果不仅表明了 VIPBoard 相对于传统键盘的巨大优势，也说明了文中算法的普适性和实际应用价值。

研究结果：

本文设计了一款面向视障用户的智能键盘，通过使用字符级的自动纠错算法减少了用户输入错误时的调整时间，提高了文本输入的效率。同时，本文精心设计了与纠错算法相对应的布局调整策略与交互方式，使得用户可以在几乎没有学习成本的前提下使用该键盘。该键盘将现代智能技术带到了视障人士的生活中，为他们提供便利。

本文提出了世界上首款面向视障用户的智能键盘，开创性地将字符级的自动纠错思想引入到了视障用户的日常使用中，提高了输入效率。在用户评估中，整个技术体现出了相比于传统显著的性能优势和用户积极的主观偏好。值得一提的

是，本文在人机交互顶级会议 CHI2019 上获得了最佳论文提名的荣誉，究其原因在于从特殊群体的特殊使用方式出发将已有的解决问题的思路迁移到该特定的场景下，提出了一种优雅的解决方案。同时，针对用户的学习成本进行的交互设计优化和详实的实验设计，让该工作实际应用价值体现地更为明显，从技术角度给视障群体带来一丝光明。

论文题目：*EarTouch: Facilitating Smartphone Use for Visually Impaired People in Mobile and Public Scenarios*

中文题目：耳势交互：提升视力障碍用户在移动和公众场景下使用手机的体验

论文作者：Ruolin Wang, Chun Yu, Xing-Dong Yang, Weijie He, Yuanchun Shi

论文出处：ACM CHI Conference on Human Factors in Computing Systems 2019(CHI 2019)

论文地址：https://dl.acm.org/citation.cfm?id=3300254

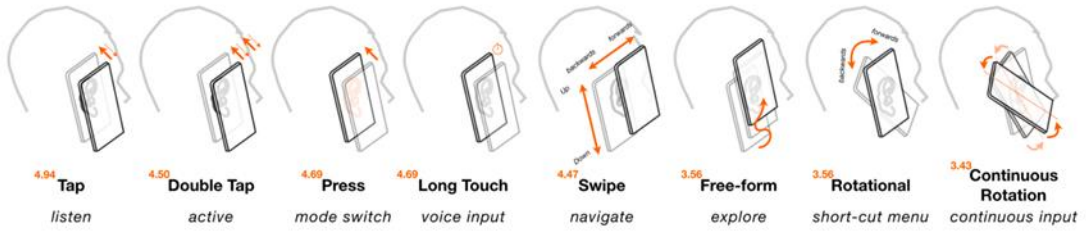
研究问题：

智能触屏手机已经成为视力障碍人群同世界沟通的重要工具。视力障碍用户使用手机时主要采用的姿势需要占用双手：一只手握持手机，用另一只手的手指在屏幕上触摸，读屏软件能够把交互内容转化为语音读出来。在公众场景下，通过手机扬声器听取语音反馈会受到噪音的影响，引起泄漏隐私等担忧，用户往往需要将手机扬声器举近耳边进行交互。而在移动出行的场景下，用户一只手被盲杖等物品占用时，完成交互需要腾出双手，转换姿态，费时费力。此前针对视障用户单手使用移动设备的研究工作集中于盲文的阅读和输入等有限的应用场景，而缺少一套面向手机上日常交互的无障碍解决方案。本文设计和开发了 EarTouch，一种基于耳势的单手交互技术，能够支持视力障碍用户在移动和公众场景下以一种相对稳定的使用姿态便捷地完成常用交互任务并轻松有效地获取语音反馈。

研究方法：

本研究开展了有 30 位视障用户参与的用户访谈和 23 位视障用户参与的设计工作坊，提出了一套包含八类耳势动作的交互设计；基于电容传感器数据和惯性传感器数据识别耳势动作，并在 16 位视力障碍用户的操作数据集上实现了高

准确度的评估效果；最后通过有 22 位视力障碍用户参与的用户实验对可用性进行评估。结果表明，EarTouch 易于学习，在多数任务上执行效率更高，富有趣味性，能够保护用户的隐私并被用户所接受，用户期待将来能够在自己的手机上使用到这项技术。EarTouch 选用的八类耳势交互动作、功能，及视障用户对于各类动作易于操作的评分的平均值（5 点李克特量表，5=非常同意该动作易于操作）。



EarTouch 拓展了智能手机的输入能力，使得耳朵能够代替手作为触摸输入，因而用户只需要占用一只手来握持手机。特殊地，耳朵靠近屏幕进行操作时也能够同时用来从听筒更加隐秘地接收语音反馈。另外，使用 EarTouch 时的握持姿态（耳朵位于听筒附近、嘴位于话筒附近）也支持用户在同一姿态下完成语音输入。

研究结果：

基于分析电容传感器数据和惯性传感器数据的智能算法，EarTouch 支持手机识别耳朵在屏幕上的接触、相对移动等动作，从而支持一系列交互操作如单击、双击、旋转等，使得视力障碍用户在公众和移动场景下能够便捷地单手完成高频、紧急的交互任务如接打电话、发送语音消息、快捷导航等。EarTouch 支持语音反馈从手机上方的听筒播放，无论是触摸输入、语音输入或是获取反馈，用户都可以在一个相对稳定轻松的类似打电话的单手姿态下完成。在不便使用耳机时，从听筒听取反馈也能解决公众场合下泄漏隐私和引起他人关注等问题。怀着基于用户能力水平进行设计的理念，EarTouch 可以使更多用户更容易地 与智能手机进行交互。虽然本文为视力障碍智能手机用户设计了 EarTouch，但该技术也可能使运动障碍的用户（如单臂残疾人等）和具有情境障碍的非残疾用户（如父母用一只手抱着婴儿时）受益。从更长远的层面上来讲，EarTouch 也为智能手机面向不同交互能力的所有用户的包容性设计迈出了重要的一步。

9.5 人机交互进展

最近的十年，是人机交互向自然交互蓬勃发展的十年。毋庸置疑，计算机是世纪最伟大的发明，其作用从科学计算工具迅速发展为信息处理和信息交互工具，起引领作用的则是人机交互技术的变革，即以鼠标发明为标志的图形用户界面（Graphical User Interface, GUI）的产生，一改规范命令与计算机交互的命令行界面模式（Command Line Interface，CLI），GUI 提供了普通人与计算机便捷交互的工具和方法，让计算机从实验室走进办公室、走入家庭，十多年前，触屏技术成为产品技术，GUI 中的鼠标被人的天然指点（pointing）工具——手指所取代，计算机又变身出手机，成为更多人方便使用的随身掌上工具。更少依赖操控工具，发展学习和使用成本更小的自然交互技术，一直是人机交互研究的价值追求，最近十年，随着感知和计算技术的进步，自然交互技术创新层出不穷，并能迅速成为新型产品技术，《麻省理工科技评论》总结和评论人机交互领域的突破技术（breakthroughs），为人机交互技术、未来终端技术的发展建立了一个高端的技术论坛，影响深远。我把这些突破技术分为 3 大类：支持自然动作的感知技术，面向穿戴的新型终端和基于语音识别的对话交互。

人体动作蕴含丰富的语义，动作交互技术一方面需要感知技术的进步，另一方面需要发现或设计有明确交互语义的动作（gesture，姿态，由于人手的灵巧性，手势成为主要的交互动作，通常叫做手势），如今，二维表面上，多指触摸动作在触屏上已普遍可用，三维空间中，嵌入了深度摄像头的手持和固定设备，能比较准确识别人的姿态和动作，做出响应。不同于人脸识别等目标明确的视觉识别任务，动作交互不仅要求视觉识别的准确度，更需要研究基于交互任务的动作表达的自然性与一致性，难以发现和突破，所以，除了动作语义很直白的动作游戏（body game），三维动作交互尚缺少普遍认知和接受的交互动作语义。而无论二维还是三维，手势的不可见性，是动作交互的主要难题。

穿戴（wearable）取代手持（handheld）曾是前几年的一个革命口号，目前看，市场上的确出现了一定规模的新产品，但穿戴仍是补充的地位。穿戴设备中，手环设备基本只有健康和活动检测功能，智能手表可以算做创新终端，但作为缩小版的手机，由于交互界面的缩小和操作方式的限制（通常是小界面上双手参与操

作），其承载功能也较手机缩减很多。VR／AR（虚拟现实／增强现实）的一个理想载体是头戴式设备，最近几年，多款智能眼镜产品面世，较之前笨重的头盔轻便了许多，逼真的虚拟场景和准确的现实对象识别信息都可以清晰呈现在眼前，并在特定领域开拓着增强体验的应用；然而，智能眼镜尚缺少与其三维真实显示匹配的准确的自然输入技术，以及从眼手绑定在手机上转变到眼手分离的眼镜设备上时，尚未建立起相应的交互模式。

自然语言对话式交互得益于大数据和智能技术的进步，多语言的自然语音识别技术在用户终端上都达到了很高的可用水平，并且，语音识别超越文本输入方式，成为智能软件助理的使能技术，近两年，更是有基于语音接口的家居产品如雨后春笋般出现，VUI（Voice User Interface，语音用户界面）已经成为交互术语。然而，VUI 的局限也是显而易见的，相对并行模式的视觉通道，串行模式的语音通道的带宽显然窄的多，出声的使用方式在很多场合是不合适的，但作为一种可用的自然交互技术，有效提升了用户体验。

人机交互作为终端产品的引领技术的作用已经是产业界的普遍认识，欣喜看到很多种自然交互技术和新型交互终端面世，但 GUI 仍是交互的主导模式。计算无所不在，交互自然高效是发展趋势，人机交互的研究和开发空间很大，需要综合地探索自然交互技术的科学原理，建立明确的优化目标，结合智能技术，发展高可用的自然交互技术。

10 机器人

10.1 机器人概念

机器人广义上包括一切模拟人类行为或思想以及模拟其他生物的机械（如机器狗，机器猫等）。狭义上对机器人的定义还有很多分类法及争议，有些电脑程序甚至也被称为机器人（例如爬虫机器人）。联合国标准化组织采纳了美国机器人协会给机器人下的定义：“一种可编程和多功能的操作机；或是为了执行不同的任务而具有可用电脑改变和可编程动作的专门系统。一般由执行机构、驱动装置、检测装置和控制系统和复杂机械等组成”。机器人是综合了机械、电子、计算机、传感器、控制技术、人工智能、仿生学等多种学科的复杂智能机械。

目前，智能机器人已成为世界各国的研究热点之一，成为衡量一国工业化水平的重要标志。机器人是自动执行工作的机器装置，因此，它既可以接受人类指挥，又可以运行预先编排的程序，也可以根据以人工智能技术制定的原则纲领行动。在当代工业中，机器人指能自动执行任务的人造机器装置，用以取代或协助人类工作，一般会是机电装置，由计算机程序或电子电路控制。机器人的范围很广，可以是自主或是半自主的，从本田技研工业的 ASIMO 或是 TOSY 的 TOPIO 等拟人机器人到工业机器人，也包括多台一起动作的群机器人，甚至是纳米机器人。借由模仿逼真的外观及自动化的动作，理想中的高仿真机器人是高级整合控制论、机械电子、计算机与人工智能、材料学和仿生学的产物。机器人可以作一些重复性高或是危险，人类不愿意从事的工作，也可以做一些因为尺寸限制，人类无法作的工作，甚至是像外太空或是深海中，不适人类生存的环境。机器人在越来越多方面可以取代人类，或是在外貌、行为或认知，甚至情感上取代人类。

机器人技术最早应用于工业领域，但随着机器人技术的发展和各行业需求的提升，在计算机技术、网络技术、MEMS 技术等新技术发展的推动下，近年来，机器人技术正从传统的工业制造领域向医疗服务、教育娱乐、勘探勘测、生物工程、救灾救援等领域迅速扩展，适应不同领域需求的机器人系统被深入研究和开发。过去几十年，机器人技术的研究与应用，大大推动了人类的工业化和现代化进程，并逐步形成了机器人的产业链，使机器人的应用范围也日趋广泛[63]。

10.2 机器人发展历史

“机器人”一词最早出现在 1920 年捷克斯洛伐克剧作家 Karel Capek 的科幻情节剧《罗萨姆的万能机器人》中。

机器人从幻想世界真正走向现实世界是从自动化生产和科学研究的发展需要出发的。1939 年，纽约世博会上首次展出了由西屋电气公司制造的家用机器人 Elektro，但它只是掌握了简单的语言，能行走、抽烟，并不能代替人类做家务。

现代机器人的起源则始于二十世纪 40-50 年代，美国许多国家实验室进行了机器人方面的初步探索。二次世界大战期间，在放射性材料的生产和处理过程中应用了一种简单的遥控操纵器，使得机械抓手就能复现人手的动作位置和姿态，代替了操作人员的直接操作。在这之后，橡树岭和阿尔贡国家实验室开始研制遥控式机械手作为搬运放射性材料的工具。1948 年，主从式的遥控机械手正式诞生于此，开现代机器人制造之先河。美国麻省理工学院辐射实验室（MIT Radiation Laboratory）1953 年研制成功数控铣床，把复杂伺服系统的技术与最新发展的数字计算机技术结合起来，切削模型以数字形式通过穿孔纸带输入机器，然后控制铣床的伺服轴按照模型的轨迹作切削动作。

上世纪 50 年代以后，机器人进入了实用化阶段。1954 年，美国的 GeorgeC. Devol 设计并制作了世界上第一台机器人实验装置，发表了《适用于重复作业的通用性工业机器人》一文，并获得了专利。GeorgeC. Devol 巧妙地把遥控操作器的关节型连杆机构与数控机床的伺服轴连接在一起，预定的机械手动作一经编程输入后，机械手就可以离开人的辅助而独立运行。这种机器人也可以接受示教而能完成各种简单任务。示教过程中操作者用手带动机械手依次通过工作任务的各个位置，这些位置序列记录在数字存储器内，任务执行过程中，机器人的各个关节在伺服驱动下再现出那些位置序列。因此，这种机器人的主要技术功能就是“可编程”以及“示教再现”。

上世纪 60 年代，机器人产品正式问世，机器人技术开始形成。1960 年，美国的 Consolidated Control 公司根据 GeorgeC. Devol 的专利研制出第一台机器人样机，并成立 Unimation 公司，定型生产了 Unimate（意为“万能自动”）机器人。同时，美国“机床与铸造公司”（AMF）设计制造了另一种可编程的机器人

Versatran（意为“多才多艺”）。这两种型号的机器人以“示教再现”的方式在汽车生产线上成功地代替工人进行传送、焊接、喷漆等作业，它们在工作中表现出来的经济效益、可靠性、灵活性，使其它发达工业国家为之倾倒。于是 Unimate 和 Versatran 作为商品开始在世界市场上销售，日本、西欧也纷纷从美国引进机器人技术。这一时期，可实用机械的机器人被称为工业机器人。

在机器人崭露头角于工业生产的同时，机器人技术研究不断深入。1961 年，美国麻省理工学院 Lincoln 实验室把一个配有接触传感器的遥控操纵器的从动部分与一台计算机连结在一起，这样形成的机器人可以凭触觉决定物体的状态。随后，用电视摄像头作为输入的计算机图像处理、物体辩识的研究工作也陆续取得成果。1968 年，美国斯坦福人工智能实验室（SAIL）的 J. McCarthy 等人研究了新颖的课题——研制带有手、眼、耳的计算机系统。于是，智能机器人的研究形象逐渐丰满起来。

上世纪 70 年代以来，机器人产业蓬勃兴起，机器人技术发展为专门的学科。1970 年，第一次国际工业机器人会议在美国举行。工业机器人各种卓有成效的实用范例促成了机器人应用领域的进一步扩展；同时，又由于不同应用场合的特点，导致了各种坐标系统、各种结构的机器人相继出现。而随后的大规模集成电路技术的飞跃发展及微型计算机的普遍应用，则使机器人的控制性能大幅度地得到提高、成本不断降低。于是，导致了数百种类的不同结构、不同控制方法、不同用途的机器人终于在 80 年代以来真正进入了实用化的普及阶段。进入 80 年代后，随着计算机、传感器技术的发展，机器人技术已经具备了初步的感知、反馈能力，在工业生产中开始逐步应用。工业机器人首先在汽车制造业的流水线生产中开始大规模应用，随后，诸如日本、德国、美国这样的制造业发达国家开始在其他工业生产中也大量采用机器人作业。

上世纪 80 年代以后，机器人朝着越来越智能的方向发展，这种机器人带有多种传感器，能够将多种传感器得到的信息进行融合，能够有效的适应变化的环境，具有很强的自适应能力、学习能力和自治功能。智能机器人的发展主要经历了三个阶段，分别是可编程试教、再现型机器人，有感知能力和自适应能力的机器人，智能机器人。其中所涉及到的关键技术有多传感器信息融合、导航与定位、路径规划、机器人视觉智能控制和人机接口技术等。

进入 21 世纪，随着劳动力成本的不断提高、技术的不断进步，各国陆续进行制造业的转型与升级，出现了机器人替代人的热潮。同时，人工智能发展日新月异，服务机器人也开始走进普通家庭的生活。世界上许多机器人科技公司都在大力发展机器人技术，机器人的特质与有机生命越来越接近。

最近，波士顿动力公司在机器人领域的成就已经成为人们的焦点，其产品机器狗 Spot 和双足人形机器人 Atlas 都让人大为惊叹。Spot 的功能十分先进，可以前往你告诉它要去的目的地，避开障碍，并在极端情况下保持平衡。Spot 还可以背负多达四个硬件模块，为公司提供其他多款机器人完成特定工作所需的任何技能；Atlas 已经掌握了倒立、360 度翻转、旋转等多项技能，继表演跑酷、后空翻等绝技之后，Atlas 又掌握了一项新技能—体操，再次让我们大开眼界。



图 10-1 波士顿动力机器人 Spot 与 Atlas

经过几十年的发展，机器人技术终于形成了一门综合性学科——机器人学（Robotics）。一般地说，机器人学的研究目标是以智能计算机为基础的机器人的基本组织和操作，它包括基础研究和应用研究两方面内容，研究课题包括机械手设计、机器人动力和控制、轨迹设计与规划、传感器、机器人视觉、机器人控制语言、装置与系统结构和机械智能等。由于机器人学综合了力学、机械学、电子学、生物学、控制论、计算机、人工智能、系统工程等多种学科领域的知识，因此，也有人认为机器人学实际上是一个可分为若干学科的学科门类。同时，由于机器人是一门不断发展的科学，对机器人的定义也随着其发展而变化，目前国际上对于机器人的定义纷繁复杂，RIA、JIRA、NBS、ISO 等组织都有各自的定义，迄今为止，尚没有一个统一的机器人定义。

10.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为机器人领域全球学者分布情况：

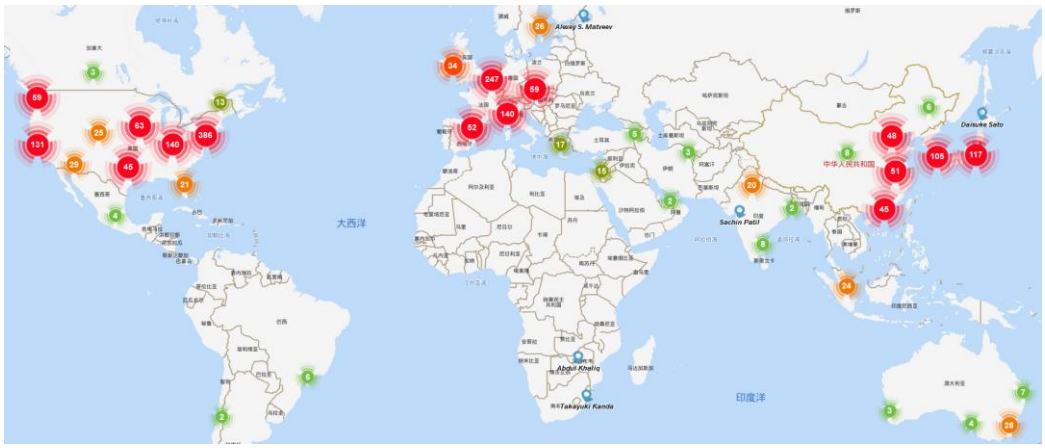


图 10-2 机器人领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显；欧洲也有较多的人才分布；亚洲的人才主要集中在我国东部及日韩地区；其他诸如非洲、南美洲等地区的学者非常稀少；机器人领域的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，机器人领域中男性学者占比 90.7%，女性学者占比 9.3%，男性学者占比远高于女性学者。

机器人领域学者的 h-index 分布如下图所示：

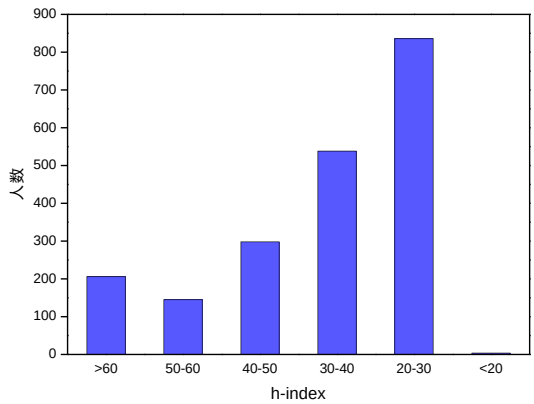


图 10-3 机器人领域学者 h-index 分布

从上图可以看出，大部分学者的 h-index 分布在中低区域，其中 h-index 在 20-30 区间的人数最多，有 836 人，占比 41.3%，小于 20 区间的人数最少，只有 3 人。

- 中国人才分布

我国专家学者在机器人领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是珠三角和长三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩等地相比，中国在机器人领域学者数量较少。



图 10-4 机器人领域中国学者分布

中国与其他国家在机器人领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 10-1 机器人领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	445	6606	15	756

中国-日本	90	998	11	196
中国-加拿大	69	761	11	128
中国-新加坡	62	586	9	123
中国-德国	57	780	14	98
中国-英国	50	1270	25	94
中国-法国	41	860	21	63
中国-澳大利亚	27	351	13	45
中国-瑞典	20	246	12	27
中国-意大利	18	318	18	33

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在机器人领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席；中国与英国合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中英合作达到了较高的水平。

10.4 论文解读

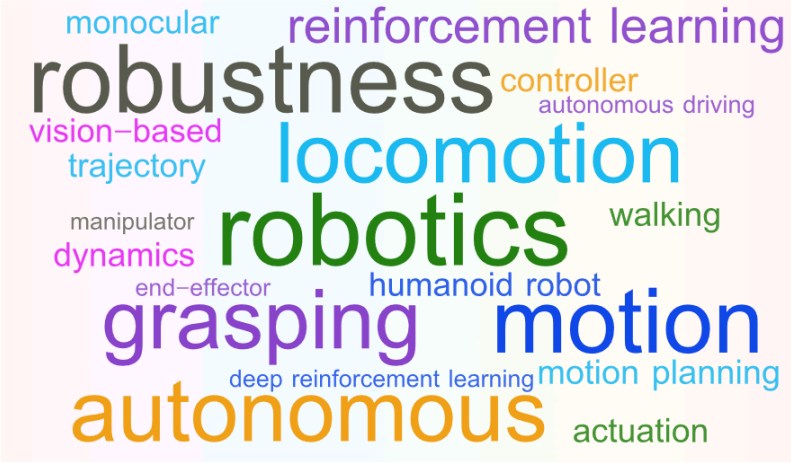
本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

IEEE International Conference on Robotics and Automation

IEEE/RSJ International Conference on Intelligent Robots and Systems

Robotics: Science and Systems·A Robotics Conference

IEEE Transactions on Robotics



我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如上图所示。其中，机器人（robotics）、鲁棒性（robustness）、动作（motion）是本领域中最热的关键词。

论文题目：*Robotic Pick-and-Place of Novel Objects in Clutter with Multi-Affordance Grasping and Cross-Domain Image Matching*

中文题目：通过多 affordance 抓取和跨域图像匹配完成杂乱环境下对新物体的捡放操作

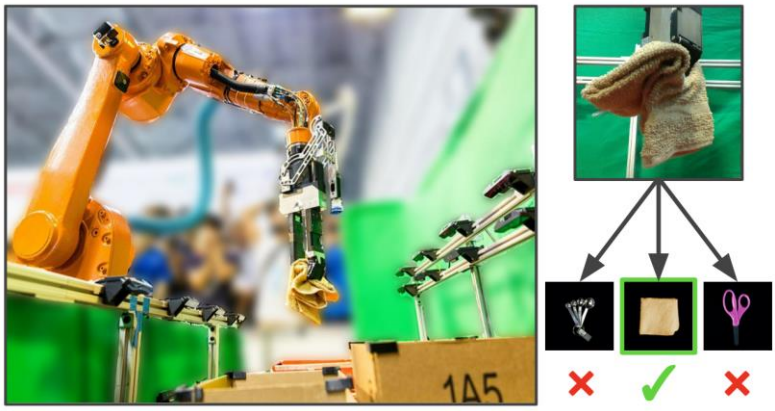
论文作者：Andy Zeng, Shuran Song, Kuan-Ting Yu, Elliott Donlon, Francois R. Hogan, Maria Bauza, Daolin Ma, Orion Taylor, Melody Liu, Eudald Romo, Nima Fazeli, Ferran Alet, Nikhil Chavan Dafle, Rachel Holladay, Isabella Morona, Prem Qu Nair, Druck Green, Ian Taylor, Weber Liu, Thomas Funkhouser, Alberto Rodriguez

论文出处：IEEE International Conference on Robotics and Automation, 2018

论文地址：https://ieeexplore.ieee.org/abstract/document/8461044

研究问题：

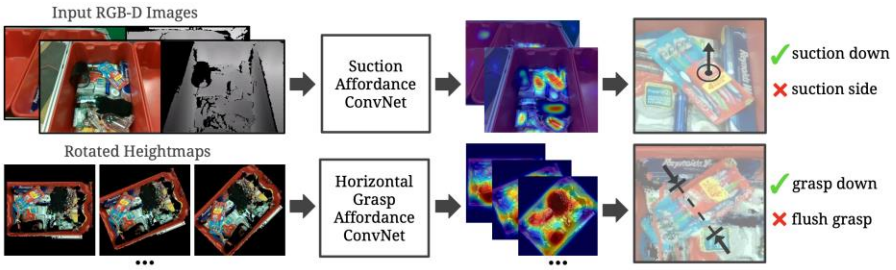
人类可以在仅掌握少量先验知识的前提下识别和抓取陌生目标物，这一能力一直是机器人研究的灵感来源，也是很多实际应用的核心。为此，提出一种能在杂乱环境下对新目标物进行识别和捡放操作的机器人系统，整个系统可直接用于新目标物（在测试过程中首次出现），而无需额外的数据收集或重新训练，如下图所示。



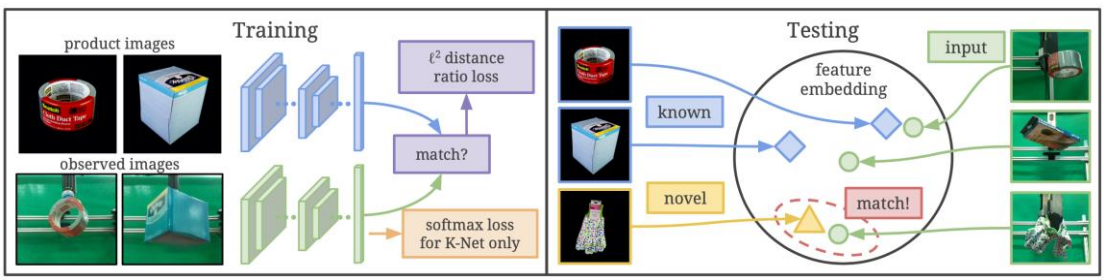
研究方法：

专门设计该机器人识别捡放系统，该系统由两部分组成：1）具有四个基础行为的多模式抓取框架，该框架使用深度卷积神经网络（ConvNets）来预测场景 affordance，而无需事先对目标物进行分割和分类。2）跨域图像匹配框架，用于通过将抓取的对象与产品图像进行匹配来识别抓取的对象，该框架使用了 ConvNet 架构，该架构可直接用于新目标物而无需重新进行训练。这两部分互相配合，可以在杂乱的环境中完成对新目标物的抓取操作。

机器人吸取和抓取的 affordance 预测如下图所示，考虑各个视角的 RGBD 图像，可通过一个全卷积残差网络估算出每张图片的吸取 affordance。然后，将预测汇总在 3D 点云上，并基于表面法线生成向下吸取或侧向吸取的建议。并行地，我们将 RGB-D 图像合并为 RGB-D 高度图，将其旋转 16 个不同的角度，并估计每个高度图的水平抓取。这有效地生成了针对 16 个不同抓取角度的 affordance 图，从中可得到向下抓取和其他抓取的建议。



新物体的识别框架如下图所示。训练一个双流的卷积神经网络，其中一个流计算得到产品图像的 2048 维特征向量，而另一个流计算得到观察图像的 2048 维特征向量，并对两个流进行优化，以使相同图像的特征更加相似，反之则不同。在测试期间，已知对象和新对象的图像都映射到公共特征空间上。通过在相同的特征空间找到与其最近的特征来匹配来识别它们。



研究结果：

提出一种系统，该系统能够以很少的先验信息（少数产品图片）来拾取和识别新对象。该系统首先使用与类别无关的 affordance 预测算法在四种不同的抓取动作元之间进行选择，然后通过将抓取的对象与它们的产品图像进行匹配来识别抓取的对象。通过评估证明，该机器人系统可以拾取并在杂乱无章的环境中识别出新物体。

论文题目：*Using Simulation and Domain Adaptation to Improve Efficiency of Deep Robotic Grasping*

中文题目：使用仿真和领域适应来提高深度机器人抓取的效率

论文作者：Konstantinos Bousmalis, Alex Irpan, Paul Wohlhart, Yunfei Bai, Matthew Kelcey, Mrinal Kalakrishnan, Laura Downs, Julian Ibarz, Peter Pastor, Kurt Konolige, Sergey Levine, Vincent Vanhoucke

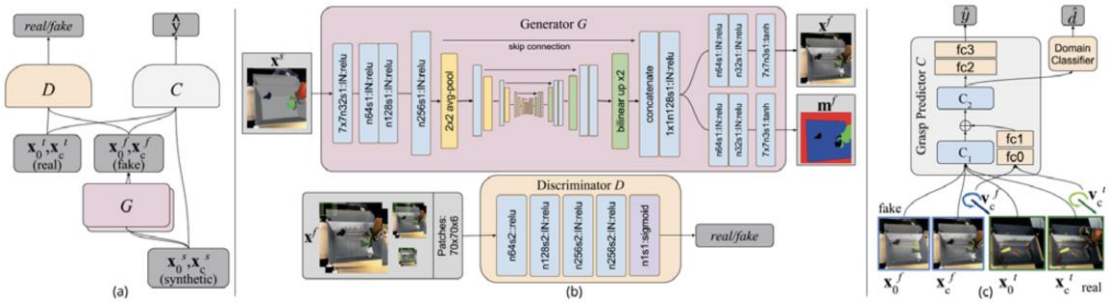
论文出处：IEEE International Conference on Robotics and Automation, 2018

论文地址：https://ieeexplore.ieee.org/document/8460875

研究问题：

收集带注释的视觉抓取数据集以训练现代机器学习算法可能是非常耗时的。一个替代方法是使用现成的模拟器来合成数据，这样就可以自动产生这些数据的真实标注。不幸的是，仅基于模拟数据训练的模型通常无法泛化到现实世界。研究如何将随机模拟环境和域适应方法应用到真实场景，训练一种抓取系统，可以通过原始 RGB 图像中进行规划，抓取新的目标物。

研究方法：



研究模拟环境中的 3D 目标模型、模拟的场景和模拟的动力学如何影响机器人最终的抓取性能，以及将模拟与真实场景集成以实现最大程度的迁移。具体方法如上图所示。

（a）为像素级的域适应模型 GraspGAN 的概述。从仿真器中得到的图像元组 x^s 并输入到生成器 G 中，生成真实版本的图像 x^f，鉴别器 D 获得未标注的真实世界图像 x^t 和 x^f，并经过训练以区分它们。真实的和经过适应的图像也被送到抓取预测网络 C 中，一同进行并行训练。因此，生成器 G 从鉴别器 D 和预测器 C 中获得反馈，以使适应的图像看起来更加真实并保持其语义信息；（b）为生成器 G 和鉴别器 D 的体系结构；（c）为 DANN 模型，其中 C_1 包含 7 个卷积层，C_2 包含 9 个卷积层。

研究结果：

研究将模拟数据合并到基于学习的抓取系统中的方法，以提高抓取性能并减少数据需求。通过使用合成数据和域适应，仅使用少量随机生成的模拟数据，就可以达到给定性指标的 50 倍。还表明，仅使用未标注的真实数据和 GraspGAN 的方法，就可以在没有任何真实数据标注的情况下获得与真实世界相同的抓取性能。

论文题目：*Dex-Net 2.0: Deep Learning to Plan Robust Grasps with Synthetic Point Clouds and Analytic Grasp Metrics*

中文题目：Dex-Net 2.0：利用合成点云进行鲁棒抓取和分析抓取指标的深度学习

论文作者：Bohg Jeffrey Mahler, Jacky Liang, Sherdil Niyaz, Michael Laskey, Richard Doan, Xinyu Liu, Juan Aparicio Ojea, and Ken Goldberg

论文出处：Robotics: Science and Systems, 2017

论文地址：https://arxiv.org/pdf/1703.09312.pdf

研究问题：

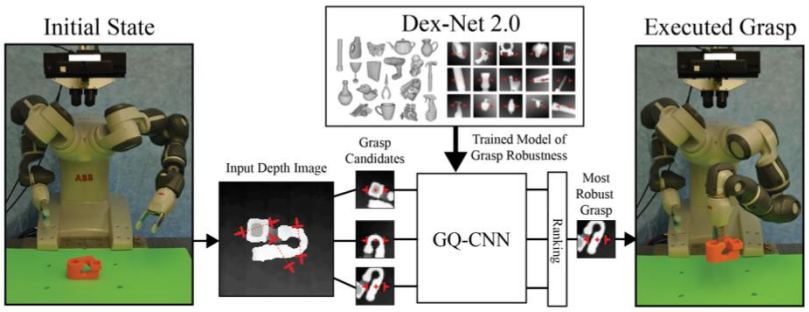
为了减少采用深度学习进行鲁棒机器人抓取策略学习所需的数据收集时间，探索了从 670 万点云，抓取姿态和抓取指标的合成数据集进行训练，这些数据是从 Dex Net 1.0 的数千个三维模型中以随机姿势在桌子上生成的。利用得到的数

据集 Dex-Net 2.0 训练抓取质量卷积神经网络（GQ-CNN）模型，该模型可快速从深度图像预测抓取成功的概率，其中抓取被指定为相对于 RGB-D 传感器的夹持器的平面位置、角度和深度。

研究方法：

研究基于深度点云的处于桌面上的单刚体的平行爪抓取规划问题。学习一个函数，它以一个候选抓取和一个深度图像作为输入，并输出一个鲁棒性的估计值，或者在传感和控制的不确定性下的成功概率。

Dex Net 2.0 的架构如下图所示。GQ-CNN 是抓取质量卷积神经网络，它是经离线训练的，使用由 670 万个合成点云、相关鲁棒抓取指标的数据集 Dex-Net 1.0 数据集，可从深度图像预测鲁棒候选抓取。当一个物体呈现给机器人时，深度相机会返回一个三维点云，识别出一组几百个候选抓取点。GQ-CNN 迅速确定了最稳健的抓取位姿，并由 ABB YuMi 机器人执行操作。



研究结果：

开发了一个抓取质量卷积神经网络（GQ-CNN）体系结构，它可以预测基于点云模型抓取的稳定性，并在 Dex-2.0 数据集上对其进行训练，它是一个包含 670 万点云、平行抓取和稳定性抓指标的数据集。在 1000 多个物理评估中，发现 Dex-Net 2.0 抓取规划器是一种可靠的、速度比基于点云配准方法快 3 倍的，并且在 40 个新目标的测试集上具有 99%的精度的抓取规划器。

论文题目：*Deep Predictive Policy Training using Reinforcement Learning*

中文题目：深度预测策略的强化学习训练方法

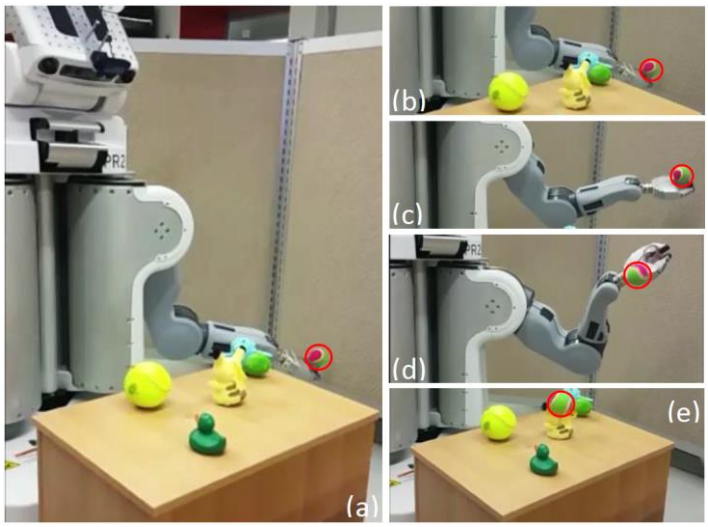
论文作者：Ali Ghadirzadeh, Atsuto Maki, Danica Kragic and Marten Bjorkman.

论文出处：Robotics: Science and Systems，2019

论文地址：https://arxiv.org/pdf/1903.11239.pdf

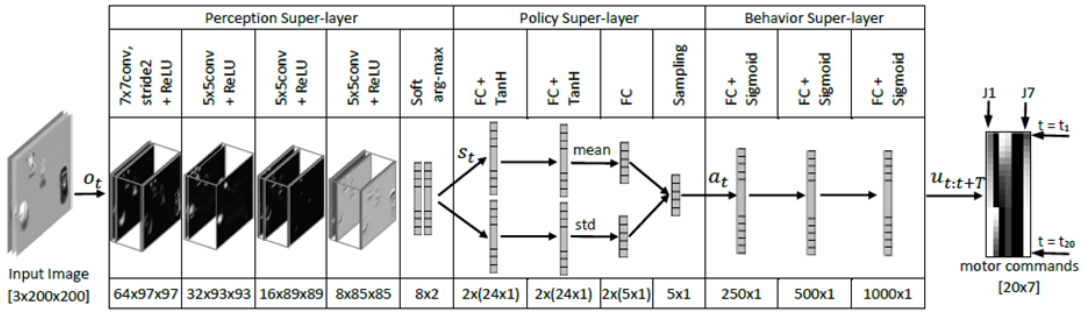
研究问题：

由于感知运动过程的固有延迟，机器人任务学习最好通过预测动作策略来实现。然而，训练这样的预测策略是具有挑战性的，因为它涉及到在整个动作过程中找到运动激活的轨迹。本文中，提出一个基于深度神经网络的数据高效深度预测策略训练（DPPT）框架，将图像观测映射到一系列的运动激活。该体系结构由三个子网络组成，分别称为感知层、策略层和行为层。感知层和行为层迫使我们对视觉和行为进行抽象分别用合成训练样本和模拟训练样本训练数据。策略层是一个较小的子网络，具有较少的参数来映射抽象流形之间的数据。使用策略搜索强化学习的方法对每个任务进行训练。通过在 PR2 机器人上训练熟练抓取和投球的预测策略。下图表示在机器人抛掷 ball 过程的瞬间图。



研究方法：

由感知层、策略层和行为层组成的深度预测策略体系结构如下图所示。作为网络输入，给出了一个中心 RGB 图像。感知层将图像数据抽象为与任务相关的对象对应的若干空间位置。策略层将抽象状态随机映射到操作流形中的一个点。最后，针对给定的采样动作，行为层生成一长轨迹的电机指令，并应用于机器人连续 T 个时间步长。



研究结果：

文章证明了所提出的结构和学习框架的适用性。该方法的有效性通过以下事实得到了证明:这些任务仅使用 180 次真正的机器人进行训练，并提供定性的最终奖励。

论文题目：*Learning Agile and Dynamic Motor Skills for Legged Robots*

中文题目：面向腿式机器人的敏捷动态特性的技能学习

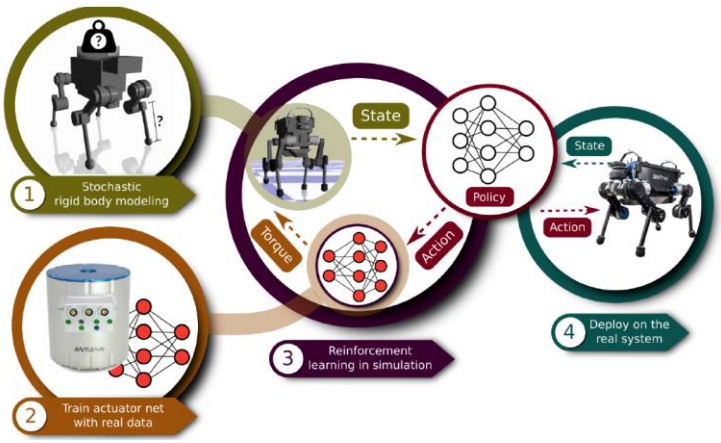
论文作者：Jemin Hwangbo, Joonho Lee, Alexey Dosovitskiy, Dario Bellicoso, Hoonho Lee, Vassilios Tsounis, Vladlen Koltun and Marco Hutter.

论文出处：Science Robotics, 2019

论文地址：https://arxiv.org/pdf/1901.08652.pdf

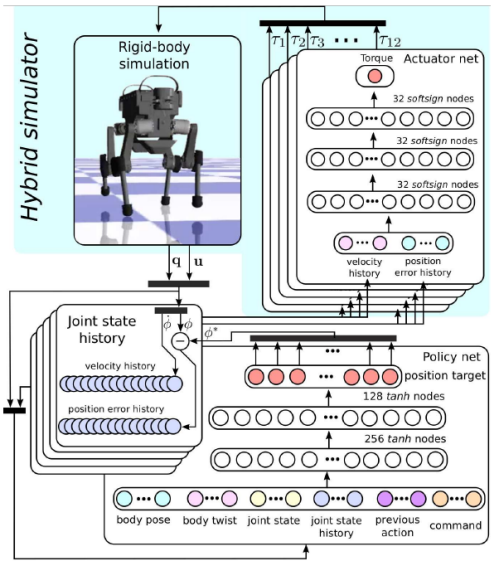
研究问题：

近年来，腿式机器人是机器人技术中最大的挑战之一。动物的动态和敏捷的动作是现有的方法无法模仿的，而这些方法是由人类精心设计的。一个令人信服的替代方案是强化学习，它需要最少的技能并促使控制策略的自然演化更新。然而，到目前为止，对腿式机器人的强化学习研究主要局限于仿真，在实际系统中应用比较简单的例子较少。主要原因是，使用真正的机器人进行训练，尤其是使用动态平衡系统，既复杂又昂贵。在本论文中，我们提供了一种新的方法，在模拟中训练一个神经网络策略，并将其迁移到一个最先进的腿系统，因此我们利用快速、自动化和经济有效的数据生成方案。



研究方法：

对于腿式机器人的敏捷动态性技能学习的过程，首先是系统建模，针对于四足机器人的物理参数的辨识以及确定机器人动态参数的不确定性指标，这个过程可能需要环境参数估计，物理动态性能估计等；其次是训练驱动神经网络，这个过程一般通过构建机器人状态到机器人电机控制的映射函数实现，随着深度神经网络的广泛认可，这样的非线性映射函数现大多采用深度神经网络拟合；然后在仿真中完成基于强化学习的驱动神经网络的学习过程，最后将训练好的驱动神经网络拟合的控制策略应用在实际的系统中。



整个系统的控制网络由三部分构成，首先是策略网络，用于将当前的观测量和之前的关节状态量映射到目标关节量（下一时刻关节控制量），然后是驱动网络，用于在刚体关节控制中将历史关节状态映射到关节力矩控制量上，机器人状态量有各关节的位置信息 q 与速度信息 u。

研究结果：

应用于一个复杂的中型犬大小的四足系统 ANYmal 机器人，使得在模拟中训练的四足机器人的运动策略超越了以前的方法，ANYmal 能够精确和高效地遵循高水平的身体速度指令，比以前跑得更快，甚至在复杂的配置中也能从跌倒中恢复过来。

论文题目：*Making Sense of Vision and Touch: Self-Supervised Learning of Multimodal Representations for Contact-Rich Tasks*

中文题目：理解视觉和触觉：接触任务多模态表达的自监督学习

论文作者：Michelle A. Lee, Yuke Zhu, Krishnan Srinivasan, Parth Shah, Silvio Savarese, Li Fei-Fei, Animesh Garg, and Jeannette Bohg

论文出处：IEEE International Conference on Robotics and Automation，2019

论文地址：https://ieeexplore.ieee.org/abstract/document/8793485

研究问题：

非结构化环境中需要接触的操作任务通常需要触觉和视觉反馈。但是，人工设计融合各个不同模态的机器人控制器并非易事。尽管深度强化学习已经成功地应用于针对高维输入的控制策略学习，但由于样本复杂性，这些算法通常难以部署在实际的机器人上。提出使用自监督来学习感官输入的紧凑和多模态表示，以用来提高策略学习的样本效率。

研究方法：

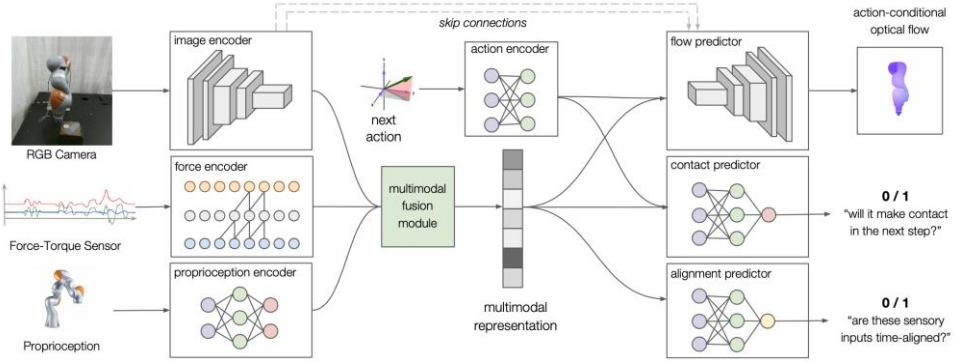
以学习机器人执行需要接触操作任务的策略为目标。希望评估融合多传感器信息的价值以及多模态表示在跨任务传递的能力。为了提高样本效率，首先学习了基于神经网络的多传感器数据特征表示。得到的压缩特征向量用作通过强化学习学习到的策略的输入。

将操作任务建模为一个有限时间的离散马尔科夫决策过程 M，状态空间 S，动作空间 A，状态转移动力学 $T: S \times A \rightarrow S$，初始状态分布 ρ_0，回报函数 $r: S \times A$

→ R，时间 T，折扣系数 $\gamma \in$（0,1]，为了确定最优随机策略 $\pi : S \rightarrow P$（A），我们希望最大化期望折扣奖励

$$J(\pi) = \mathbb{E}_{\pi}\left[\sum_{t=0}^{T-1} \gamma r(\mathbf{s}_t, \mathbf{a}_t)\right]$$

自监督多模态表示学习的神经网络结构如下图所示，该网络将来自三个不同传感器的数据作为输入：RGB 图像、32ms 窗口上的 F/T 读数以及末端执行器的位置和速度。它将这些数据编码并融合到一个多模态表示中，基于此，可以学习包含接触操作的控制器。这种表示学习网络是通过自监督形式进行端到端训练的。



我们将具有接触的操作作为一个无模型强化学习问题，研究它在依赖多模态反馈以及在几何、间隙和构型不确定的情况下的性能。由于选择无模型，还消除了对精确动力学模型的需要，这是存在接触的操作中的典型困难。

研究结果：

提出了一种新颖的模型，将异构感官输入编码为多模态表示。一旦经过训练，当用作用于强化学习的浅层神经网络策略的输入时，该表示就保持固定。通过自我监督来训练表示模型，从而无需手动标注。实验表明，需要接触的任务需要视觉和触觉的多模式反馈，此外，还进一步证明了多模态表示可以很好地迁移到其他新任务中。

论文题目：A Magnetically-Actuated Untethered Jellyfish-Inspired Soft Milliswimmer

中文题目：一个受水母启发的磁力驱动软体游泳机器人

论文作者：Ziyu Ren , Tianlu Wang，Wenqi Hu , and Metin Sitti

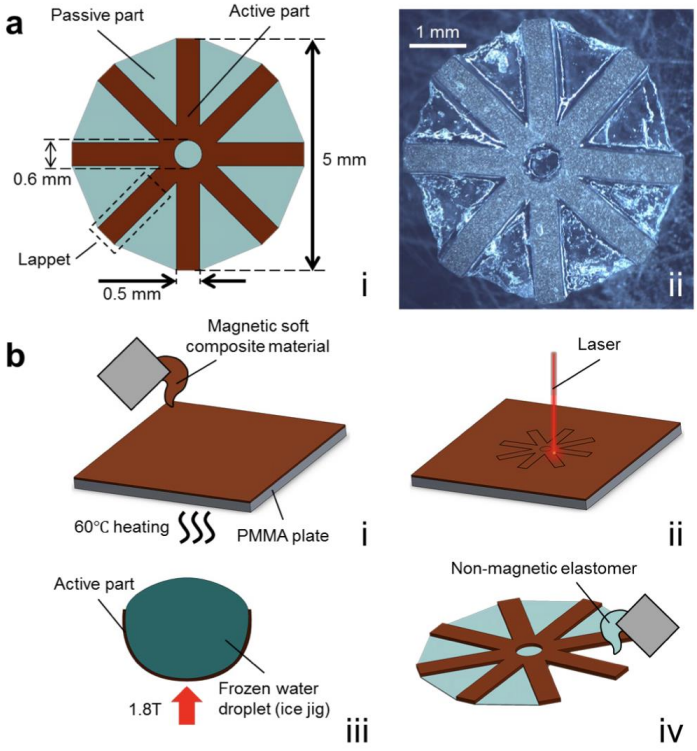
论文出处：Robotics: Science and Systems, 2019

论文地址：http://www.roboticsproceedings.org/rss15/p13.pdf

研究问题：

不受限制的小型软机器人可以用于医疗和生物医学应用。他们可以进入狭小空间并以可编程方式改变形状，以适应非结构化环境并具有多种动态行为。但是，当前的微型软机器人的功能有限，从而限制了它们在医疗中的应用。利用磁性软复合材料形状可编程的优势，提出一种不受束缚的软体机器人，它可以像水母一样在时间和轨迹上不对称地上下跳动，可以通过调节外部振荡磁场的大小，频率和方向来控制其游泳速度和方向。

研究方法：



该机器人的设计如上图所示，机器人主体由两部分组成：主动部分像肌肉一样工作以实现划桨运动，而被动部分则填充了主动部分的间隙，使身体成为连续的流体动力表面。身体的主动部分由软磁性材料制成，可在外部 B 场下变形。通过将钕铁硼（NdFeB）磁性微粒（MQP-15-7，Magnequench；平均直径：5μm）与聚合物（Ecoflex 00-10，Smooth-On Inc.）混合来制备材料，质量比为 1：1。将

该混合物浇铸到涂覆有聚对二甲苯-C 的聚合物（甲基丙烯酸甲酯）（PMMA）板上。聚合物在 60°C 固化形成厚度约为 96μm 的薄膜（下图 b-i）。使用激光切割机从该薄膜上切出主动部分的几何形状（下图 b-ii）。从平板上移开主动部分后，使用移液器将水滴滴在其上。活性部分可以立即包裹水滴并在表面张力作用下形成椭圆形（下图 b-iii）。然后将有效成分放入冰箱进行冷冻，以保持椭圆形的形状。椭圆形主动部分在振动样品磁力计（VSM，EZ7，Microsense）内部被 1.8T 均匀磁场磁化。磁化后，用非磁性弹性体（Ecoflex 00-10）填充主动部分的间隙以形成厚度约为 20μm 的薄层薄膜（下图 b-iv）。最终机器人如下图 a-ii 所示。

研究结果：

提出了一种使用磁性软复合材料制作的软体游泳机器人。只需调节外部磁场的波形，频率和振荡方向即可实现对其控制。已经进行了初步研究以发现其推进速度与输入控制信号之间的关系。当驱动频率增加时，由于流体动力阻尼力，机器人的跳动幅度单调减小。实验数据和模型预测都显示了对于特定控制波形的最佳驱动频率的存在。实验表明，该机器人可用于多种潜在医疗功能。

论文题目：*Robust Visual-Inertial State Estimation with Multiple Odometries and Efficient Mapping on an MAV with Ultra-Wide FOV Stereo Vision*

中文题目：鲁棒多测度视觉惯性状态估计及其在具有超广角立体视觉的微型飞行器上的高效映射

论文作者：M. G. Mu¨ller, F. Steidle, M. J. Schuster, P. Lutz, M. Maier, S. Stoneman, T. Tomic, and W. Sturzl

论文出处：IEEE International Conference on Intelligent Robots and Systems，2018

论文地址：https://ieeexplore.ieee.org/document/8594117

研究问题：

近年来，微型飞行器（MAV）已用于各种各样的应用中。他们能够快速到达感兴趣的点或获得以前难以或不可能到达的视角，这使它们对于诸如勘探，检查，搜索和救援之类的任务变得非常有用。提出了一种配备两对广角立体相机和一个

惯性测量单元（IMU）的多旋翼系统，以实现强大的视觉惯性导航和省时的全向3D 映射，如下图所示。



研究方法：

四个摄像头垂直覆盖了 240 度的立体视野（FOV），这使得该系统也适用于狭窄和狭窄的环境，例如洞穴。在所提出的方法中，从四个广角摄像头合成了八个虚拟针孔摄像头。所得的四个合成针孔立体系统中的每一个都为独立的视觉测距法（VO）提供输入。随后，基于它们与状态估计的一致性，将四个单独的运动估计与来自 IMU 的数据融合。

研究结果：

提出了配备有四个广角摄像机的 MAV。多达 240°的垂直立体视野使 MAV 能够感知其下方，上方和前方的对象，这与避障，路径规划和有效的映射等任务有关。实验表明，由四个具有独立关键帧的立体测距仪提供的鲁棒运动估计，也可以从较大的视野中受益，从而可以进行状态估计。

10.5 机器人进展

- 机器人学习

在 AI 兴起的时代，机器人拥有了一种新型的学习方式：深度强化学习。这一新方式借助通用化的神经网络表示，处理复杂的传感器输入，来让机器人从自己的经验活动中直接学习行为。相比传统方式，它解放了工程设计人员们的双手，不再需要程序员们手动设计机器人每一个动作的每一项精确参数。但是，现有的强化学习算法都还不能够适用于有复杂系统的机器人，不足以支撑机器人在短时

间内就学习到行为，另外在安全性上也难以保障。针对这种困境，2019 年初，谷歌 AI 与 UC 伯克利大学合作研发了一种新的强化学习算法：SAC（Soft Actor-Critic）。SAC 非常适应真实世界中的机器人技能学习，可以在几个小时内学会解决真实世界的机器人问题，而且它的一套超参数能够在多种不同的环境中工作，效率十分之高。SAC 的开发基于最大熵强化学习这个框架。此框架尝试让预期回报最大化，同时让策略的熵最大化。一般而言，熵更高的策略具有更高的随机性。从直觉上看，这意味着，最大熵强化学习能取得高回报策略中具有最高随机性的那个策略。SAC 学习一个随机策略，这个策略会把状态映射到动作，也映射到一个能够估计当前策略目标价值的 Q 函数，这个 Q 函数还能通过逼近动态编程来优化它们。SAC 通过这样的方式，来让经过熵强化的回报最大化。此过程中，目标会被看作一个绝对真的方法，来导出更好的强化学习算法，它们有足够高的样本效率，且表现稳定，完全可以应用到真实世界的机器人学习中去。

- 机器人应用

2019 年 6 月，亚马逊在 MARS 人工智能大会上最新发布的仓库机器人 Pegasus，该机器人已正式加入亚马逊 Kiva 机器人行列。Pegasus 是一种新型包裹分拣机器人，外观上看，Pegasus 机器人十分类似亚马逊既有的 Kiva 机器人，外观还是橙色不变，2 英尺高，3 英尺宽，约相当于一个手提包的大小。Pegasus 机器人更像是对原有 Kiva 机器人的改良版，在原有机器人底座上增加了一个载货平台+皮带传送带对各个包裹进行分类和移动，有助于最大限度地减少包裹损坏并缩短交货时间。Pegasus 机器人可以自主将右侧盒子放在正确的位置。仓库作业人员将包裹扫描完放到 Pegasus 机器人上，Pegasus 机器人载着包裹到指定地点。机器人配备的摄像机可以感知任何意外障碍。到了指定地点，机器人载货平台上的传送带将包装从机器人上移开，然后包裹沿着滑槽向下移动，准备送出。机器人在大约 2 分钟内完成整个包裹运送过程。据亚马逊介绍，Pegasus 机器人具有与 Kiva 机器人驱动器相同的容量。Pegasus 机器人目前已经在在丹佛分拣中心上线的六个多月，行驶约 200 万英里，经测试，它能将当前系统的包裹分拣错误率大幅降低 50%。本次 MARS 人工智能大会上，除了推出 Pegasus 机器人，亚马逊还发布了一种大型模组化运输机器人 Xanthus。依据上方安装的模组，执行多种不同的任务 Xanthus 拥有透过改变上方配备，胜任不同任务的能力。相较过

去使用的系统，Xanthus 不仅用途更为广泛，体积也只有前辈的 1/3，成本甚至直接砍半。

- 机器人平台

如何将机器人技术落地、实践商业化一直是备受关注的问题。波士顿动力的策略是要希望其成为平台公司，通过授权或开源方式，使其技术能被广为被使用。2018 年这个传言似乎得到了证实，在《连线》杂志举办的峰会上，波士顿动力创始人暨首席执行官 Marc Raibert 指出，他们的定位是成为平台公司，让生态圈包括第三方伙伴、客户，一起来找到技术真正适合使用的地方。Marc Raibert 表示波士顿动力在开发机器人时是以“平台”的概念来出发，客户可以增加硬件，例如手臂及其他组件，“当然，我们也可以针对单一领域打造一个有特殊应用的机器人方案，但我们不知道哪一个领域合适，所以我们从平台的角度出发，希望生态圈帮我们一起来找到技术真正可落地之处”、“我们要打造的是‘通用用途的平台’（general purpose platform），让第三方伙伴、客户、波士顿动力自己的应用开发团队，可以一同来设计产品以符合定制化需求。

11 数据库技术

11.1 数据库概念

数据库是按一定的结构和规则组织起来的相关数据的集合，是综合各用户数据形成的数据集合,是存放数据的仓库（我国数据库的发展现状与趋势--陈黎）。随着计算机技术与网络通信技术的快速发展,数据库技术已经成为当今信息社会中对大量数据进行组织与管理的重要技术手段,是网络信息化管理系统的基础。目前,新一代数据库系统不仅保持和继承了传统数据库系统的各项功能,支持知识管理、数据管理和对象管理,而且还对其它应用系统开放,在网络上支持标准网络协议,具有良好的可连接性、可移植性、可互操作性和可扩展性。

数据库技术与网络通信技术、人工智能技术、面向对象程序设计技术、并行计算技术等互相渗透和结合,是当前数据库技术应用的主要特征，当前具有此类特征的新型数据库系统包括如分布式数据库系统、知识库系统和主动数据库系统等[64]。

- 分布式数据库系统

分布式数据库系统由一组分布在网络中的不同计算机上的数据组成。系统中每台服务器有自己的数据库系统及若干台客户机，3 台服务器之间通过网络相连。网络中的每个节点具有独立处理的能力,可以执行局部应用,同时每个节点也能通过网络子系统执行全局应用。用户通过客户机可以对本地服务器中的数据库执行某些应用,也可以对 2 个或 2 个以上节点中的数据库执行某些应用。

- 知识库系统

知识库系统是数据库和人工智能两种技术相结合的产物,简单来说就是在数据库技术中引入人工智能技术,把数据库看作一个人工智能系统,利用人工智能技术来提高 DBMS 的表达、推理和查询能力。其功能主要体现在数据库系统推理能力的扩充、语义知识的引入、知识的获取、知识和数据的有效组织及管理等方面；而效率则体现在数据库对用户查询的快速响应和查询优化上。

- 主动数据库

在实际应用中，如计算机集成制造系统、管理信息系统、办公自动化系统等通常希望数据库系统在紧急情况下能根据数据库的当前状态主动做出反应，并执行相应的操作，向用户提供特定信息。主动数据库是相对于传统数据库的被动性而言的，其主要任务是提供对紧急情况的及时反应能力，同时提高数据库管理系统的模块化程度。主动数据库通常采用的方法是在传统数据库系统中嵌入事件、条件、动作规则,在某一事件发生时引发数据库管理系统去检测数据库的当前状态；判断是否满足设定的条件,如果条件满足便触发规定动作执行。

11.2 数据库技术历史

在处理器、计算机内存、计算机存储和计算机网络等领域的技术进步之后，数据库的大小、性能和性能以及它们各自的 DBMS 都以数量级增长。数据库技术的发展可以根据数据模型或结构划分为三个时代：导航数据库时代、SQL/关系时代、后关系时代。其中后关系包含了面向对象型数据库及新型数据库技术的发展。下面按照时间顺序介绍每个时期的主流数据库技术[65]。

数据库的起源

60 多年前，数据以一种原始与粗糙的方式进行管理。通过大量的分类、比较和表格绘制的机器运行数百万穿孔卡片来进行数据的处理，其运行结果在纸上打印出来或者制成新的穿孔卡片。而数据管理就是对所有这些穿孔卡片进行物理的储存和处理。但是这样并不能满足对日益增长数据管理控制的需求。数据以某种抽象的方式组织起来，使存储和检索更加有效，这是不可避免的。

20 世纪 60 年代，基于导航的数据库管理系统

20 世纪 60 年代计算机开始广泛地应用于数据管理，但对数据的共享提出了越来越高的要求。传统的文件系统已经不能满足人们的需要。能够统一管理和共享数据的数据库管理系统（DBMS）应运而生。数据模型是数据库系统的核心和基础，各种 DBMS 软件都是基于某种数据模型的。所以通常也按照数据模型的特点将传统数据库系统分成网状数据库（Network database）、层次数据库（Hierarchical database）和关系数据库（Relational database）三类。

20 世纪 70 年代初，关系数据库的发展

1970 年，Codd 撰写了大量论文，概述了数据库建设的新方法，最终实现了大型共享数据库的开创性关联数据模型。Codd 描述了一个用于存储和处理大型数据库的新系统。关系模型通过将数据分解为一系列规范化表，可选元素从主表移出到只有在需要时才占用空间的位置。可以在这些表中自由插入，删除和编辑数据，DBMS 可以进行任何维护，以向应用程序/用户呈现表视图。关系模型还允许数据库的内容发展，而不必不断重写链接和指针。关系部分来自引用其他实体的实体，称为一对多关系，如传统的层次模型，以及像导航（网络）模型这样的多对多关系。因此，关系模型既可以表达层次结构模型，也可以表示导航模型，也可以表示其原生表格模型，从而根据应用程序的需要许对这三种模型进行纯粹的或组合的建模。

20 世纪 70 年代后期，SQL 的发展

在 20 世纪 70 年代早期，IBM 开始基于 Codd 的概念松散地开发原型系统 SystemR。第一个版本于 1974 年 5 月准备就绪，然后在多表系统上开始工作，在该系统中可以拆分数据，以便记录的所有数据不必存储在单个大块“块”。随后的多用户版本在 1978 年和 1979 年由客户进行了测试，此时添加了一种标准化的查询语言——SQL（Structured Query Language）。结构化查询语言是高级的非过程化编程语言，允许用户在高层数据结构上工作。它不要求用户指定对数据的存放方法，也不需要用户了解具体的数据存放方式，所以具有完全不同底层结构的不同数据库系统,可以使用相同的结构化查询语言作为数据输入与管理的接口。结构化查询语言语句可以嵌套，这使它具有极大的灵活性和强大的功能。

20 世纪 90 年代，面向对象数据库

数据库面临的首要问题之一就是非字符（字母数字）数据的崛起。越来越多的图像、声音文件、地图和视频需要存储、操作和检索。这导致了大量的创新,但也支离破碎的标准。这时面向对象数据库应运而生。面向对象数据库系统（OODBS）支持定义和操作 OODB，应满足两个标准：首先它是数据库系统，其次它也是面向对象系统。第一个标准即作为数据库系统应具备的能力（持久性、事务管理、并发控制、恢复、查询、版本管理、完整性、安全性）。第二个标准

就是要求面向对象数据库充分支持完整的面向对象（OO）概念和控制机制。综上所述，我们将面向对象数据库简写为：面向对象数据库=面向对象系统+数据库能力。

21 世纪初，NoSQL 与 NewSQL 技术的发展

NoSQL，泛指非关系型的数据库。NoSQL 数据库的产生是为了解决大规模数据集合多重数据种类带来的挑战，尤其是大数据应用难题。NoSQL 早期就有人提出，发展至 2009 年趋势越发高涨。NoSQL 的拥护者们提倡运用非关系型的数据存储，相对于铺天盖地的关系型数据库运用，这一概念无疑是一种全新的思维的注入。NewSQL 一词是由 451Group 的分析师 Matthew 在研究论文中提出的，是对各种新的可扩展/高性能数据库的简称，这类数据库不仅具有 NoSQL 对海量数据的存储管理能力，还保持了传统数据库支持 ACID 和 SQL 等特性。

图数据库也是非关系型的数据库的一种，是一个使用图结构进行语义查询的数据库，它使用节点、边和属性来表示和存储数据。该系统的关键概念是图，它直接将存储中的数据项，与数据节点和节点间表示关系的边的集合相关联。这些关系允许直接将存储区中的数据链接在一起，并且在许多情况下，可以通过一个操作进行检索。图数据库将数据之间的关系作为优先级。查询图数据库中的关系很快，因为它们永久存储在数据库本身中。可以使用图数据库直观地显示关系，使其对于高度互连的数据非常有用。

NewSQL 属于分布式数据库，分布式数据库是指利用高速计算机网络，将物理上分散的多个数据存储单元，连接起来组成一个逻辑上统一的数据库。分布式数据库的基本思想是，将原来集中式数据库中的数据，分散存储到多个通过网络连接的数据存储节点上，以获取更大的存储容量和更高的并发访问量。

11.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为数据库领域全球学者分布情况：



图 11-1 数据库领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布在其东西海岸；欧洲也有较多的人才分布；亚洲的人才主要集中在我国东部；其他诸如非洲、南美洲等地区的学者非常稀少；数据库领域的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，数据库领域中男性学者占比 91.7%，女性学者占比 8.3%，男性学者占比远高于女性学者。

数据库领域学者的 h-index 分布下图所示，分布情况大体呈阶梯状，其中 h-index 小于 20 区间的人数最多，有 921 人，占比 45.4%，50-60 区间的人数最少，有 87 人。

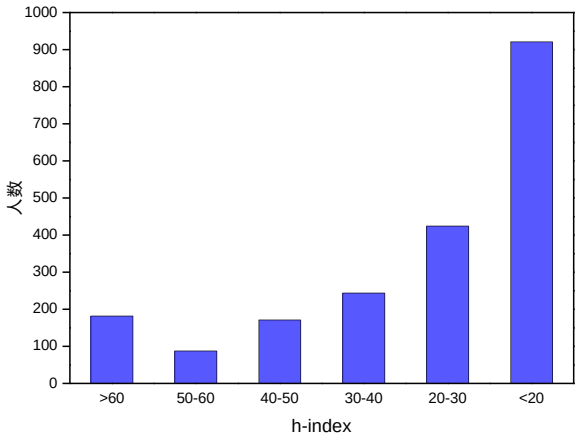


图 11-2 数据库领域学者 h-index 分布

- 中国人才分布

我国专家学者在数据库领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是珠三角和长三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等地相比，中国在数据库领域学者数量较多但差距不大。



图 11-3 数据库领域中国学者分布

中国与其他国家在数据库领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 11-1 数据库领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	259	9679	37	584
中国-新加坡	89	3968	45	139
中国-澳大利亚	47	1565	33	81
中国-丹麦	25	318	13	45
中国-加拿大	17	290	17	41

中国-法国	15	181	12	14
中国-希腊	11	911	83	17
中国-卡塔尔	10	206	21	17
中国-韩国	7	227	32	16
中国-瑞士	6	66	11	21

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在数据库领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席；中国与希腊合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中希合作达到了较高的水平。

11.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

ACM SIGMOD International Conference on Management of Data

International Conference on Very Large Data Bases

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，大数据（big data）、数据库系统（database systems）、解析法（analytics）是本领域中最热的关键词。



论文题目：Self-Driving Database Management Systems

中文题目：自动驾驶的数据库管理系统

论文作者：Andrew Pavlo, Gustavo Angulo, Joy Arulraj and, Haibin Lin, Jiexi Lin, Lin Ma, et al.

论文出处：7th Biennial Conference on Innovative Data Systems Research (CIDR) – CIDR 2017

论文地址：https://www.pdl.cmu.edu/PDL-FTP/Database/p42-pavlo-cidr17.pdf

研究问题：

在过去的二十年中，研究人员和数据库系统供应商都尝试开发了各式辅助工具以在数据库系统的调优和物理设计等各个方面协助数据库管理员（Database Administrator，DBA）。但是，大多数的工作还是不足够完善的，因为它们仍然需要 DBA 对数据库的任何更改做出最终决定，并且是在问题发生后解决问题的反应性措施。尤其是随着云数据库的发展，不需要人工干预的 DBMS 就成为了一个迫切的需求，于是能“自动驾驶”的数据库管理系统（Database Management System，DBMS）便成为了必然的选择。真正地能“自动驾驶”的数据库管理系统所需要的是一种为自治操作而设计的新体系结构。与早期的各种 DBMS 不同的是，该类系统的所有方面都由集成的计划组件控制，该组件不仅可以针对当前工作负载（Workload）优化系统，而且还能预测未来的工作负载的变化趋势，以便系统可以相应地进行准备。这样，DBMS 可以支持所有以前的调优技术，而无需人工确定正确的方式和适当的时间来部署它们。

研究方法：

该论文指出“自动驾驶”（Self-Driving）的 DBMS 会面临以下三点问题：

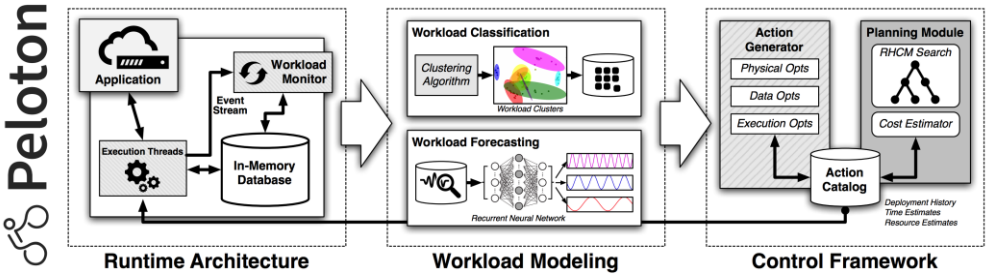
1）需要理解业务应用的工作负载（Workload）：简单的来说，我们可以将应用分为联机事务处理过程（Online Transaction Processing，OLTP）、联机分析处理过程（Online Analytical Processing，OLAP）和混合交易-分析处理过程（Hybrid Transaction-Analytical Processing，HTAP）三种类型。不同类型的应用需要不同的设置，OLTP 以交易型场景为主，数据采用行存储，便于快速的对一条记录进行增删改查；HATP 则是混合型场景，不仅有频繁的条目增删改查，也有大量的单列分析应用。

2）需要能够预测资源的使用趋势：有助于系统根据未来可能的情况和资源需求进行动态资源调配以确保对性能的影响保持最低。由于很多应用的使用模式与人的活动模式密切相关，数据库管理员通常会在业务低峰期进行一些优化操作，避免影响应用的服务质量。不可否认，现实中难免会存在一些工作负载的异常状况是 DBMS 无法避免的。但是，这些预测模型是可以尽早地提供一些预警使得 DBMS 尽可能完成相应操作。有了预测模型之后，DBMS 可以通过一些调优操作使数据库在预期工作负载上工作的更好。自动驾驶的 DBMS 可以支持的操作主要有如下几种：（1）数据库的物理设计（Database's Physical Design）；（2）数据组织形式的变更（Data Organization）；（3）DBMS 运行时的行为（DBMS's Runtime Behavior）。这三类操作的详细内容如下表所示，这里不一一赘述。

	Types	Actions
PHYSICAL	Indexes	AddIndex, DropIndex, Rebuild, Convert
	Materialized Views	AddMatView, DropMatView
	Storage Layout	Row→Columnar, Columnar→Row, Compress
DATA	Location	MoveUpTier, MoveDownTier, Migrate
	Partitioning	RepartitionTable, ReplicateTable
RUNTIME	Resources	AddNode, RemoveNode
	Configuration Tuning	IncrementKnob, DecrementKnob, SetKnob
	Query Optimizations	CostModelTune, Compilation, Prefetch

3）要一个灵活的，内存级的架构，便于快速应用优化操作：如果 DBMS 不够灵活，那么模型推荐的一些优化措施不能及时地实施，也就失去了优化的意义。同时，也不能算得上是真正意义的“自动驾驶”数据库。

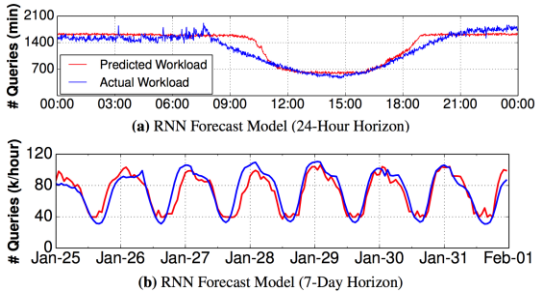
现有的 DBMS 对于自动化操作的支持都不太好友，经常需要通过重启来使得配置生效。针对这些，Peloton 采用了多版本并发控制的模式，可以在不阻塞 OLAP 的情况下，提供 OLTP 的服务，并采用了无锁数据结构的内存存储管理器，能支持快速地执行 HTAP 操作。



上图是 Peloton 系统的“自动驾驶”工作流程的概述。除了环境设置（例如，内存阈值，目录路径）以外，Peloton 的目标在没有任何人工提供指导信息的情况下能有效的运行。系统会自动学习如何改善应用程序查询和事务的延迟（Latency）。由于延迟是反映 DBMS 性能的最重要指标，Peloton 系统主要将延迟作为优化的指标。除此之外，该系统还可以优化在分布式环境中其它的重要指标，例如服务成本和能源。 Peloton 还包含一个嵌入式监视器，该监视器遵循已执行查询的系统内部事件流。在 Workload 建模部分，该系统首先对工作负载进行聚类，这样子可以减少 DBMS 管理的预测模型的数量，降低了预测应用行为的复杂度，同时采用了 DBSCAN 聚类算法，使得效果进一步提升。在工作负载预测部分，该系统可以预测周期性的工作负载和数据变化趋势，能更好的提供服务性能。例如，每当 DBMS 执行完一个查询操作之后，对系统该查询的聚类中心进行表示，并按照预定义的统计区间去记录这些查询的请求次数。基于此，系统可以使用这些数据来预测“未来”的查询数量，并可能提前执行相关的优化。通过系统的控制模块，可以提供对系统的持续监控并选择合适的优化措施去提升应用的服务性能。

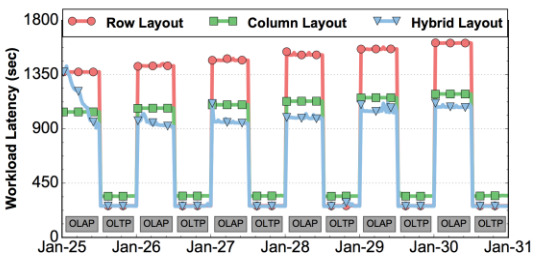
研究结果：

Peloton 系统首先对工作负载的不同粒度和时间跨度进行预测。如下图（a）和（b）所示，Peloton 系统对于短期内的查询预测是比较准确的，但是长期的预测效果不佳。从下图（c）中可以看出，该实验主要测试了 OLAP、OLTA 和 HTAP 场景下的优化效果。可以看出，系统可以看出 Peloton 会随时间的的推移而收敛。例如，在第一个 OLAP 操作之后，系统会将存储的元组改为列存储的布局，这是 OLTP 查询的理想选择，因此其延迟下降。本篇论文开创性地提出了 Self-Driving DBMS 的概念，并提出了一个初步的系统验证了该概念的可行性。



(a) RNN Forecast Model (24-Hour Horizon)

(b) RNN Forecast Model (7-Day Horizon)



(c) Adaptive Storage Layouts

论文题目：*Neo: A Learned Query Optimizer*

中文题目：Neo：学习型查询优化器

论文作者：Ryan Marcus, Parimarjan Negi, Hongzi Mao, Chi Zhang, Mohammad Alizadeh, Tim Kraska, Olga Papaemmanouil, Nesime Tatbul

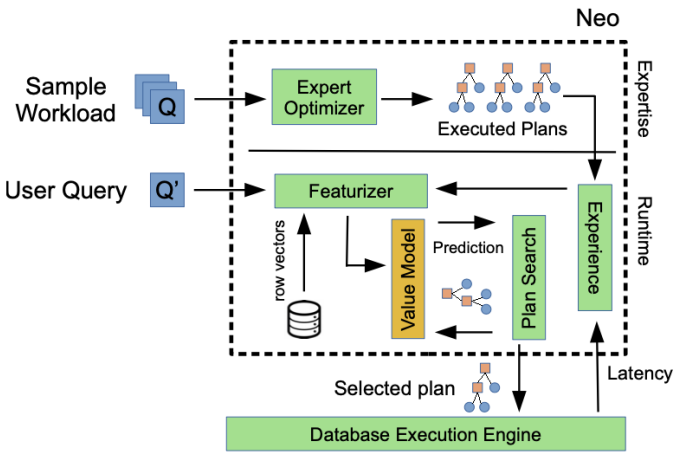
论文出处：45th International Conference on Very Large Data Bases – VLDB 2019

论文地址：http://www.vldb.org/pvldb/vol12/p1705-marcus.pdf

研究问题：

查询优化是数据库系统中最具挑战性的问题之一。尽管在过去几十年中取得了进步，但是查询优化器仍然是极其复杂的组件，需要针对特定工作负载和数据集进行大量手动调整。受此缺点的激励，并受到将机器学习应用于数据管理挑战的最新进展的启发，我们引入了 Neo（神经优化器），这是一种新型的基于学习的查询优化器，它依赖于深度神经网络来生成查询执行计划。Neo 从现有的优化器中引导其查询优化模型，并继续从传入的查询中学习，以成功为基础，从失败中学习。此外，Neo 会自然地适应基础数据模式，并且对估计错误具有鲁棒性。实验结果表明，即使从 PostgreSQL 之类的简单优化器启动后，Neo 仍可以学习一种模型，该模型可提供与最新的商业优化器相似的性能，甚至在某些情况下甚至可以超越它们。

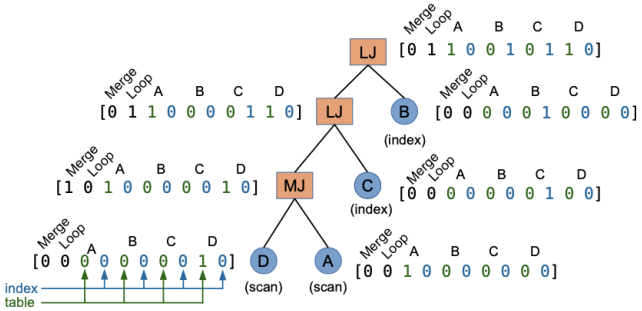
研究方法：



Neo 的设计模糊了传统查询优化器主要组件（基数估计，成本模型和计划搜索算法）之间的界限。Neo 将基数估计和成本模型两个功能组合在一个价值网络中，在价值网络的指导下，Neo 对查询计划空间进行了简单的搜索以制定决策。随着 Neo 发现更好的查询计划，Neo 的价值网络会不断完善，将搜索重点放在更好的计划上，这是一个强化学习的过程。强化学习过程将会持续进行，直到 Neo 的决策政策收敛为止。

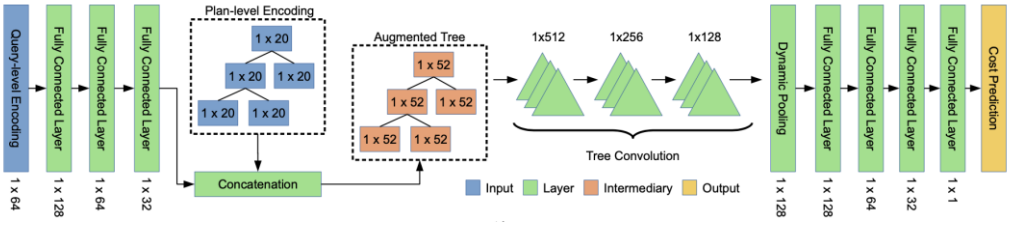
Neo 的运行分为两个阶段：初始阶段和运行时阶段，如上图所示。在初始阶段，Neo 使用传统的查询优化器为样本工作负载中的每个查询创建查询执行计划。这些执行计划及其延迟会添加到 Neo 的经验中。这些经验被用作价值模型训练的起点。建立初始价值模型可预测给定部分或完整执行计划的最终执行时间。在运行时阶段，Neo 使用价值模型在查询执行计划的空间中进行搜索，并以最短的预测执行时间发现计划。随着 Neo 优化更多查询，价值模型将得到改进并针对用户数据库进行量身定制。

为了对查询进行合理的表示，Neo 对查询本身和查询计划分别进行了编码。查询编码由两个部分组成，首先将查询的连接图编码为邻接矩阵，由于该矩阵是对称的，因此仅对上三角部分进行编码；为了对列谓词向量进行编码，Neo 基于自然语言处理模型 word2vec，将列谓词向量中的每个条目编码为包含与谓词相关的包含语义信息的向量。这种编码需要在数据库中的数据上建立模型进行训练，行中频繁出现的值被映射成相似的向量。查询计划编码将执行计划的每个节点编码为一个向量，并按照节点在查询计划中的位置将这些向量组织成向量树，如下图所示。



Neo 的价值网络结构如下图所示，Neo 通过称为树卷积的技术从树状查询计划表示中进行学习。查询编码首先通过多个全连接层的，每个层的大小都减小。

第三个全连接层输出的矢量与树状查询计划编码的每一个向量连接在一起。与查询编码合并以后的树状查询计划编码接着就通过几个树卷积层，之后，进入动态池化层，将树结构展平为单个向量。最后，几个附加的全连接层用于将此向量，并将此向量映射成单个值，用作输入计划的代价估计值。



研究结果：

如下图所示，显示了 Neo 在测试数据集上进行 100 次训练迭代后 Neo 的相对性能（越低越好）。例如，利用 PostgreSQL 和 JOB 工作负载，Neo 生成的查询仅比原始 PostgreSQL 优化器创建的查询占用平均执行时间的 60%。此外，对于 SQL Server 以及 JOB 和 Corp 工作负载，Neo 生成的查询计划也比 SQL Server 商业优化程序创建的计划快 10%。总体而言，该实验表明 Neo 的表现，与开源优化器以及同类的商业产品一样好，有时甚至好于这些产品。



论文题目：*SageDB: A learned database system*

中文题目：SageDB：学习型数据库系统

论文作者：T Kraska, M Alizadeh, A Beutel, EH Chi, J Ding, et al.

论文出处：9th Biennial Conference on Innovative Data Systems Research (CIDR) – CIDR 2019

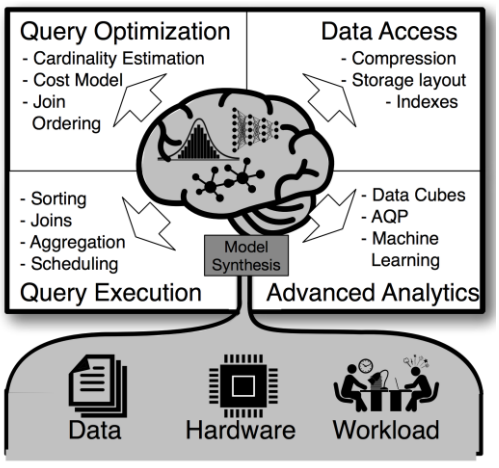
论文地址：https://ai.google/research/pubs/pub47669

研究问题：

现代数据处理系统通常被设计为通用的，因为它们可以处理各种不同的数据模式、数据类型和数据分布，并旨在通过使用优化器（Optimizer）和代价模型（Cost Model）来提供对数据的高效访问。这种通用数据库参数设定自然导致数据库系统无法利用特定应用程序和用户数据的特征来进行特定的优化。该论文的研究者提出了 SageDB，一种对新型数据处理系统。该系统高度专注于通过代码合成和机器学习来开发应用程序。通过对数据分布，工作负载和硬件进行建模，SageDB 可以学习数据的结构以及最佳访问方法和查询计划。这些学习的模型通过代码合成被深深地嵌入到数据库的每个组件中。

研究方法：

SageDB 的核心思想是建立一个或多个数据和工作负载分布有关的模型，并基于这些模型自动地为数据库系统的所有组件构建最佳的数据结构和算法。这种称为“合成数据库”的方法将使每个数据库组件的实现都专用于特定的数据库，查询工作负载和执行环境，从而使数据库系统的性能大大提高。如下图所示，SageDB 从总体上构建了一个能对数据负载分布感知的模型，并自动地为数据库每个组件选择合适的算法和数据特征。尽管这套系统目前还停留在理论论证层面，还有很多突出的问题没有彻底解决，但是它为未来的数据库架构和组织管理方法提供了非常大胆的方向。当每个组件乃至于决策系统都具有自主学习和优化能力，数据库将能更好的为用户服务。



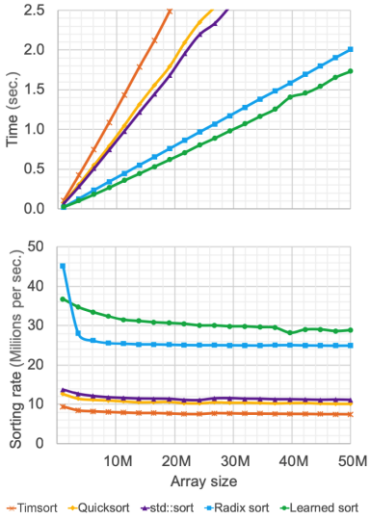
接下来，我们以学习排序模型为例，详细介绍 SageDB 的工作原理。加快排序的基本思想是使用 CDF 模型将记录大致按排序顺序放置，再对该结果进行更

正成几乎完美排序的结果。根据模型的执行代价及其精度，这种排序技术可以在性能方面优于其他分类技术。例如，假定以下查询：“SELECT * FROM customer c ORDER BY c.name “，并假定表的大小为 N，在 c.name 这一列上有一个学习索引（Learned Index），且该列的数据是以 customer id 而不是名称的顺序存储。为了快速按 customer name 的数据进行排序，我们可以通过 CDF 模型将每条记录的键 k 放置到合适的位置，这样子可以得到一个大致的排序结果。至此，我们将每个记录映射到一个在输出数组中的位置，如果发生冲突，我们将其存储一些溢出数组（算法 1 中的第 5-10 行）。在为了减少碰撞次数，我们可以改为分配一个比 N 大 m 倍的输出数组（例如 m = 1.37），然后删除此映射中所有为空的位置。注意，在这种情况下，预测的位置的计算方式为 pos = F（k）* m * N（算法 1 中的第 6 行）。假设如果我们将存储桶设为缓存行的倍数，那么我们可以在桶中进行快排，同时减少冲突的可能性。一个类似的想法用作 Cuckoo Hashing 的一部分。无论哪种情况，如果我们模型是非单调的，我们必须使用高效的 localsort 算法（即插入排序）以纠正所有排序错误（算法 1 中的第 12 行）。最后，我们对溢出数组进行排序（算法 1 中的第 13 行）并且合并排好序的数组和溢出数组，同时删除空元素（算法 1 中的第 15 行）。

Algorithm 1 Learned sorting algorithm

Input a - the array to be sorted
Input F - the CDF model for the distribution of a
Input m - the over-allocation parameter
Output o - the sorted version of array a

```
1:  procedure LEARNED-SORT(a, F, m)
2:      o ← [∞] * (a.length * m)
3:      s ← {}
4:      // STEP 1: Approximate ordering
5:      for i in a do
6:          pos ← F(i) * a.length * m
7:          if o[pos] = ∞ then
8:              o[pos] ← i
9:          else
10:             s ← s ∪ {i}
11:     // STEP 2: Touch-up
12:     INSERTION-SORT(o)
13:     QUICKSORT(s)
14:     // STEP 3: Merging
15:     return MERGE-AND-REMOVE-EMPTY(o, s)
```



研究结果：

以学习排序为例，该论文以数组大小为变量，对比了多种排序方法。从上图的结果可以看出，在排序时间（Time）和排序比率（Sorting Rate）两个指标上，SageDB 提出的 Learned Sort 方法均比其他方法好，验证了所提方法的有效性。

SageDB 为数据库的设计提供了一种新的思路，并在初步的实验上验证了这种思路的可行性。

论文题目：*QTune: A Query-Aware Database Tuning System with Deep Reinforcement Learning*

中文题目：QTune：一种基于深度强化学习的对查询感知的数据库调优系统

论文作者：Guoliang Li, Xuanhe Zhou, Shifu Li, Bo Gao.

论文出处：45th International Conference on VeryLarge Data Bases – VLDB 2019

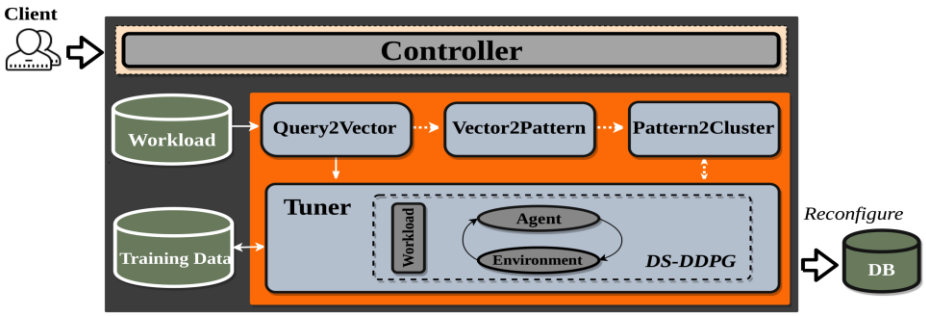
论文地址：http://www.vldb.org/pvldb/vol12/p2118-li.pdf

研究问题：

传统的数据库调优通常依靠雇佣专家（DBA）来完成：专家针对指定的负载，在线下反复进行瓶颈检测、参数调整和性能对比，直至达到满意效果，这是一项非常耗时的工作，而且严重依赖于 DBA 自身的经验和知识。此外，在云数据库环境下，要面临更加繁多的数据库状态和负载类型，极大地增加了这项工作的难度。虽然目前有一些自动配置工具可以供使用，但是这些调优工具多是由特定供应商创建，很难支持其他 DBMS，如 MySQLTuner。此外，这些调优工具都是基于有限的规则，很容易重复推荐错误的方案，无法自动的从以前的调优工作中获得知识来优化规则。因此该论文的研究者提出了 QTune，一种基于学习的数据库自动调优系统。该系统高度专注于利用历史数据学习负载和数据库环境到最优配置的映射策略，从查询编码、参数学习、多粒度调参三个方面提供智能的数据库调优服务。

研究方法：

QTune 是一套对查询感知的数据库调优系统，其核心思想是基于深度强化学习算法学习在不同的环境条件下推荐参数的策略。如下图所示，QTune 主要从三个方面提供智能的数据库调优服务。



其一，为了提高对不同负载的适应能力，QTune 在查询计划级别对作业进行编码，对执行开销作预评估。其二，为了提高调优表现，QTune 的调优模型基于一种面向调优问题的深度强化学习方法——DS-DDPG，不仅基于 Actor-Critic 算法，大大提高学习效率，而且能够根据当前负载、数据库状态特征综合决策，提高调优表现的同时增强了对不同负载、数据库实例等的适应能力。其三，为了更好地满足不同用户对数据库性能的需求，QTune 增加 Query2Cluster 模块，先对作业进行配置类别划分、批量聚类，再根据结果进行查询簇级别的批量调优，能够有效地平衡吞吐量和延迟的优化程度。此外，QTune 还支持查询级别、负载级别的调优模式，分别满足延迟敏感型、吞吐量敏感性用户的需求。

接下来，我们详细介绍 QTune 是如何训练调优模型的。整个调优模型的训练分为两个部分。首先，预测模型（Predictor）是一个五层的神经网络，负责将负载特征（v）转换成数据库的状态变化量ΔS。因此，QTune 用批量梯度下降（Batch Gradient Descent）的方法训练预测模型：批量梯度算法每一轮累积所有样本的误差值。对于负载样本 W，QTune 将 W 解析成特征向量 v；然后预测模型输入特征向量 v、数据库外部状态 s、内部配置 I，输出估计的状态变化量ΔS（函数 TrainPredictor 中的第 4 行）；然后计算ΔS相对于真实值得平方差值，并累加到总误差 E 上（函数 TrainPredictor 中的第 5 行）。预测完每个样本后，QTune 用整体的平均误差 E 更新网络权重（函数 TrainPredictor 中的第 6 行）。其次，代理模型（Agent）用于决定如何调整数据库参数，其训练采用 Actor-Critic 算法。在代理模型中，行动方（Actor）负责实际调参，而评判方（Critic）负责给每个调参决策打分。训练中，对于每个样本（S'_i, A_i, R_i），A_i 是一个推荐的参数集合，S'_i 是查询在 A_i 下执行后的数据库状态，R_i 是查询执行的表现增益。对于每个（S'_i, A_i），评判方先输出一个 Q 值（函数 TrainAgent 中的第 5 行），行动方根据 Q 值更新自己的网络权重；然后评判方根据蒙特卡洛理论用 R_i 计算长期受益 Y_i（函

数 TrainAgent 中的第 6 行），用 Y_i 计算误差值 L；最后 QTune 用误差值 L 更新评判方的网络权重（函数 TrainAgent 中的第 7 行）。

Algorithm 1: Training DS-DDPG

Input: U: the query set $\{q_1, q_2, \cdots, q_{|U|}\}$
Output: π_P, π_A, π_C

1 Generate training data T_P;
2 TrainPredictor(π_P, T_P);
3 Generate training data T_A;
4 TrainAgent(π_A, π_C, T_A);

Function TrainPredictor(π_P, T_P)

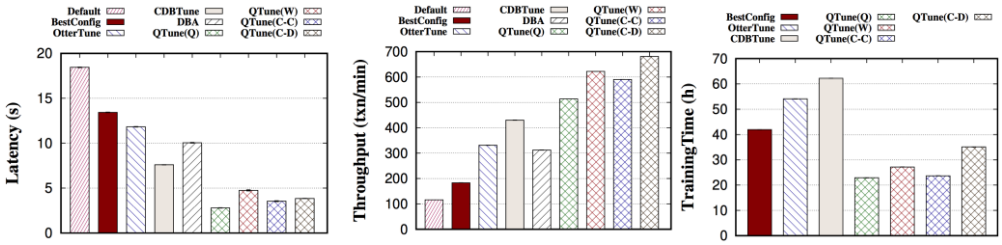
Input: π_P: The weights of a neural network; T_P: The training set

1 Initiate the weights in π_P;
2 **while** *!converged* **do**
3 **for** *each* $(v, S, I, \Delta S) \in T_P$ **do**
4 Generate the output G of $\langle v, S, I \rangle$;
5 Accumulate the backward propagation error: $E = E + \frac{1}{2}||G - \Delta S||^2$;
6 Compute gradient $\nabla_{\theta_\pi}(E)$, update weights in π_P;

Function TrainAgent(π_A, π_C, T_A)

Input: π_A: The actor's policy; π_C: The critic's policy; T_A: training data

1 Initialize the actor π_A and the critic π_C;
2 **while** *!converged* **do**
3 Get a training data $T_A^1 = (S'_1, A_1, R_1), (S'_2, A_2, R_2), \ldots, (S'_t, A_t, R_t)$;
4 **for** $i = t - 1$ *to* 1 **do**
5 Update the weights in π_A with the action-value $Q(S'_i, A_i|\pi_C)$;
6 Estimate an action-value $Y_i = R_i + \tau Q(S'_{i+1}, \pi_A(S'_{i+1}|\theta^{\pi_A})|\pi_C)$;
7 Update the weights in π_C by minimizing the loss value $L = (Q(S'_i, A_t|\pi_C) - Y_i)^2$;



研究结果：

如上图所示，经过在 TPC-H、JOB、Sysbench 等标准测试集的测试，QTune 在延迟、吞吐量、训练时间等方面都有较大的提升，如同比业界最优的 CDBTune 吞吐量平均有 23.4%的提高，延迟平均有 29.3%的降低。目前 QTune 支持 GaussDB、PostgreSQL、MySQL 三种关系型数据库和 MongoDB 一种非关系型数据库，主要在单节点上提供自调优服务。

论文题目：*An End-to-End Learning-based Cost Estimator*

中文题目：一个端到端的基于学习的代价估计器

论文作者：Ji Sun, Guoliang Li.

论文出处：46th International Conference on Very Large Data Bases (VLDB 2020)

论文地址：http://www.vldb.org/pvldb/vol13/p307-sun.pdf

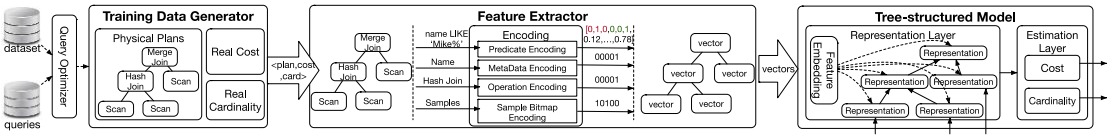
研究问题：

代价和基数估计是数据库查询优化的核心问题，他用于指导优化器选择最高效的执行计划。其中，基数指的是子查询计划得到的结果的行数，而代价指的是

子查询计划所消耗的 CPU 和 IO 时间。由于传统依赖于经验的基数和代价估计方法无法识别多个数据表之间的相关关系，所以他们无法提供高质量的多表连接基数（代价）估计。最近数据库社区越来越多的研究显示基于学习的基数估计和基于经验的方法相比，性能有显著提升。但是，已有的基于学习的方法依然存在一些局限性。比如，他们主要关注估计基数，而忽略了对于代价的估计，他们使用的模型要么无法表示复杂的查询计划，要么模型过于复杂而无法高效完成代价估计的任务。

研究方法：

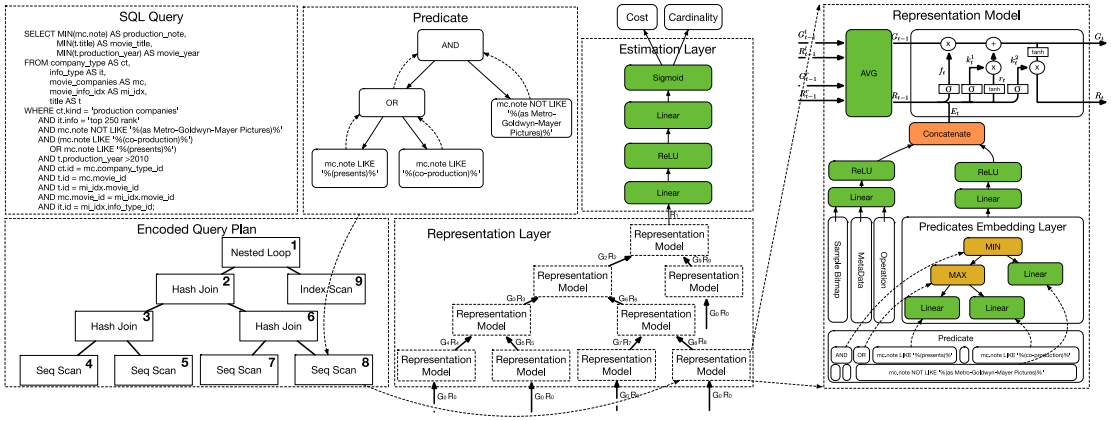
文章提出一个高效的基于树形递归神经网络的代价估计框架，这个框架支持同时估计查询基数和代价。对于估计算法，文章提出了包含查询以及具体物理执行操作的高效的特征抽取和编码技术。对于包含字符串模式匹配的查询条件，文章提出一个高效的基于词向量的编码方法以及对应高效词典抽取方法用于从数据库文本中抽取训练词向量所需要的字符子串。



如上图所示，文章提出的框架包含四个部分，1）训练数据生成器用于根据数据表连接图随机生成一系列的查询以及对应的真实基数和代价，输入模型进行训练。2）特征抽取模块从查询计划中抽取重要的特征并且编码成浮点向量作为输入。3）树形结构的模型是一个随着执行计划变化而变化的动态递归神经网络，他的每个节点用子节点的输出和对应执行计划节点的特征作为输入，输出作为上层节点的输入，依次循环直到根节点。4）表示记忆池用来记录已经计算过的子查询的表示，这样对于执行优化器选择计划的过程中对于同样的子查询不需要重复计算。

如下图所示，神经网络模型也分为三个层次，1）嵌入层将操作符，元数据，谓词和采样从系数矩阵编码成稠密矩阵。特别的，对于复合查询谓词，使用与或池化进行嵌入编码。2）表示层，这一层使用 LSTM 神经元组，接受孩子节点的输出表示，和本节点的特征进行线性非线性融合变换，输入当前的子查询表示。

3）输出层，将查询根节点的输出表示作为输入，通过两层全连接神经网络输出要估计的基数和代价。



Synthetic	median	90th	95th	99th	max	mean
PostgreSQL	1.69	9.57	23.9	465	373901	154
MySQL	2.07	22.6	50.6	625	458835	353
Oracle	1.97	12.4	40.1	473	545912	378
MSCN-NoSamp	2.14	6.72	11.5	114	1870	23.6
TLSTM-NoSamp	1.97	5.53	9.13	81.5	988	10.3
MSCN	1.19	3.32	6.84	30.51	1322	2.89
TNN	1.40	5.51	10.7	43.1	441	3.57
TLSTM	1.20	3.21	6.12	25.2	357	2.87
TPool	**1.18**	**3.19**	**6.05**	**24.5**	**323**	**2.81**

Scale	median	90th	95th	99th	max	mean
PostgreSQL	2.59	200	540	1816	233863	568
MySQL	3.08	90.1	329	7534	54527	426
Oracle	2.43	114	482	3412	102833	397
MSCN-NoSamp	2.33	96.1	257	1110	4013	131
TLSTM-NoSamp	2.06	69	176	931	3295	78.2
MSCN	**1.42**	37.4	140	793	3666	35.1
TNN	1.59	58.7	141	573	2238	31.3
TLSTM	1.43	38.8	139	469	1892	28.1
TPool	**1.42**	**37.3**	**125**	**345**	**1813**	**26.3**

JOB-light	median	90th	95th	99th	max	mean
PostgreSQL	7.93	164	1104	2912	3477	174
MySQL	9.55	303	685	2256	2578	149
Oracle	8.32	374	976	2761	3331	157
MSCN-NoSamp	5.43	126	978	1310	2020	100
TLSTM-NoSamp	5.18	97.3	613	864	1541	72.3
MSCN	3.82	78.4	362	927	1110	57.9
TNN	**2.95**	76.8	275	799	902	49.8
TLSTM	3.73	50.8	157	256	289	24.9
TPool	3.51	**48.6**	**139**	**244**	**272**	**24.3**

Synthetic	median	90th	95th	99th	max	mean
PostgreSQL	15.1	65.1	173	1200	8040	62.7
MySQL	4.51	39.7	94.7	449	7203	32.4
Oracle	6.72	41.1	124	796	6674	56.1
MSCN-NoSamp	10.3	24.7	234	569	2110	31.6
TLSTM-NoSamp	5.34	21.2	153	328	1345	19.8
MSCN	3.14	7.43	18.1	65.8	739	10.3
TNN	1.49	4.50	10.6	61.5	718	4.35
TLSTM	1.56	4.47	10.7	57.7	689	4.45
TLSTM-Multi	1.49	4.33	10.2	55.8	624	4.16
TPool	**1.48**	**4.12**	**10.1**	**47.6**	**532**	**3.99**

Scale	median	90th	95th	99th	max	mean
PostgreSQL	13.3	38.9	81.1	718	1473	35.7
MySQL	4.25	37.4	131	577	5157	40.7
Oracle	6.49	27.7	61.4	623	3612	31.5
MSCN-NoSamp	3.32	20.9	30.5	274	1173	21.2
TLSTM-NoSamp	2.19	13.4	21.7	228	1162	14.9
MSCN	1.79	10.6	27.1	88.8	1027	8.22
TNN	1.61	5.37	13.5	72.7	714	5.53
TLSTM	1.58	5.51	14.4	70.1	611	5.21
TLSTM-Multi	1.56	5.56	12.2	68.6	**254**	4.41
TPool	**1.54**	**5.29**	**11.9**	**67.6**	**254**	**4.39**

JOB-light	median	90th	95th	99th	max	mean
PostgreSQL	26.8	332	696	2740	3020	173
MySQL	9.47	102	342	1293	2228	84.5
Oracle	12.3	157	278	1366	1825	102.1
MSCN-NoSamp	12.4	152	231	1071	1553	62.7
TLSTM-NoSamp	10.4	103	217	986	1271	38.3
MSCN	4.75	11.3	40.1	563	987	27.4
TNN	2.06	25.5	134	293	401	19.1
TLSTM	3.66	32.1	80.3	445	583	17
TLSTM-Multi	**1.85**	13.2	22.9	**95**	**123**	5.81
TPool	**1.85**	**11.1**	**20.3**	101	125	**5.76**

Cardinality	median	90th	95th	99th	max	mean
PostgreSQL	184	8303	34204	106000	670000	10416
MySQL	104	28157	213471	1630689	2487611	60229
Oracle	119	55446	179106	697790	927648	34493
TLSTM-Hash	11.1	207	359	824	1371	83.3
TLSTM-Emb	11.6	181	339	777	1142	70.2
TLSTM-EmbRule	10.9	136	227	682	904	55.0
TPool	**10.1**	**74.7**	**193**	**679**	**798**	**47.5**

Cost	median	90th	95th	99th	max	mean
PostgreSQL	4.90	80.8	104	3577	4920	105
MySQL	7.94	691	1014	1568	1943	173
Oracle	6.63	149	246	630	1274	55.3
TLSTM-Hash	4.47	53.6	149	239	478	24.1
TLSTM-Emb	4.12	18.1	44.1	105	166	10.3
TLSTM-EmbRule	4.28	13.3	22.5	104	126	8.6
TPool	**4.07**	**11.6**	**17.5**	**63.1**	**67.3**	**7.06**

研究结果：

文章提出了完整的对于查询计划的基数和代价估计方法，如上图实验结果显示，对于 IMDB 数据集，无论是在 Synthetic, Scale, JOB-light 还是 JOB-full 查询上，该方法对于基数和代价的估计都要比传统方法和目前最新的基于深度学习的方法准确高效。

论文题目：*Software Engineering for Machine Learning: a Case Study*

中文题目：面向机器学习的软件工程：一个案例研究

论文作者：Amershi, Saleema, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann

论文出处：41st International Conference on Software Engineering (ICSE) ---- ICSE 2019

论文地址：https://dl.acm.org/citation.cfm?id=3339967

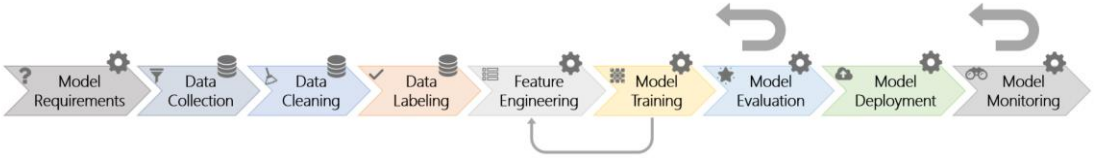
研究问题：

从个人计算、互联网、移动计算到云计算，每个转变都为软件工程带来了新的目标，这些目标促使研究新的开发模式，以解决软件工程在这些新领域遇到的问题。该论文首先概述了冲击软件行业的最新趋势，即在软件开发周期中集成人工智能（AI）的功能。该论文调研了一些使用机器学习来开发应用程序（例如 Bing Search 或 Cortana 虚拟助手）的 Microsoft 产品团队，和使用实时翻译文本，语音和视频的平台（例如 Microsoft Translator），以及为客户构建自己的机器学习应用程序 Azure AI 的平台。

研究方法：

在该论文中，通过案例分析，了解到了一些 Microsoft 软件团队是如何构建以客户为中心的 AI 功能的软件应用程序。为此，Microsoft 将现有的敏捷开发软件工程流程与 AI 专用工作流集成在一起，这些工作流得益于开发早期 AI 和数据科学应用程序的先前经验。在这项研究中，他们调研了 Microsoft 员工是如何应对 AI 软件开发以及大规模 AI 基础架构和应用程序开发所带来的日益增长的挑战。研究指出，由于整个公司的团队在 AI 方面拥有不同的工作经验，随着团队的成熟，新团队报告中出现许多问题的在这些成熟的团队中的重要性急剧下降，而某些问题对于大规模 AI 的实践仍然至关重要。这项研究尝试创建新的工作流程度量标准，以帮助开发团队在开发 AI 应用程序周期中做到进度管理。如下图所示，该论文调研了面向机器学习的软件工程的开发周期，这个开发周期通常包含了九个阶段，有些阶段是面向数据的（例如，收集，清洁和标记），而另一些

阶段是面向模型的（例如，模型需求，功能工程，培训，评估，部署和监视）。工作流中有许多反馈循环。较大的反馈箭头表示模型评估和监视可能会循环回到之前的任何阶段，例如在模型评估阶段，如果发现当前的模型效果不理想，有可能会回到数据清洗的阶段来进一步清洗数据以提高模型的质量。较小的反馈箭头说明模型训练可以循环回到特征工程（例如，在表示学习中）。



该论文在进行案例分析研究的过程中，主要有两个阶段收集数据：1、通过一组访谈来收集与研究问题相关的主要主题；2、以及对已收集的有意义的主题进行大规模调查。

Id	Role	Product Area	Manager?
I1	Applied Scientist	Search	Yes
I2	Applied Scientist	Search	Yes
I3	Architect	Conversation	Yes
I4	Engineering Manager	Vision	Yes
I5	General Manager	ML Tools	Yes
I6	Program Manager	ML Tools	Yes
I7	Program Manager	Productivity Tools	Yes
I8	Researcher	ML Tools	Yes
I9	Software Engineer	Speech	Yes
I10	Program Manager	AI Platform	No
I11	Program Manager	Community	No
I12	Scientist	Ads	No
I13	Software Engineer	Vision	No
I14	Software Engineer	Vision	No

1. Part 1
   1.1. Background and demographics:
      1.1.1. years of AI experience
      1.1.2. primary AI use case*
      1.1.3. team effectiveness rating
      1.1.4. source of AI components
   1.2. Challenges*
   1.3. Time spent on each of the nine workflow activities
   1.4. Time spent on cross-cutting activities
2. Part 2 (repeated for two activities where most time spent)
   2.1. Tools used*
   2.2. Effectiveness rating
   2.3. Maturity ratings
3. Part 3
   3.1. Dream tools*
   3.2. Best practices*
   3.3. General comments*

受访者的基本信息如上图左图所示，可以看出，受访者有从事一线软件开发工作的工程师、应用科学家、项目经理、架构师以及研究员，从事的领域包括了搜索、语音和视频等。采访的内容如上图右图所示，主要询问受访者的 AI 工作背景、目前工作中遇到的难题、开发周期中主要的工具和一些开发诉求和感想等。该论文研究者将调查问卷以内部邮件的形式发送给有关 AI 和 ML 相关的 4195 位员工。551 名员工填写了调查问卷，这项问卷的回应率为 13.6%。受访者分布在公司的所有部门中，并且来自各种职位：数据和应用科学（42%），软件工程（32%），程序管理（17%），研究（7%）和其他（1%）。21%的受访者是经理，这有助于该论文在访谈中平衡大多数经理的观点。

研究结果：

如下表所示，可以看出对于开发经验较少（Low）、中（Medium）和高（High）的受访者来说，最困难的就是数据的收集、清洗和管理工作。然而，对于 AI 领域的新知识的教育和训练，则认为没有那么重要。通过这项案例研究，我们可以看到数据管理在面向机器学习的软件工程周期中是十分重要的，也是十分具有挑战性的。

	Frequency			Rank		
				Experience		
Challenge	Medium vs. Low	High vs. Low	Trend	Low	Medium	High
Data Availability, Collection, Cleaning, and Management	-2%	60%		1	1	1
Education and Training	-69%	-78%		1	5	9
Hardware Resources	-32%	13%		3	8	6
End-to-end pipeline support	65%	41%		4	2	4
Collaboration and working culture	19%	69%		5	6	6
Specification	2%	50%		5	8	8
Integrating AI into larger systems	-49%	-62%		5	16	13
Education: Guidance and Mentoring	-83%	-81%		5	21	18
AI Tools	144%	193%		9	3	2
Scale	154%	210%		10	4	3
Model Evolution, Evaluation, and Deployment	137%	276%		15	6	4

论文题目：*Declarative Recursive Computation on an RDBMS, or, Why You Should Use a Database for Distributed Machine Learning*

中文题目：RDBMS 上的声明式递归计算，或者为什么要使用数据库进行分布式机器学习

论文作者：Jankov, Dimitrije, Shangyu Luo, Binhang Yuan, Zhuhua Cai, Jia Zou, Chris Jermaine, and Zekai J. Gao

论文出处：45th International Conference on VeryLarge Data Bases – VLDB 2019

论文地址：http://www.vldb.org/pvldb/vol12/p822-jankov.pdf

研究问题：

诸如 TensorFlow 之类的现代机器学习（ML）平台主要是为了支持数据并行性而设计的，其中在一组计算单元上并行执行一组几乎相同的计算（例如，梯度计算）。计算之间的唯一区别是，每个计算都针对不同的训练数据（称为“批次”）进行操作。对于异步数据并行性而言，在每次计算完成后，都将本地结果加载到参数服务器；对于同步数据并行性而言，在每次计算完成后，都将本地结果全局

聚合并用于更新模型。但问题是，数据并行性有其局限性。例如，数据并行性隐含地假设正在学习的模型（以及使用批处理更新模型时生成的中间数据）可以适合计算单元（可以是服务器或 GPU）的 RAM。但是，该假设有时候不成立。例如，目前最好的 NVIDIA Tesla V100 Tensor Core GPU 具有 32GB 的 RAM，然而这 32GB 的 RAM 无法存储全连接层的 200,000 个类别的条目的矢量编码。处理这样规模的模型需要并行化，在这种模型中，要学习的统计模型不是简单地在不同的计算单元上进行复制，而是要进行分区和并行操作，并由一系列批量同步操作执行。然而，现有的分布式机器学习系统为模型并行性提供的支持十分有限。该论文研究如何对现代关系数据库管理系统（RDBMS）进行少量更改，以使其适合于分布式学习计算。

研究方法：

该论文认为可以使用关系型数据库的相关技术来实现模型并行化。更加具体的，模型的不同部分可以存储在一组表中，部分模型上的计算通常可以通过一些 SQL 查询来表示。实际上，对于开发人员来说，学习算法的模型并行 SQL 与数据并行实现很相似。关系数据库管理系统（RDBMS）提供了声明性的编程接口，这意味着程序员或自动算法生成器，如果通过自动微分自动生成了 ML 算法，则只需指定他/她/它想要的内容，而无需指定需要写出如何计算它。计算将由系统自动生成，然后进行优化和执行以匹配数据大小，布局和计算硬件。无论计算是在本地计算机上还是在分布式环境中运行，代码都是相同的。相反，诸如 TensorFlow 之类的系统提供的声明性相对较弱，因为必须在某个物理计算单元（例如 GPU）上指定并执行计算图中的每个逻辑运算（例如矩阵乘法）。使用关系技术的另一个好处是，RDBMS 中的分布式计算已经研究了三十多年，并且速度快，功能强大。RDBMS 附带的查询优化器对于优化分布式计算非常有效。但是，有两个原因导致 RDBMS 无法作为大多数大型 ML 算法的平台直接使用。首先，RDBMS 缺乏足够的递归支持。在深度学习中，“遍历”深度神经网络的各个层，然后向后“遍历”整个网络以传播错误是必要的操作。虽然在 RDBMS 中，这种“循环”可以通过表之间的递归依赖关系来声明性地表示，但是 RDBMS 对递归的支持通常仅限于通过诸如传递闭包之类的集合来计算 fixed-points。另一个挑战是，典型的深度学习计算的查询计划可能会运行到成千上万的运算符，而现

有的 RDBMS 优化器将无法处理该运算符。简单来说，这篇文章向数据库表引入了多维的，类似数组的索引。当一组表共享相似的计算模式时，可以将它们压缩并替换为具有多个版本（由其索引指示）的表。此外，该论文还修改了数据库的查询优化器，以使其能够应对非常大的查询图（Query Graphs）。一个查询图被划分为一组可运行的 *frames*，优化了操作的代价，并将图划分问题（Grpah-cutting Problem）形式化为广义二次赋值（Generalized Quadratic Assignment）问题。

FFNN			
Hidden Layer Size	RDBMS (CPU)	RDBMS (GPU)	TensorFlow (GPU)
$3 per hour budget			
10000	04:50	06:25	00:24
20000	07:07	07:12	Fail
40000	11:52	11:48	Fail
80000	16:30	Fail	Fail
160000	Fail	Fail	Fail
$7 per hour budget			
10000	04:53	04:58	00:15
20000	05:54	06:08	Fail
40000	09:32	08:26	Fail
80000	12:03	17:50	Fail
160000	Fail	Fail	Fail
$15 per hour budget			
10000	05:12	5:00	00:12
20000	05:36	06:30	Fail
40000	09:08	08:39	Fail
80000	12:24	12:20	Fail
160000	39:40	Fail	Fail

Word2Vec		
Embedding Dimensions	RDBMS	TensorFlow
100	00:16:43 (00:01:59)	00:08:03
1000	00:17:05 (00:01:53)	01:14:58
10000	00:29:18 (00:01:53)	Fail

Collapsed LDA			
Number of Topics	RDBMS	TensorFlow	Spark
1000	00:06:25	00:05:06	00:00:39
5000	00:06:54	00:25:22	00:03:03
10000	00:07:05	00:52:35	00:06:39
50000	00:08:32	04:51:51	00:55:27
100000	00:09:58	Fail	01:42:35

研究结果：

该论文在 SimSQL 的基础上进行了实验，主要测试了三种机器学习算法的分布式实现：（1）多层前馈神经网络（FFNN）；（2）Word2Vec 算法；（3）分布式的 LDA 话题模型。如上图所示，在 FFNN 上，当隐层的大小较小的时候（10000），该论文提出的方法并不能超过 TensorFlow，但是当隐层的大小增大的时候，TensorFlow 并不能完成计算，而该论文提出的方法能进行有效地计算，同样的结论也可以在 Word2Vec 算法和 LDA 话题模型上得出。研究结果表明，该论文提出的方法能够处理需要在机器或计算单元之间分布的大型复杂模型。与 TensorFlow 相比，基于 RDBMS 的模型并行机器学习计算框架具有很好的伸缩性，对于 Word2Vec 和 LDA，基于 RDBMS 的计算可以比 TensorFlow 更快。但是，对于基于 GPU 的神经网络实现，RDBMS 的速度比 TensorFlow 慢。

11.5 数据库技术重要进展

步入大数据时代，面对 PB 乃至 EB 级海量数据、复杂多变的应用场景、异构的硬件架构和层次不齐的用户使用水平，传统的数据管理技术难以满足新时代的需求。例如，一个云数据库系统通常具有百万级别的数据库实例，每一个数据

库实例通常都有各自的应用场景、不同用户的使用水平往往也有着比较大的差别，数据库中传统的启发式算法在这些场景中难以取得较好的效果，而有经验的数据库管理员也难以直接干预和优化数量如此之多的数据库实例。

近年来，以机器学习为代表的人工智能技术因其强大的学习和适应能力，在多个领域都大放异彩。同样的，在数据管理领域，传统机器学习和深度学习等技术也有着巨大的潜力和广阔的应用前景。例如，数据库系统所积累的海量历史查询记录可以为基于学习的数据库智能优化技术提供数据支撑。一方面，我们可以构建包含查询、视图或数据库状态的有标签数据，比如，在视图选择问题中，这个标签是指每个候选视图是否被选中。另一方面，在缺乏标签数据的时候，我们可以利用（深度）强化学习技术探索性地（从选择结果的反馈中学习）选择最优的候选视图。此外，人工智能技术让自治数据库的自动决策管理、自动调优和自动组装等需求成为可能。在以深度学习为代表的人工智能技术的加持下，让数据库朝着更加智能的方向发展，数据管理技术也随之智能化。近些年涌现的自治数据库和人工智能原生数据库（如 SageDB，XuanyuanDB），通过融合人工智能技术到数据库系统的各个模块（优化器、执行器和存储引擎等）和数据管理的生命周期，可以大幅度提升数据库各方面的性能，为下一代数据库和人工智能技术的发展指明了一个方向。

在另外一方面，数据管理技术也能以基础设施的身份来支持人工智能的发展。目前的人工智能在落地过程中还面临着一些挑战性。例如，人工智能算法训练效率较低，现有人工智能系统缺少执行优化技术（如大规模缓存、数据分块分区、索引等），不仅会导致大量的计算、存储资源浪费，而且会提高程序异常的发生率（如内存溢出、进程阻塞等），严重影响单个任务的执行效率。其次，人工智能技术往往依赖高质量的训练数据，现实中的训练数据往往是包含很多缺失值、异常值和别名等类型的错误，这些错误通常会影响训练效率，对模型的质量造成干扰。面向人工智能的数据管理技术可以为解决上述挑战做出贡献。

12 可视化技术

12.1 可视化技术概念

可视化技术是把各种不同类型的数据转化为可视的表示形式，并获得对数据更深层次认识的过程。可视化将复杂的信息以图像的形式呈现出来，让这些信息更容易、快速地被人理解，因此，它也是一种放大人类感知的图形化表示方法。

可视化技术充分利用计算机图形学、图像处理、用户界面、人机交互等技术，以人们惯于接受的表格、图形、图像等形式，并辅以信息处理技术（例如：数据挖掘、机器学习等）将复杂的客观事物进行图形化展现，使其便于人们的记忆和理解。可视化为人类与计算机这两个信息处理系统之间提供了一个接口，对于信息的处理和表达方式有其独有的优势，其特点可总结为可视性、交互性和多维性。

目前，数据可视化针对不同的数据类型及研究方向，可以进一步划分为科学数据可视化、信息可视化，以及可视分析学三个子领域。这三个领域既紧密相关又分别专注于不同类型的数据及可视化问题。具体而言，科学可视化是针对科学数据的可视化展现技术。科学数据，例如，医疗过程中由 CT 扫描生成的影像数据、风洞实验而产生的流体数据、以及分子的化学结构等，是对物理世界的客观描述，往往是通过科学仪器而测量得到的数据。这类数据的可视化主要关注于如何以清晰直观的方式展现数据所刻画的真实物理状态。因此，科学可视化往往呈现的是三维场景下的时空信息。信息可视化注重于如何以图形的方式直观展现抽象数据，它涉及到了对人类图形认知系统的研究。在这里，抽象数据（例如：图形数据、多维度数据、文本数据等）往往是对各应用领域所产生数据的高层次概括，记录的是抽象化的信息。针对这样的数据，信息可视化着眼于多维度信息的可视编码技术，即如何以低维度（2D）的图形符号来直观展现并揭示抽象数据中所隐藏的潜在规律与模式；可视分析学是多领域技术结合的产物，旨在结合并利用信息可视化、人机交互、以及数据挖掘领域的相关技术，将人的判断与反馈作为数据分析中重要的一环，从而达到精准数据分析、推理及判断的目的。

可视化技术的重要性在于，通过提供对数据和知识的展现，建立用户与数据系统交互的良好沟通渠道，利用人类对图形信息与生俱来的模式识别能力，通过

以直观的图像化方式展现数据，从而帮助用户快速发觉数据中的潜在规律，并借助分析人员的领域知识与经验，对模式进行精准分析、判断、推理，从而达到辅助决策的目的。

目前可视化技术在各行各业中均得到了广泛的应用。其中，可视化技术在信息安全、智慧医疗、电子商务、机器学习、智慧城市、文化体育、数字新闻、气象预报、地质勘测等诸多领域产生了非常广泛的应用，并逐渐成为这些领域当中越来越重要的组成部分。

当下可视化存在的挑战是：如何进一步深入挖掘人类对于图形、动画、以及交互的感知及认知模式，从而进一步完善可视化的相关理论；如何打破“手工作坊”式的针对每一个问题，单独定制数据可视化设计方案的传统模式，大规模批量创造生成风格化的可视展现；以及，如何根据用户的数据分析任务与需求自动推荐合适的可视化展现方式。

主要的研究趋势：海量、异构、时变、多维数据的可视化展示方案；可视化在可解释性深度学习领域的应用；自动可视化生成技术的研究；基于形式概念分析理论的知识可视化方法；可视化模式识别；整体可视与局部详细可视相结合的新方法研究等。

12.2 可视化技术发展历史

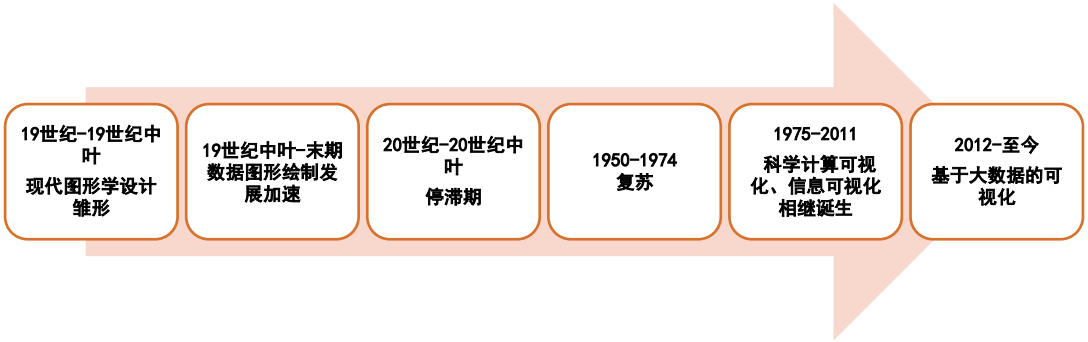


图 12-1 可视化发展历程

19 世纪-19 世纪中叶：现代图形学设计雏形

十九世纪前叶，因为受视觉表达方法创新的影响，统计图形及专题绘图领域应用得到快速发展。目前，我们看到的绝大多数统计图形都是在这一时间段被发

明的。同期，因政府开始着重关注人口、教育、犯罪、疾病等领域，数据的收集整理范围明显扩大，超乎以往的社会管理方面的数据被收集起来用于社会分析。1801 年英国地质学家 William Smith 绘制了第一幅地质图，引领了一场在地图上表现量化信息的潮流。这一时期，数据的收集整理从科学技术和经济领域扩展到社会管理领域，对社会公共领域数据的收集标志着人们开始以科学手段进行社会研究。与此同时科学研究对数据的需求也变得更加精确，研究数据的范围也有明显扩大，人们开始有意识地使用可视化方式来尝试研究、解决更广泛领域的问题。

19 世纪中叶-末期：数据图形绘制发展加速

在十九世纪中叶，统计图形、概念图等概念迅猛发展，此时的人们已经掌握了整套统计数据可视化工具，数据可视化领域发展进入了加速期，随着数字信息对社会、工业、商业直至交通规划的影响不断增大，欧洲开始着力发展数据分析技术。一群学者发起的统计理论给出了多种数据的意义，数据可视化迎来了它历史上的第一个发展加速期。统计学理论的建立是可视化发展的重要一步，此时数据由政府机构进行收集，数据的来源变得更加规范化。随着社会统计学的影响力越来越大，在 1857 年维也纳的统计学国际会议上，学者就已经开始对可视化图形的分类和标准化进行讨论。不同数据图形开始出现在书籍、报刊、研究报告和政府报告等正式场合之中。这一时期法国工程师 Charles Joseph Minard 绘制了多幅有意义的可视化作品，被称为“法国的 Playfair”，他最著名的作品是用二维的表达方式，展现六种类型的数据，用于描述拿破仑战争时期军队损失的统计图，如下图所示。并且在这一时期出现了三维的数据表达方式，这种创造性的成果对后来的研究有十分突出的作用。

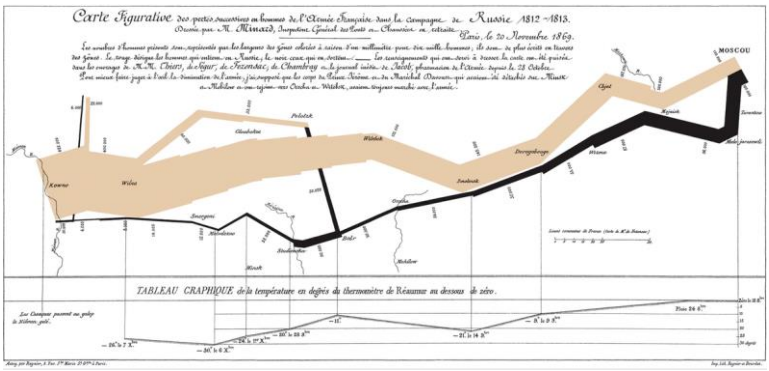


图 12-2 拿破仑进军莫斯科的历史事件

20 世纪-20 世纪中叶：停滞期

20 世纪的上半叶，随着数理统计这一新数学分支的诞生，追求数理统计严格的数学基础并扩展统计的疆域成为这个时期统计学家们的核心任务。数据可视化成果在这一时期得到了推广和普及，并开始被用于尝试着解决天文学、物理学、生物学的理论新成果，Hertzsprung-Russell 绘制的温度与恒星亮度图成为了近代天体物理学的奠基之一；伦敦地铁线路图的绘制形式如今依旧在沿用（如下图所示）；E.W.Maunder 的“蝴蝶图”用于研究太阳黑子随时间的变化。然而，这一时期人类收集、展现数据的方式并没有得到根本上的创新，统计学在这一时期也没有大的发展，所以整个上半叶都是休眠期。但这一时期的蛰伏与统计学者潜心的研究才让数据可视化在本世纪后期迎来了复苏与更快速的发展。



图 12-3 HenryBeck 设计的伦敦地铁图

1950-1974：复苏

从 20 世纪上半叶末到 1974 年这一时期被称为数据可视化领域的复苏期，在这一时期引起变革的最重要的因素就是计算机的发明，计算机的出现让人类处理数据的能力有了跨越式的提升。在现代统计学与计算机计算能力的共同推动下，数据可视化开始复苏。随着计算机的普及，上世纪六十年代末，各研究机构逐渐开始使用计算机程序取代手绘的图形。由于计算机的数据处理精度和速度具有强大的优势，高精度分析图形已不能用手绘制。在这一时期，数据缩减图、多维标度法 MDS、聚类图、树形图等更为新颖复杂的数据可视化形式开始出现。人们尝试着在一张图上表达多种类型数据，或用新的形式表现数据之间的复杂关联，

这也成为这一时期数据处理应用的主流方向。数据和计算机的结合让数据可视化迎来了新的发展阶段。

1975-2011：科学计算可视化、信息可视化相继诞生

这段时期，计算机成为数据处理的一个重要工具，数据可视化进入了新的黄金时代，随着应用领域的增加和数据规模的扩大，更多新的数据可视化需求逐渐出现。二十世纪七十年代到八十年代，人们主要尝试使用多维定量数据的静态图来表现静态数据，八十年代中期出现了动态统计图，最终在上世纪末两种方式开始合并，致力于实现动态、可交互的数据可视化，动态交互式的数据可视化方式成为新的发展主题。数据可视化的这一时期的最大潜力来自动态图形方法的发展，允许对图形对象和相关统计特性的即时和直接的操纵。这一时段初期就已经出现交互系统，通过调整控制来选择参考分布的形状参数和功率变换。这可以看作动态交互式可视化发展的起源，并推动了这一时期数据可视化的发展。

2012-至今：基于大数据的可视化

步入 21 世纪互联网数据量猛增，人们逐渐开始对大数据的处理进行了重点关注。之后全球每天的新增数据量就已经开始以指数倍膨胀，用户对于数据的使用效率也在日益提升，数据的服务商开始需要从多个维度向用户提供服务，大数据时代就此正式开启。2012 年，我们进入数据驱动的时代。人们对数据可视化技术的依赖程度也不断加深。大数据时代的到来对数据可视化的发展有着冲击性的影响，继续以传统展现形式来表达庞大的数据量中的信息是不可能的，大规模的动态化数据要依靠更有效的处理算法和表达形式才能够传达出有价值的信息，因此大数据可视化的研究成为新的时代命题。我们在应对大数据时，不但要考虑快速增加的数据量，还需要考虑到数据类型的变化，这种数据扩展性的问题需要更深入的研究才能解决；互联网的加入增加了数据更新的频率和获取的渠道，而实时数据的巨大价值只有通过有效的可视化处理才可以体现，于是在上一历史时期就受到关注的动态交互的技术已经向交互式实时数据可视化发展。综上，如何建立一种有效的、可交互式的大数据可视化方案来表达大规模、不同类型的实时数据，成为了数据可视化这一学科的主要的研究方向[66-67]。

12.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为可视化领域全球学者分布情况：



图 12-4 可视化领域全球学者分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显；欧洲也有较多的人才分布；亚洲的人才主要集中在我国东部地区；其他诸如非洲、南美洲等地区的学者非常稀少；可视化领域的人才分布与各地区的科技、经济实力情况大体一致。此外，在性别比例方面，可视化中男性学者占比 91.7%，女性学者占比 8.3%，男性学者占比远高于女性学者。

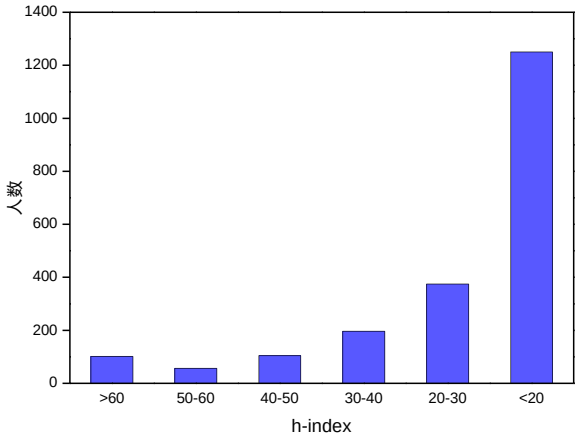


图 12-5 可视化领域学者 h-index 分布

可视化学者的 h-index 分布如上图所示，分布情况大体呈阶梯状，其中 h-index 小于 20 区间的人数最多，有 1250 人，占比 60.1%，50-60 的区间人数最少，有 56 人。

- 中国人才分布

我国专家学者在可视化领域的分布如下图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩等地相比，中国在可视化领域学者数量较多但差距较小。



图 12-6 可视化领域中国学者分布

中国与其他国家在可视化的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。从表中数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在可视化领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 5 席；中国与瑞士、中国与瑞典的平均引用数都达到了最高，说明在合作质量上中瑞合作达到了较高的水平。

表 12-1 可视化领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	152	4996	33	311
中国-新加坡	34	684	20	64
中国-英国	32	1393	44	78
中国-瑞士	31	2298	74	40
中国-瑞典	31	2298	74	40
中国-德国	23	422	18	49
中国-加拿大	19	659	35	37
中国-法国	7	83	12	17
中国-澳大利亚	2	71	36	7
中国-日本	2	58	29	7

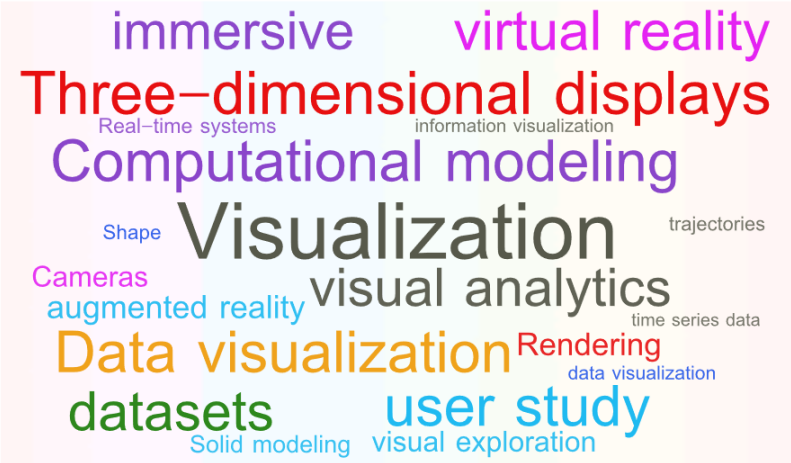
12.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

IEEE Transactions on Visualization & Computer Graphics

IEEE Visualization Conference

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，可视化（Visualization）、三维显示（Three-dimensional displays）、计算建模（Computational modeling）是本领域中最热的关键词。



论文题目：*Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips*

中文题目：城市大时空数据的可视化研究:纽约市出租车出行研究

论文作者：Nivan Ferreira, Jorge Poco, Huy T. Vo, Juliana Freire, Cláudio T. Silva

论文出处：IEEE Transactions on Visualization and Computer Graphics, 2013

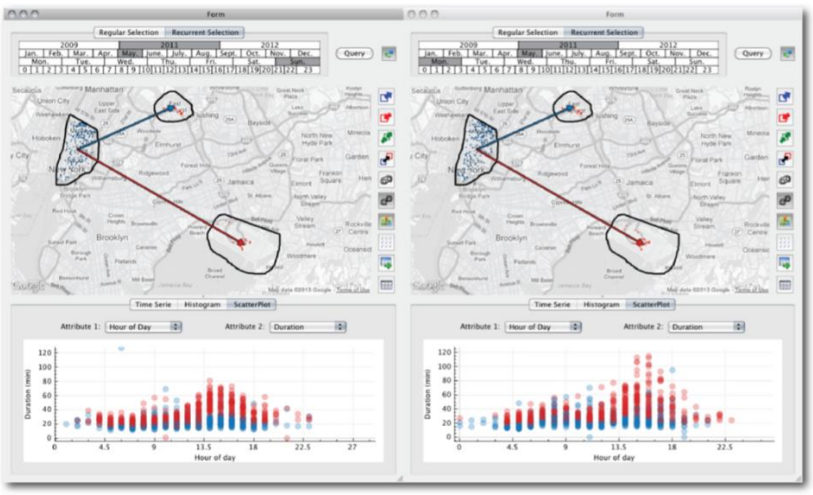
论文地址：

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6634127&tag=1

研究问题：

出租车数据是城市中极具价值的信息，收集并利用好出租车的数据可以有效的帮助决策者和社会学家理解城市的状况并做出正确的决策。但高效的探索出租车数据其实是一个充满挑战的事情。出租车数据十分复杂且庞大，包含了时间和空间上的信息，很难快速查询并进行比较。在采访城市规划和交通专家后，该文作者了解到，他们目前没有合适的工具来完成分析。一些简单的工具和语言只能分析一些小规模的数据，能分析的数据比较片面；而复杂一些的工具，虽然可以对大数据进行分析，则需要掌握高级的数据查询语言，对分析人员而言很困难。所以该文提出了一种支持在起点-终点（OD）数据上进行复杂时空可视化查询的模型。

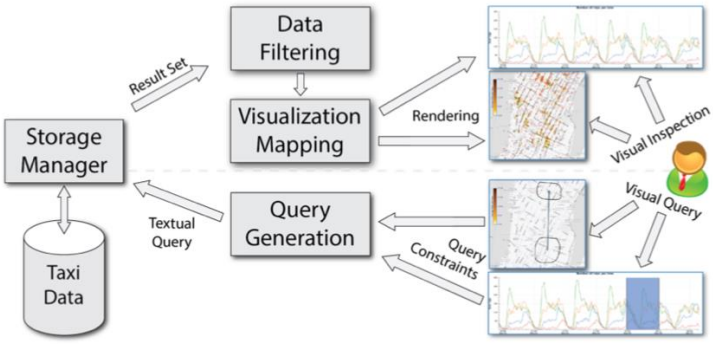
研究方法：



该文提出了一个可视化系统 TaxiVis，对纽约市的出租车数据进行可视化，让用户更便捷的进行可视查询。系统的主要界面如上图所示，主要包括三个部分，最上面的是时间选择器，中间的是可以用来查询和展示结果的地图，位于下方的是数据信息总结的一个视图，可以将查询到的结果通过不同的形式进行展示，例如，时间序列，直方图，不同属性的散点图，支持可视化的过滤。

系统的主要流程是，用户通过和地图和其他的一些视觉元素进行交互，直观地进行查询。系统接收到可视化的查询后，会在内部自动生成文本查询。关于数据查询的设计，作者对城市规划和交通专家进行了采访，将查询需求归纳成下面三点：时间、地点和对象。对应的支持三种查询：时间+地点->对象，时间+对象->地点，地点+对象->时间。作者提出了一个可视模型，SELECT * FROM trips WHERE <constraints>。Constrains 就是上述三种类别，时间，空间，以及对象属性上的限制。时间限制通过时间选择器实现；空间限制可以在地图上进行区域圈选；而属性限制可以通过在属性直方图上筛选完成。

之后，在存储管理器查询获得结果，为支持时空数据的交互式查询，作者采用一种高效的数据存储模式，构建了一个基于 k-d 树的特殊索引。有了查询结果后，系统会立刻将结果显示在地图上。用户可以依照这个查询结果不断的来优化查询。由于查询到的结果可能会很大，因此作者还利用了自适应的细节层次（LOD）和密度热力图来可视化地展示结果。具体的系统结构如下所示：



研究结果：

该文构建了一个可视化系统 TaxiVis，系统支持多种交互方式，帮助用户对数据的所有维度进行查询，更加便捷的探索与出租车行程相关的属性。作者通过一系列案例研究验证了系统的有效性，如比较了不同区域上下客流量随时间的变

化，交通枢纽地区的交通流量变化，以及飓风 Sandy 和 Irene 对纽约市出租车系统的影响。案例表示了作者提出的系统确实可以帮助行业专家完成之间无法实现的研究任务，通过作者设计的可视化系统，专家们能更高效的挖掘出更多的有效信息。

论文题目：*UpSet: Visualization of Intersecting Sets*

中文题目：UpSet: 相交集的可视化

论文作者：Alexander Lex, Nils Gehlenborg, Hendrik Strobelt, Romain Vuillemot, and Hanspeter Pfister

论文出处：IEEE Transactions on Visualization and Computer Graphics（Volume: 20, Issue: 12, Dec. 31 2014）

论文地址：http://sci.utah.edu/~vdl/papers/2014_infovis_upset.pdf

研究问题：

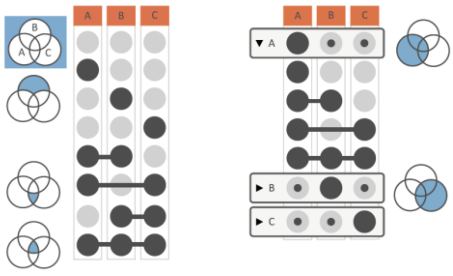
了解集合之间的关系是一项重要的数据分析任务。如果集合的数量超过一定的阈值，集合交叉点的数量就会爆炸性增长。所以对于数量多的集合数据来说，对其进行分析和可视化是具有挑战性的。为了解决这个问题，该文提出了 UpSet—一种新颖的可视化技术，用于对集合及其交集进行定量分析。

研究方法：

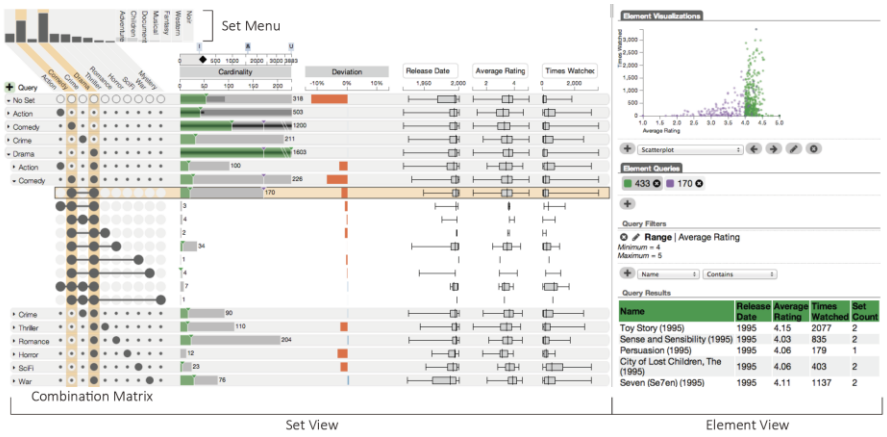
UpSet 专注于创建任务驱动的聚合，研究集合之间在大小，所包含元素及其相关属性方面的相互关系。UpSet 引入矩阵布局，可有效表示关联数据，例如交集中元素的数量，以及从子集或元素属性派生的其他摘要统计信息。根据各种度量进行分类，可以对相关的交叉集合进行任务驱动的分析。集合中表示的元素及其关联的属性可以在单独的视图中可视化。

UpSet 采用“分而治之”的概念方法，将 k 个集合的数据集划分为所有可能的 2^k 个交叉点。这些交点对应于维恩图的原子区域。将这些基本构建块被称为互斥交集。使用专有交集作为基本构建模块，使分析人员能够创建聚合。聚合是使用任务驱动方法定义的互斥交集的集合。交叉集可以根据某种聚合语义进行聚合，也可以通过查询进行聚合。

UpSet 将集合的交集绘制为矩阵，如下图所示。每一列对应于一组，并且每一行对应于维恩图中的一个分段。若单元格为空（浅灰色圆圈），表示此集合不是该交集的一部分；若单元格为已填充（黑色圆圈），表示该集合是交集的一部分。带有黑点的浅灰色圆圈表示这些集合可能是互斥交集的一部分。如下图（a）中的第一行完全为空，说明它不是交集的一部分，第二行对应于仅在集合 A 中的元素（且不在 B 或 C 中）。



下图为使用 UpSet 对电影数据集进行关于集合分析的示例，该图清楚的显示了电影流派的关系。设置视图显示了集合间的相交关系及其聚合，元素数量和属性统计信息。元素视图显示了过滤后的元素和一个散点图，比较了两组过滤后的元素。



研究结果：

该文提出一种关于集合的可视化技术 UpSet 并进行了用例研究，通过采访来自各个领域（宏观经济学、遗传学、药理学和社会网络分析）的多位研究人员，总结他们在研究中遇到的各类与集合相关的分析任务，并且让每位研究人员都利用 UpSet 都对他们提供的数据集进行了分析。最终实验结果表明，UpSet 可以有

效解决几位研究人员所描述的 26 个与集合相关的任务中的 23 个。由研究结果以及专家评论可以得出，UpSet 这一技术对于分析集合数据来说是一种很有用的工具，同时在不同领域的多个用例研究结果展示了其通用性。

论文题目：*Towards Better Analysis of Deep Convolutional Neural Networks*

中文题目：针对深度卷积神经网络的进一步分析

论文作者：Mengchen Liu, Jiaxin Shi, Zhen Li, Chongxuan Li, Jun Zhu, Shixia Liu

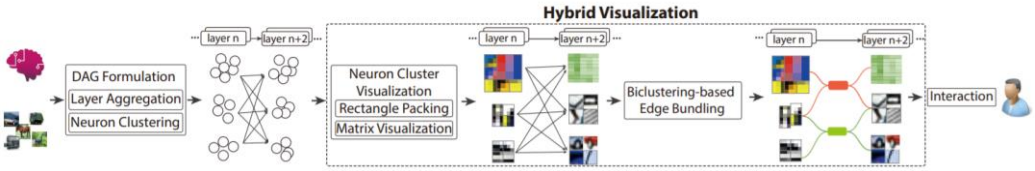
论文出处：IEEE transactions on visualization and computer graphics, 2016

论文地址：https://arxiv.org/pdf/1604.07043.pdf

研究问题：

深卷积神经网络（CNNs）在图像分类等模式识别任务中取得了突破性的性能，然而开发高质量的深度学习模型通常依赖于大量的尝试和错误。因为对于深度学习模型如何学习以及为什么有效仍然没有明确的解释。为了帮助专家们更好地探索和深入理解 CNN，并且分析每个神经元的作用和神经元之间的联系，该文提出了一个新的交互式可视化系统 CNNVis，帮助专家更好地理解、诊断和提炼深度 CNN。

研究方法：



CNNVis 是一个交互式的可视化分析系统，旨在帮助机器学习专家更好地理解、诊断和完善 CNN 模型。基于深度 CNN 的特点，CNNVis 将其表示为有向无环图（DAG），其中每个节点表示了一个神经元，每条边代表一对神经元之间的连接。若需要可视化大型 CNN，CNNVis 首先对网络中的层进行聚类，然后从每个层聚类中选择一个具有代表性的层，然后 CNNVis 在每个代表层中对神经元进行聚类，并从每个神经元聚类中选择几个代表性的神经元。在 DAG 表示的基础上，CNNVis 开发了一种混合可视化方法，通过显示不同类型图像中神经元的作

用来揭示神经元与每个神经元之间的多种相互作用关系。需要特别指出的是，CNNVis 提出了一种分层矩形排列算法来显示神经元簇的派生特征。在 hold-Karp 算法（state compression dynamic programming，状态压缩动态规划）的基础上，CNNVis 设计了一种矩阵重排序算法，演示了每个神经元素激活时的簇模式。这里，激活值是神经元的输出值。这个值由将神经元输入值转换为神经元输出值的激活函数决定。此外，CNNVis 还提出了一种双聚类边缘捆绑方法来减少神经元之间大量连线所造成的视觉混乱。

研究结果：

通过三个案例研究，该文评估了 CNNVis 的高效性与有用性。三个案例研究包括：利用 CNNVIs 分析模型结构对于模型表现的影响，利用 CNNVis 诊断一个失败的模型训练过程和利用 CNNVis 提升模型的表现。研究结果证实 CNNVis 可以帮助专家诊断模型结构的潜在问题，并对 CNN 进行优化，从而使模型构建过程的迭代速度更快，收敛速度更快。

论文题目：*Visualizing Dataflow Graphs of Deep Learning Models in TensorFlow*

中文题目：在 TensorFlow 中可视化深度学习模型的数据流图

论文作者：Kanit Wongsuphasawat, Daniel Smilkov, James Wexler, Jimbo Wilson, Dandelion Mané, Doug Fritz, Dilip Krishnan, Fernanda B. Viégas, and Martin Wattenberg

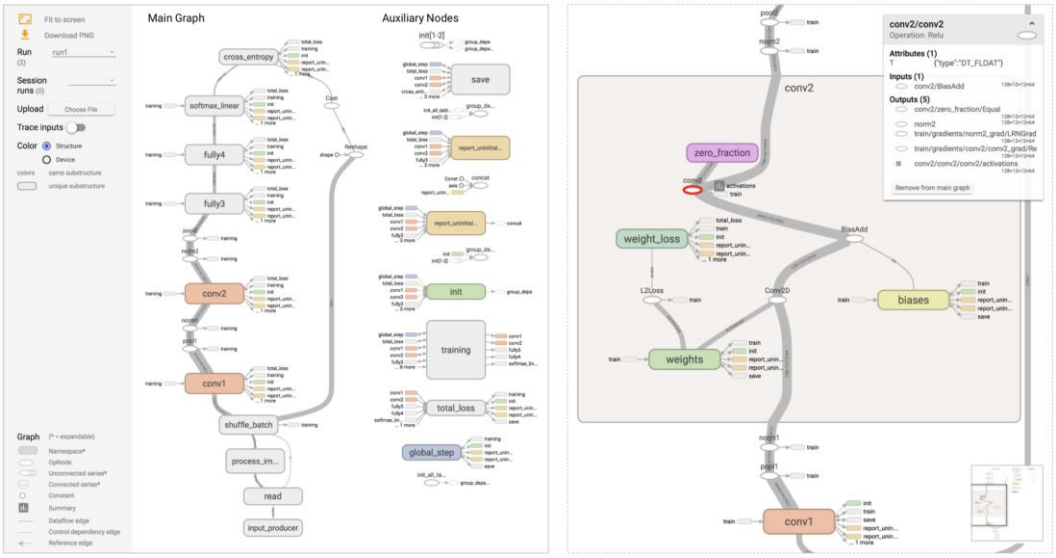
论文出处：IEEE Transactions on Visualization and Computer Graphics（Volume: 24, Issue: 1, Jan. 2018 ）

论文地址：https://idl.cs.washington.edu/files/2018-TensorFlowGraph-VAST.pdf

研究问题：

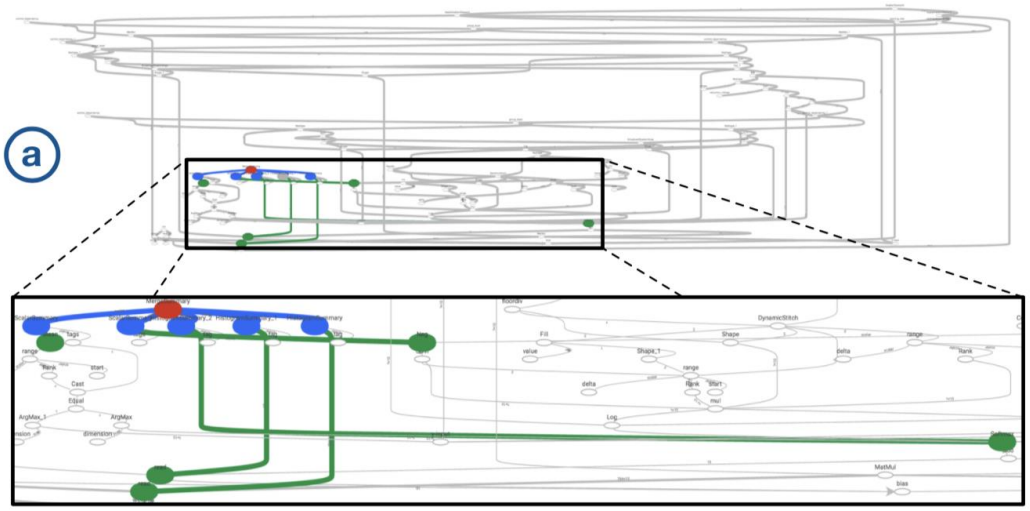
近年来，机器学习取得了一系列突破，其中，深度学习最为突出。深度学习的最大特点是其多层次的计算网络，而这些网络的复杂性导致其实现上的困难，所以 TensorFlow、Theano 和 PyTourch 等深度学习平台提供了高级的 API 来降低这些困难，开发人员可以利用这些 API 生成数据流图，以支持各种算法模型和分布式计算。然而，深度学习的网络结构通常都很复杂，开发人员凭借着自己的记

忆或代码本身，很难对算法各个部分进行调试或者互相沟通。因此，该文提出了一种基于 TensorFlow 的可视化工具 TensorFlow Graph Visualizer，旨在帮助开发者在 TensorFlow 中进行算法的分析与开发。



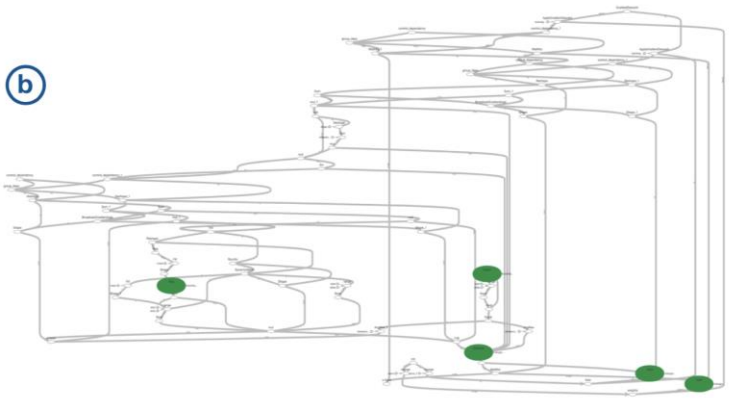
研究方法：

如下图所示，该文首先直接利用 Sugiyama 流布局对 TensorFlow 的数据流图进行布局。由于数据流图里的结构复杂，这种布局容易出现交错，并且缺乏层次分布。

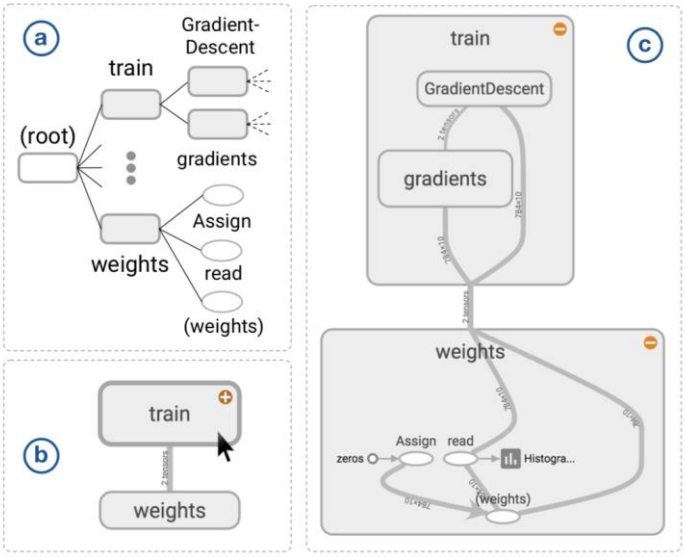


根据这种布局方式的缺点，文章提出了三点优化：

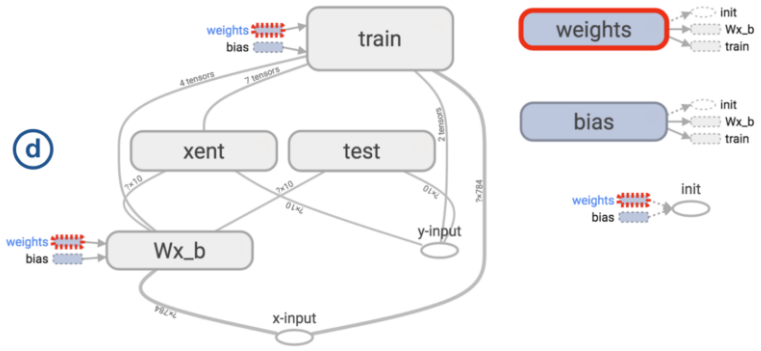
提取非重要节点。该文作者将常量节点和统计节点提取出来作为非重要节点。其原因在于，常量节点通常是某个操作的输入，只有出度，没有入度；而统计节点需要利用到负责日志记录的“中介节点”，对实际的数据流操作没有影响且容易破坏布局。将这些节点提取出来，使用小的图标进行展示，对布局效果有很大改善（见下图）。



建立层次关系。数据流本身是存在层次结构的，多种简单操作可能组合为一个复杂操作。因此，该文作者利用各个操作的命名空间这种层级关系，对布局进行优化：只显示命名空间，将底层的操作隐藏，并进行边绑定，直到用户对负节点进行点击操作才展开底层的数据操作。圆角矩形代表一组操作，节点的高度代表其包含的操作数目。对于内部结构相同的节点使用相同的颜色着色。节点间的连边根据关系有所不同：实线代表数据传输，虚线代表存在算法依赖关系，而连边的粗细则代表数据流的大小。



提取辅助节点。对于当前布局，图中仍存在许多度数高的节点，造成边缠绕，降低了图的可读性。同时，深度学习专家反馈目前图上存在着一些分析过程中不被重视的节点，比如常量或日记记录。因此，作者将这些点提取出来作为辅助节点，将其画在数据流图之外，然后在数据流图中利用小的代理图标来表示辅助节点与其他节点的相连关系，简化图的结构（见下图）。



随后，该文作者通过对数据流图结构进行分析，从组节点中提取模板和检查子图间的相似性，并且加入额外的定量数据，来帮助算法开发人员理解和分析算法。

研究结果：

该文利用经过优化后的 TensorFlow Graph Visualizer 分别展示卷积神经网络和 Inception 网络架构，展示了 TensorFlow Graph Visualizer 的应用场景，并通过用户反馈证明了 TensorFlow Graph Visualizer 对于理解、调试和共享模型的结构的有效性。

论文题目：*ACTIVIS: Visual Exploration of Industry-Scale Deep Neural Network Models*

中文题目：ACTIVIS：工业规模深层神经网络模型的可视化探索

论文作者：Minsuk Kahng, Pierre Y. Andrews, Aditya Kalro, and Duen Horng （Polo） Chau.

论文出处：IEEE Transactions on Visualization and Computer Graphics（Volume: 24, Issue: 1, Jan. 2018 ）

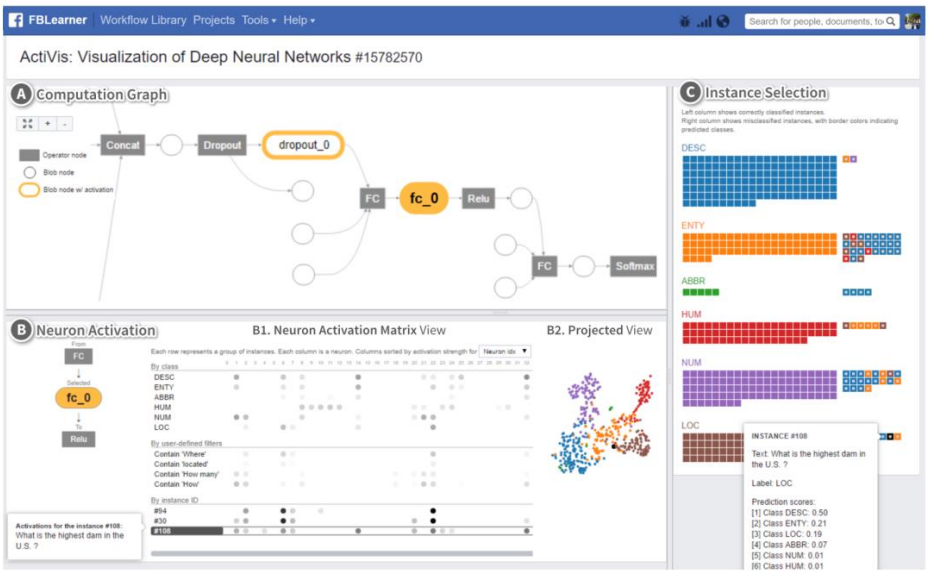
论文地址：https://arxiv.org/pdf/1704.01942.pdf

研究问题：

深度学习是机器学习的子域，它应用非监督式或监督式的特征学习和分层特征提取的方法来代替传统的手工特征提取，是一种十分高效的算法。虽然深度学习模型已经在许多预测任务中实现了一定的精准度，但是理解这些模型仍然是一个挑战。近年来，研究人员尝试着开发可视化工具来帮助人们理解深度学习模型。然而，由于工业界模型的复杂性和多样性，以及他们所使用的大规模数据集，导致现有的工作不能充分解决这些问题。因此提出了一个用于解释大规模深度学习模型和结果的交互式可视化系统（ACTIVIS）。它集成了包括模型架构视图、神经元激活矩阵视图以及实例选择视图在内的多个子视图。

研究方法：

通过与深度学习领域从业者（Facebook 员工）进行访谈，凝练出三个主要的设计目标：比较不同实例下神经元激活情况；对模型架构和神经元活动情况进行低级别检查；支持工业界模型和数据集的导入分析。根据这些设计目标，对系统做出了如下设计和开发：



如上图所示，ACTIVIS 系统主要由三个部分组成，包括模型架构视图、神经元激活矩阵视图和投影视图。模型架构视图：对整个深度学习网络结构的概览，矩形节点代表操作函数，圆形节点代表一个张量；神经元激活矩阵视图：左图中的每一列代表一个实例下神经元的激活情况，用圆形的颜色深浅来编码其激活情况，颜色越深表示激活值越大；右图为投影视图，展示实例的分类情况；实例选

择视图：每一行代表一个实例的分类情况，左半部分代表被正确分类的实例，右半部分代表被错误分类的实例，右半部分矩形颜色代表其被错误分类的类别。

研究结果：

用户在使用该系统进行神经网络模型探索时，通过概览和探索整个深度神经网络的结构，激活矩阵视图和投影视图观察分类情况，以及实例选择图中的具体情况对模型进行调整，可以在较大范围内提高开发效率。

论文题目：*FlowSense: A Natural Language Interface for Visual Data Exploration within a Dataflow System*

中文题目：FlowSense：用于可视数据的自然语言界面数据流系统内的探索

论文作者：Bowen Yu, Claudio Silva.

论文出处：EEE Transactions on Visualization and Computer Graphics

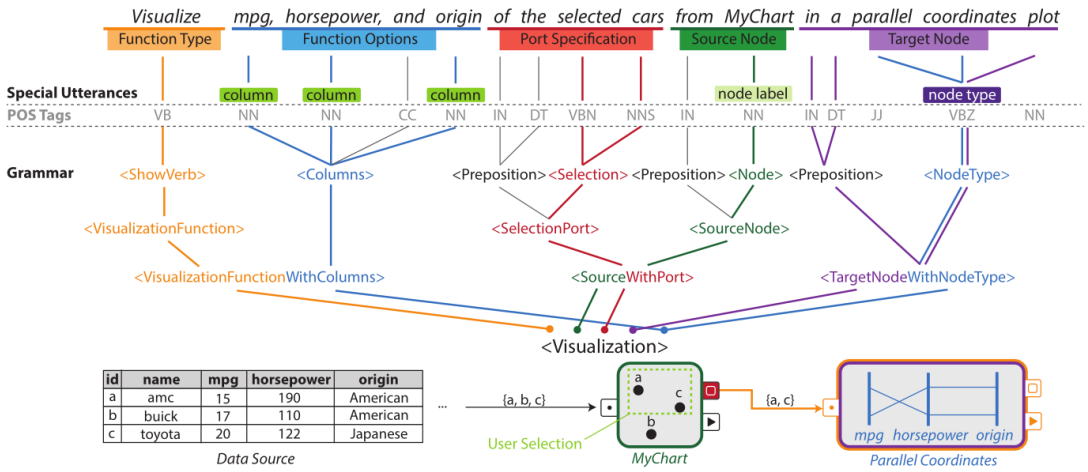
论文地址：https://arxiv.org/pdf/1908.00681.pdf

研究问题：

数据流图（DFD）是描述系统中数据流动过程的分析工具，它以图形元素和流动指向来构建数据变化的内部逻辑。与传统的流程图或框图不同，它是从数据的角度描述一个系统。这也就意味着它更抽象，不便于理解和学习。另一方面，自然语言作为人类行为学认知的基本单元，往往是最便于理解的。因此，该文提出了使用一个带有特殊文本标记和占位符的语义分析器（FlowSense）来概括不同的数据集和数据流图。FlowSense 试图将自然语言作为系统的切入口，经过自然语言处理转化为计算机语言，辅助用户完成数据流图的构建。

研究方法：

首先通过研究示例数据流图集构造 VisFlow 函数，该函数支持数据流图的自动生成。FlowSense 从用户输入的自然语言中提取一组特殊的标记，该标记是指表示为数据集或数据流图中的实体的单词（包括表列名、节点标签、节点类型和数据集名称等），它们可作为 VisFlow 函数的参数和操作数。于是，FlowSense 再通过执行 VisFlow 函数生成数据流图。具体实现流程如下图所示：



用户输入一段文字。系统会捕获到其中的特殊文本，然后利用自然语言处理技术将其转化为 VisFlow 函数可以识别的语法，即该函数的参数或操作数，最后绘制数据流图。

当用户展开数据流关系图时，总是存在一个编辑焦点，尽管该焦点通常是隐式的。FlowSense 根据用户的交互行为判断焦点，通过以下方法计算每个节点 X 的焦点得分：

$$score(X) = activeness(X,t) + \alpha(1 - \frac{1}{1+e^{-(distanceToMouse(X)/\gamma-\beta)}})$$

X 的活跃度在每次用户在系统中点击时重新迭代：

$$activeness(X,t) = activeness(X,t-1)/2 + click(X,t)$$

其中，$click(X,t)=1$ 代表用户在 t 时刻对 X 节点进行了点击操作。FlowSense 选择焦点得分最高的节点作为图表编辑焦点。如果需要多个源节点（例如，在合并查询中），那么 FlowSense 将按节点的焦点得分递减顺序选择节点。

研究结果：

该文结合 VisFlow 框架进行了用户研究来评估 FlowSense 系统的有效性。结果表明，用户通过查询得到有效数据流图的概率为 68.455%。后续研究通过研究人员对系统的改进，将有效率提高到 76.911%。这说明系统的整体运行情况良好。

论文题目：*Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco*

中文题目：以形式化可视化设计知识为约束：Draco 系统中可操作和扩展的模型

论文作者：Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith, Bill Howe, Jeffrey Heer

论文出处：IEEE Transactions on Visualization and Computer Graphics（Volume: 25, Issue: 1, Jan. 2019 ）

论文地址：https://adamsmith.as/papers/08440847.pdf

研究问题：

在可视化设计中，程序员往往根据经验对可视化图表进行设计、编写，他们无法正确判断自己编写的图表是否符合设计标准，导致最后的效果不尽如人意。虽然经验研究可以提供设计指导，但人们缺少一个系统性的判断体系和框架来表示设计标准、整合研究结果，从而促进有效的可视化编码和可视化探索。该文将可视化设计标准作为约束集，并根据实验对约束集中各个子项进行权重分配，生成符合设计标准的可视化。此外，该文还实现了一个基于答案集编程（ASP）的约束系统，旨在实用设计规范帮助用户制作好的可视化设计。

研究方法：

该文将可视化建模为事实集合，并解释了该模型的设计空间以及如何使用约束查询模型。约束求解器可以有效地在定义的空间内搜索最优的可视化规范。系统采用列举生成和测试的方法，将设计标准和算法结合起来，生成最有效的设计。

在求解最优可视化的过程中，使用了硬性可视化约束规范和软性可视化规范。

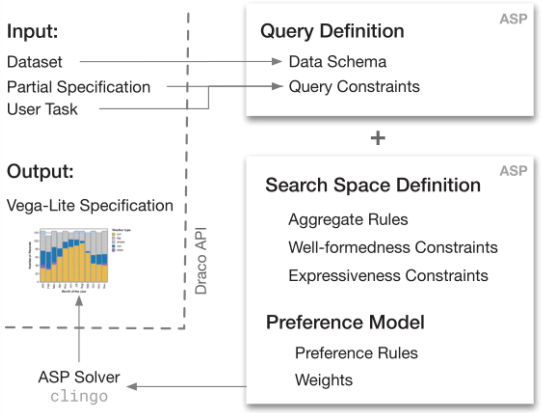
硬性可视化规范包括一些可视化图表内部的逻辑（比方说折线图无法表示种类等等），还有许多用户自己定义的偏好，这些规则必须被满足。比如 Mark ∈ {bar, line, area, point}、Channel ∈ {x, y, color, text, shape}、Field ∈ {site, year, age}、Type ∈ {categorical, continuous}、Aggregate ∈ {sum, mean, count, median} 以及 Zero ∈ {yes, no}等等。

假设有 m 个软性约束，p_i 为第 i 个约束，每个约束的惩罚为 wi，令 $S=\{(p_1, w_1)\cdots(p_i, w_i)\}$ $n_{p_i}(v)$ 是 v 这种视图违反软约束 p_i 的次数，那么 *Cost* 可以定义为：

$$Cost(v) = \sum_{i=1\ldots k} w_i \cdot n_{p_i}(v)$$

这样求解器就可以通过这些权值得到不同可视化图表的偏好。软性可视化规范是可视化的结构特征，它捕捉了可视化属性之间隐藏的关系。

使用以上约束实现了最优的编码搜索过程。Draco 将用户查询（包括数据集、部分规范和任务）编译成一组规则，并将它们与现有的知识库组合起来形成 ASP（回答集编程）程序。然后，Draco 调用 Clingo 来解决程序，以获得最佳答案集。最后，Draco 将答案集转换为 Vega - Lite 规范（Vega-Lite 是一种数据可视化的高级语法，能够快速定义一些基本的交互式数据可视化）。如下图所示：



研究结果：

该文将 Draco 应用到三个不同的场景以测试它的表达性、可拓展性和可用性。结果表明使用约束编程，使得自动化可视化设计工具的开发和维护更加容易。它还可以结合来自不同研究领域的数据学习权重，进一步加速建模工作。

论文题目：*InSituNet: Deep Image Synthesis for Parameter Space Exploration of Ensemble Simulations*

中文题目：InSituNet:用于集成仿真的参数空间探索的深度图像合成

论文作者：Wenbin He, Junpeng Wang, Hanqi Guo, Ko-Chih Wang, Han-Wei Shen, Mukund Raj, Youssef S. G. Nashed, Tom Peterka

论文出处：IEEE Transactions on Visualization and Computer Graphics （SciVis 2019）

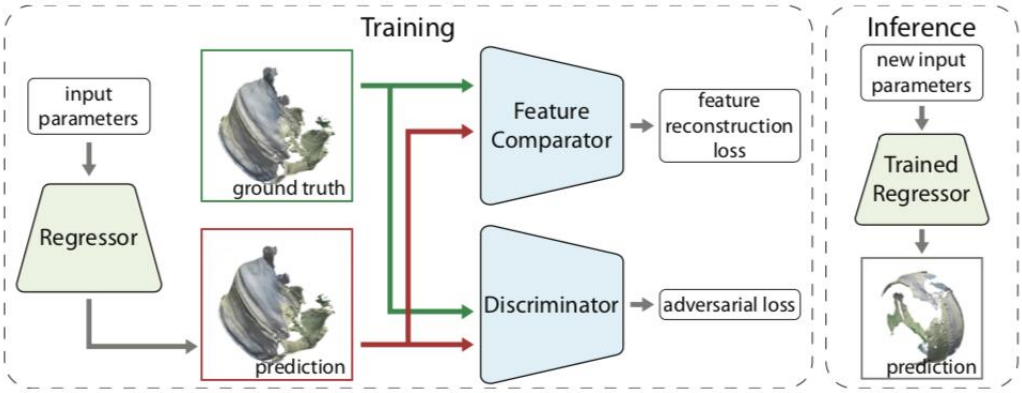
论文地址：https://arxiv.org/pdf/1908.00407.pdf

研究问题：

集成仿真在各种科学和工程学科中发挥着重要作用，如计算流体动力学、宇宙学和天气研究。由于 I/O 和存储方面的限制，在处理大规模模拟时，通过仿真生成原位可视化效果的做法越来越普遍。然而，由于无法获得原始的模拟数据，原位可视化方法限制了事后分析的灵活性。虽然现有研究已经提出了多种基于图像的方法来解决这一问题，但这些方法仍然缺乏探索仿真参数的能力。因此，该文提出了一种基于深度学习的方法——InSituNet，来实现在大规模集成仿真中灵活地探索参数空间。

研究方法：

通过现有的集成仿真，利用不同的仿真参数、视觉映射参数和视图参数生成出对应的可视化图像。这些参数和相应的可视化图像构成一组组数据对，这些数据对将被存储并用于训练 InSituNet。



InSituNet 由三部分组成，分别为回归器、特征比较器和鉴别器。

回归器R_ω是一个深度神经网络模型，将输入参数映射到可视化图像。首先，三种参数分别输入到三组全连接层中，然后将输出连接到另一组全连接层中，将

它们编码成一个隐向量。随后，潜向量被重新构造为低分辨率图像，通过 Residual blocks 进行 2D 卷积和上采样，将其映射为高分辨率输出图像。

随后，利用特征比较器和鉴别器对原始图像和预测图像进行对比。特征比较器是一个预训练的 VGG-19 模型，用于提取和比较预测图像和真实图像之间的图像特征（如边缘、形状），从而获得特征重构损失。而鉴别器是一个深度神经网络，其权值在训练过程中不断更新，来估计预测的分布与真实图像之间的差异，称为对抗损失。二者一起构成了 InSituNet 的损失函数。

利用训练后的 InSituNet，用户可以在不同的仿真与可视化参数设置下生成不同仿真图像，而不需通过真正渲染得到仿真图像，从而能够快速对集成仿真和参数设置进行深入分析。

研究结果：

通过定量和定性的评估，该文作者证明了 InSituNet 对燃烧、宇宙学和海洋三种仿真模拟的有效性，且生成结果优于现有的算法模型。

12.5 可视化进展

- 可视化语法及工具

随着大数据时代的到来，可视化已经成为一个必不可少的工具。现有的可视化软件及工具，可用于设计可视化应用程序和构建可视化分析系统，有助于可视化的广泛使用。为了减少制作可视化的技术负担，一些可视化工具提供了声明性语法，其中包括了 Vega-Lite 和 P5。声明性语法可以将可视化设计与执行细节分离，这使分析人员可以专注于特定于应用程序的设计决策。同时 Vega-Lite 和 P5 都提供了易于使用的编程接口。Vega-Lite 是一套能够快速构建交互式可视化的高阶语法，它是基于 Vega 和 D3 等底层可视化语法的上层封装。相比于其它比较底层可视化语法，Vega-Lite 可以通过几行 JSON 配置代码即可完成一些通用的图表创建，而相反地，想要用 D3 等去构建一个基础的统计图表则可能需要编写多行代码，如果涉及到交互的话代码量更是会大大增加。P5 是一个基于 Web 的可视化工具包，它能集成了 GPU 计算与渐进式处理，并且提供了带有声明性

语法的 API，可用于指定渐进式数据转换和可视化操作，从而帮助分析人员构建融合了高性能计算和渐进式分析工作流的可视化系统[68-69]。

- 可视化与故事叙述

故事叙述是可视化研究的一个重要且新兴的方向。与传统的、强调数据分析的可视化思路不同，故事叙述强调数据的传达与沟通，强调数据与人（且通常是普通人）的连结。在此思路下，研究者们致力于探究：何种数据呈现与讲述技巧，可以使数据具备吸引力、记忆度；数据故事的创作流程是怎样的，存在哪些需求和痛点；以及如何自动生成数据故事等等。用讲故事的方式来呈现可视化，本质上是体现了一种人本导向，即以人的需求出发，提取和分析数据，并以对人友好的方式，将数据中的信息传达出去。随着我们的社会越来越依赖数据赋能，更好地构建数据与人的关系，将成为一条必经之路。叙述可视化的应用，不仅在于那些以“叙述”为主业的领域，如新闻媒体、广告宣传，更在于需要用数据来影响人、说服人、打动人的各行各业。对于研究者来说，相关的研究方向则包括可视化设计、人机交互、认知与感知、智能生成与推荐等等[70-71]。

- 可视化的自动生成

数据可视化领域中大多数的可视化生成系统往往是基于数据的交互式探索，也包括商业领域的知名的可视化工具 Tableau 和 PowerBI。而近些年来，为了避免繁杂的数据分析步骤并提升用户效率，可视化的自动生成逐渐成为行业领域中的研究热点。一系列基于规则和机器学习的推荐方法层出不穷，在自动生成可视化的最新研究中，研究者希望在保证准确表现数据的同时，也能将视觉设计的因素考虑在内，确保可视化的美观性和数据的表现力。例如，DataShot 和 Text-to-Viz，分别从数据和自然语言两个角度去自动生成富有设计感的数据可视化，前者直接从表格数据生成信息简报，后者根据用户的自然语言输入生成对应的信息图。制作一个有效且美观的数据可视化往往需要跨专业领域的技能，尤其是需要同时具备数据分析能力和平面设计能力，而这对于一个没有专业训练的普通用户来说是比较困难的。DataShot 和 Text-to-Viz 等前沿的技术研究均通过自动化的方法从数据洞察和设计美学两个方面帮助用户生成可视化，降低用户制作可视化的门槛，并有效提高生产效率[72-73]。

可解释性深度学习

LSTMVis 是一个递归神经网络的可视化分析工具，它着重于对 RNNs 中的隐藏特征进行可视化分析。LSTMVis 结合了一个基于时间序列的选择界面和一个交互式的匹配工具来搜索大型数据集中相似的隐藏状态模式。系统的主要功能是理解模型中动态变化的隐藏状态。该系统允许用户选择一个假设的输入范围来关注局部的改变，将这些状态改变与大型数据集中类似的模式进行匹配，并将这些选择出来的模式进行对齐分析。RNNs 在序列建模方面有着重要的作用，但是模型中的隐藏层含义很难被解释清楚。对于一个完成训练的 RNN 模型，分析人员并不清楚这个模型是如何理解序列中不同节点之间的关系的。LSTMVis 能够帮助用户交互式地探索 RNN 模型复杂的网络结构，并将模型中抽象表示的隐藏层信息与人类可理解的原始输入进行关联[74]。

12.6 可视化应用

随着 21 世纪大数据的兴起和发展，大数据可视化广泛应用于各个领域，本节重点介绍其中的社交媒体可视化、医疗信息可视化和体育数据可视化。

12.6.1 社交媒体可视化

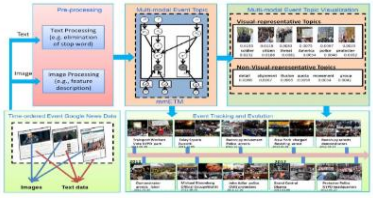
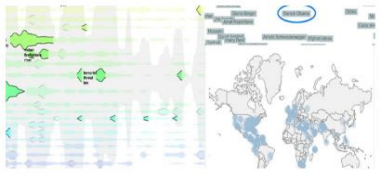
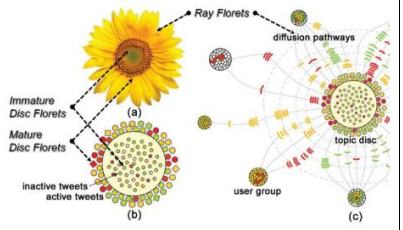
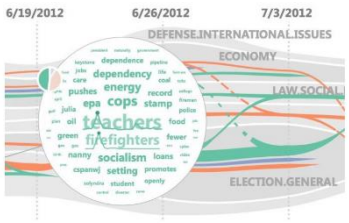
社交媒体，比如最近几年非常流行的 Twitter、Facebook、微博。它们可以作为强大的在线交流平台，允许数百万用户在任何时间、任何地点产生、传播、共享或交换信息。这些信息通常包括多种多媒体内容，如文本、图像和视频。在社交媒体上传播的大量多媒体数据，涵盖了全球范围内大规模和实时发生的社会动态信息，这种现象为社交媒体可视化提供了很多机会。

社交媒体技术层面上的可视化，主要包括：基于关键字方法的可视化，基于主题方法的可视化和多元方法的可视化。现有的研究大多集中于集体行为的可视化，这类研究的主题包括：信息扩散的可视化，社会竞争与合作的可视化，人的流动性的可视化。

社交媒体数据的可视化分析正在迅速发展，每年都有大量的新方法出现。然而，该领域仍处于起步阶段，面临许多挑战和悬而未决的问题。许多挑战不能仅使用来自一个规程的技术来解决。但是，将可视化、多媒体、NLP 和人机交互相

结合的多学科研究，将带来处理和理解社交媒体数据会有更强大、更可行的方法和技术。具体社交媒体可视化如下表所示：

表 12-2 社交媒体可视化介绍

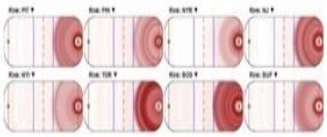
可视化方法	图例	特点
基于关键字方法的可视化		这种方法，用于跟踪和探索社交媒体上大型活动的在线对话。提供了从时间、主题、社交和图像方面对会话的可视化总结。可以灵活使用时间和空间过滤器，从社交媒体上海量的信息中获取自己想要的信息。
基于主题方法的可视化		主题方法使用基于三个标准(视觉相关性、视觉连贯性和独特性)确定的代表性图像进行可视化。一个多模态框架，用于从多模态信息中检测主题，跟踪主题的演变，并随时间的变化使用文本和图像可视化主题。
多元方法的可视化		多元方法帮助用户从多个角度获取关于社交媒体事件的信息。与其他主要侧重于理解社交媒体消息文本内容的方法相比，多元方法通过结合高级数据挖掘方法来保持态势感知，并提供全面的概述。
信息扩散的可视化		在信息扩散的可视化方法上，有不同的模型来描述信息在社交媒体上的传播，比如综合的多源驱动的异步扩散模型。往往直观地呈现出社会化媒体传播过程的三个主要特征，即时间趋势、社会空间范围和社区对某一兴趣话题的反应。
社会竞争与合作的可视化		为了便于可视化分析，引入流式可视化，用以说明每个主题的竞争力随时间的变化。并展示出促进或降低主题传播行为被可视化为线程，这表明他们的角色随着时间的推移而变化。从不同的角度展示社交媒体用户群体之间的合作与竞争行为。
人的流动性的可视化		用户可以在此基础上探索人流动性的语义，如流动方式、频繁访问序列、关键字描述等。采用启发式模型来减少数据的不确定性，从而增加对可靠数据的适当选择。

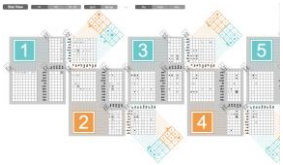
12.6.2 体育数据可视化

可视化技术在体育行业的运用越来越普遍。例如观众通过电视观看体育赛事时，经常能看到关于比赛情况的解读。以前使用技术数据统计与空间数据的方法来演示比赛实况，但用户并不能从可视化结果中看出比赛的发展趋势。现在，将

技术数据统计和时间数据相结合，就能很好地反映比赛过程。基于对整场比赛或者比赛中某个时间段的赛况的可视化，大大帮助了用户分析比赛进程的动态变化。体育可视化使用的时间概念已经对时间做了一定程度的简化，它将客观事件发生的时间作为规律的时间单元进行处理，这种可视化方法能够反映比赛的趋势,很大程度上增加了用户的体验效果。从数据角度出发，可以将体育数据可视化分为以下几类，如下表所示：

表 12-3 体育数据可视化

可视化方法	图例	特点
技术统计数据的可视化方法		根据统计数据的不同，具体的可视化方法千差万别。通常会根据数据的属性选择较好的呈现方式，并辅之以良好的交互手段。
技术统计数据和空间数据结合的可视化方法		大部分体育比赛都是在一定规模的场地中进行，球员位置和事件产生位置等都是至关重要的数据。因此仅对技术统计可视化可能会遗漏掉重要的信息。技术统计数据和空间数据结合的可视化方法能够对比赛场地进行等比例的简化绘制，并且将技术统计数据绘制到对应的空间位置上。
技术统计数据和时间数据结合的可视化方法		如果采用技术统计数据与空间数据的方法展示比赛，用户并不能从可视化结果中看出比赛的发展趋势。然而技术统计数据和时间数据相结合的方法就能很好地反映比赛过程。基于对整场比赛或者比赛中某个时间段的赛况的可视化，有利于帮助用户分析比赛进程的动态变化。
技术统计数据和时空数据可视化方法		足球比赛中阵型是一个时空数据，随比赛的进行在时间和空间维度上都会有很大的变化，球队分析师很难通过观看视频和查看统计数据看出球队整场比赛下来阵型的一个变化。使用流视图对阵型时空数据进行可视化能给出阵型在整场比赛的一个直观的变化概览。分析师通过查看概览发现异常的地方后，还可以对流进行选定，进一步查看这一段时间内球队的技术统计数据以及视频回放，进行细致的分析。

技术统计数据的关联可视化方法		单靠技术统计数据的整体时序变化趋势或者两两对比结果，很难找到比赛中运用的高层次的技战术策略，而不同技术统计数据之间的关联分析则可以帮助用户发现许多技术数据之间的关系，从而揭示出比赛中涉及的一些技战术策略。尤其是对于像乒乓球这样包含丰富战术变化的运动来说，不同种类技术之间的关联分析可以帮助分析运动员的打法风格和技术上的优势与劣势。
轨迹可视化		对于体育数据分析而言，比赛的轨迹绘制不容置疑是非常重要的。在球类比赛中，球的轨迹或者运动员的轨迹都能够反映出动态规律，通过对重要轨迹的分析可以掌握重要的比赛和球员线索。 再现比赛：通常低精度的重现可以作为一种更生动的图文直播；高精度的再现则可以做到对体育比赛的完全掌握和分析，它具有强大的分析能力。在一场冰球比赛中，这个系统可以实时地显示球场上发生的事件，新的事件通过特定标志表示，显示在球场上的对应位置以及下方的时间线上。

12.6.3 医疗数据可视化

医疗健康领域是与每一个直接密切相关的重要科学领域。一直以来，科学家在探索生命奥秘，及疾病产生机理的过程中，一直重视对跨学科技术的运用。从基于虚拟现实技术的仿真手术到手术机器人，从医学成像技术到医学图像处理，从大数据分析到人工智能，越来越多的新兴科技正在被应用到医疗领域，每一次技术上的革新与成功应用，都给医疗领域带来了全新诊疗技术，及科研手段，也提高了就诊治疗过程中患者的安全。

2010 年，医疗 2.0 伴随着互联网 2.0 的热潮应运而生，旨在利用一切先进科学技术帮助提升诊疗手段，攻克医学难题。2011 年，美国国家医学院（MOI）发布的该年度报告中特别指出，相对于其他技术在医疗领域的引用，信息可视化技术在医疗领域的应用显得尤为滞后，现有技术无法满足信息展现、用户交互、数据分析等众多需求。近些年来，伴随着大数据及人工智能技术在医疗领域的应用

与普及，信息可视化及相应的可视分析技术在医疗领域的应用也得到了长足的发展。

可视化技术，尤其是科学可视化技术在医疗领域长久以来一直扮演者重要角色。例如，无论是平面 X 光扫描，还是三维 CT 影像，都应用到了科学可视化的相关技术。然而这些技术仍然局限于对于具象数据（例如，人体的骨骼、器官组织结构等）的展现。随着互联网的普及以及可穿戴设备的广泛应用，越来越多的与医疗相关的抽象数据被采集了上来，因此对信息可视化技术提出新的需求，这也正是 MOI 报告中所指出来的相关问题。针对这些新的数据与需求，在医疗 2.0 概念的范畴下，新的信息可视化技术被主要开发用来：

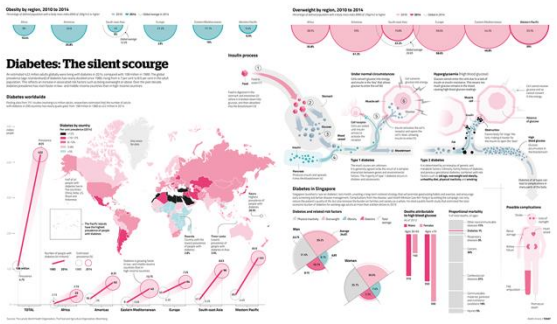
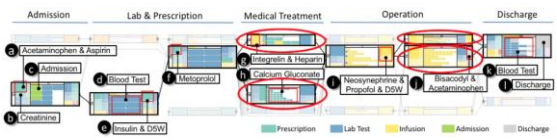
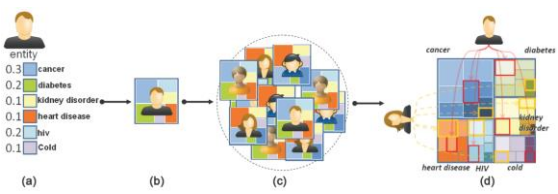
1) 展现用户的个人健康信息。例如，用于展现用户心跳、血压等状态的可视化用户界面；

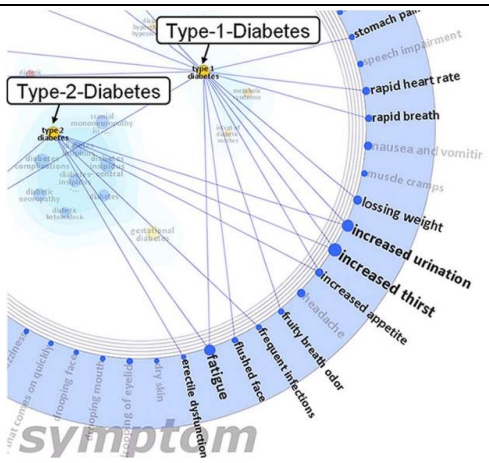
2) 汇总并展现公众健康信息。例如，用于汇总展现禽流感在扩散趋势，或不同地区的人民健康状况等统计信息的可视化大屏；

3) 分析并展现临床电子病历记录中的规律与模式。例如，疾病的演变过程，以及诊疗方案的疗效等。前两类可视化面向的一般是不具备医疗知识的普通用户，因此，往往采用传统直观的信息可视化形式，如，柱状统计图，折线图等，以便于用户的理解与阅读。第三类技术，主要面向的是医生等具有专业背景，需要对数据进行深入调查，并做出职业判断的用户，因此这类可视化更具有针对性，可视化及相关分析技术的设计也更具挑战性。

表 12-4 医疗数据可视化

可视化方法	案例	特点
医疗图像数据的可视化		采用针对标量场的可视化技术，通过颜色映射、等值线和高度图三类等方式展现数据在空间的分布，常见于呈现 X 光及 CT 成像。

针对个人统计健康信息的可视化方法		采用简单的以统计图、可视化图标等简单形式为主的可视化技术，针对无技术背景的终端用户，展现其个人健康数据。这类可视化的设计旨在采用最直观易懂的方式表达并展现如心跳、血压、体温等信息。
针对公众信息的可视化		采用信息图或可视化大屏的直观表达方式展现疾病的分布、传播等公众健康数据。
针对电子病例中时间序列数据的可视化方法		采用时间序列可视化的展现形式表达心电、血压、脑电波等随时间连续变化的数据。
针对电子病历中事件序列数据的的可视化方法		汇总并展现电子病历中所记录的医疗事件发生发展的阶段性过程，从而揭示疾病在不同人群中发展的规律，以及不同诊疗方案所带来的不同疗效。
针对电子病历中病人个体及群体特征的可视化		采用多维度数据可视化技术，展现电子病历数据中病人的个体及群体多维度特征，主要用于群体分析（Cohort Analysis）以区分不同类型的病人。

医疗知识图谱的可视化		采用图的可视化技术，展现大规模异构医疗知识图谱，用于方便知识的检索与查询。

13 数据挖掘

13.1 数据挖掘概念

数据挖掘（Data Mining），是指从大量的数据中自动搜索隐藏于其中的有着特殊关系性的数据和信息，并将其转化为计算机可处理的结构化表示，是知识发现的一个关键步骤。数据挖掘的广义观点：从数据库中抽取隐含的、以前未知的、具有潜在应用价值的模型或规则等有用知识的复杂过程，是一类深层次的数据分析方法。数据挖掘是一门综合的技术，涉及统计学、数据库技术和人工智能技术的综合，它的最重要的价值在于用数据挖掘技术改善预测模型。数据挖掘涉及的常见的任务有[75]：

数据表征：是对目标类数据的一般特征或特征的总结。对应于用户指定类的数据通常通过数据库查询收集。例如，要研究上一年销售额增长 10％的软件产品的特征，可以通过执行 SQL 查询来收集与此类产品相关的数据。

异常检测：数据库可能包含不符合数据一般行为或模型的数据对象，这些数据对象即被成为异常值。大多数数据挖掘方法将异常值视为噪声或异常。但是，在诸如欺诈检测等应用中，罕见事件可能比更常见的事件更有价值。异常值数据的分析被称为异常值挖掘。

关联规则学习：搜索变量之间的关系。例如，一个超市可能会收集顾客购买习惯的数据。运用关联规则学习，超市可以确定哪些产品经常一起买，并利用这些信息帮助营销。这有时被称为市场购物篮分析。

聚类：是在未知数据的结构下，发现数据的类别与结构。聚类算法基于最大化类内相似性和最小化类间相似性的原则对对象进行聚类或分组。也就是说，形成对象集群，使得集群内的对象彼此之间具有较高的相似性，但与其他集群中的对象非常不相似。每个形成的集群都可以被视为一类对象，从中可以派生出规则。聚类还可以促进分类的形成，也就是将观察组织成一个将类似事件归类在一起的类的层次结构。

分类：分类是查找描述和区分数据类别或概念的模型（或函数）的过程，目的是为了能够使用模型来预测类别标签未知的对象的类别。例如，一个电子邮件程序可能试图将一个电子邮件分类为“合法的”或“垃圾邮件”。

回归：试图找到能够以最小误差对该数据建模的函数。回归分析是最常用于数字预测的统计方法，但也存在其他方法。预测还可以根据现有数据确定趋势。

数据演化分析：描述并建模其行为随时间变化的对象的规则或趋势。虽然这可能包括时间相关数据的表征，区分，关联和相关分析，分类，预测或聚类，但这种分析的明显特征包括时间序列数据分析，序列或周期性模式匹配以及基于相似性的数据分析。

本报告分析了近年来数据挖掘领域的高水平学术论文，挖掘出了包括社交网络、大数据、情报分析、聚类分析、文本挖掘、用户行为、推荐系统、离群检测、专家系统等相关关键词近年来全球活跃的学术研究。此外，结合知识图谱技术，本报告将以上研究领域表示为如下图谱结构

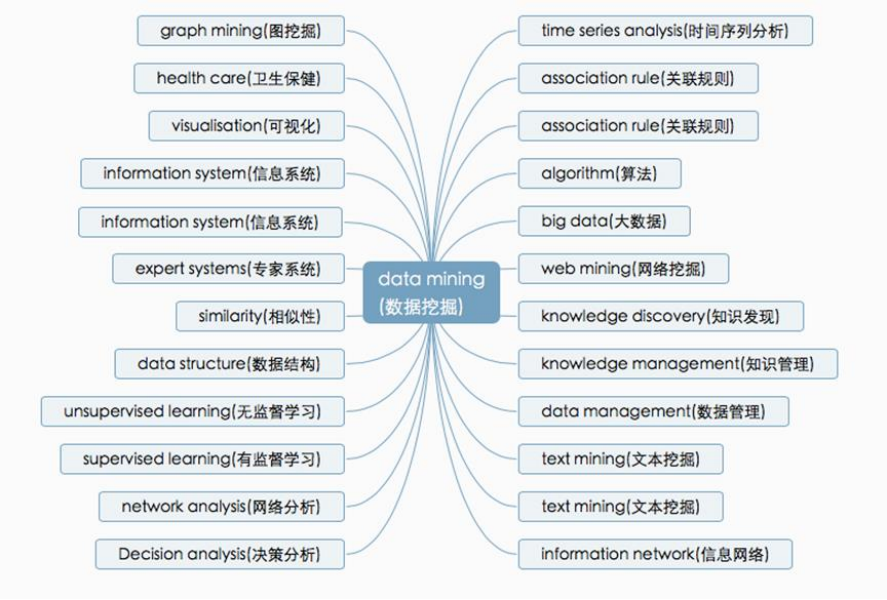


图 13-1 Data Mining 知识图谱

具体分析和处理的方法如下：

1. 使用自然语言处理技术，提取每篇论文文献的关键词，据此，结合学科领域知识图谱，将文章分配到相应领域；

2. 依据学科领域对论文文献进行聚类，并统计论文数量作为领域的研究热度；

3. 领域专家按照领域层级对学科领域划分等级，设计了三级图谱结构，最后根据概念热度定义当前研究热点。三级详细数据可以参见本报告附录，或到 https://www.aminer.cn/data 中直接下载原始数据。

13.2 数据挖掘的发展历史

随着数据体量的快速增加，人们希望有一种方法可以帮助处理这些纷繁复杂的数据，从中发现有价值的信息或知识为决策服务，数据挖掘在此背景下应运而生，下文将介绍数据挖掘的起源和发展历程。

20 世纪 60 年代，数据搜集阶段。在这个阶段受到数据存储能力的限制，特别是当时还处在磁盘存储的阶段，因此主要解决的是数据搜集的问题，而且更多是针对静态数据的搜集与展现，所解决的商业问题，也是基于历史结果的统计数据上的。

20 世纪 70 年代，数据存储阶段。随着数据库管理系统趋于成熟，存储和查询百万兆字节甚千万亿字节成为可能。而且，数据仓库允许用户从面向事物处理的思维方式向更注重数据分析的方式进行转变。然而，从这些多维模型的数据仓库中提取复杂深度信息的能力是非常有限的。

20 世纪 80 年代，数据分析访问阶段。关系性数据库与结构性查询语言的出现，使得动态的数据查询与展现成为可能，人们可以用数据来解决一些更为聚焦的商业问题。在这个阶段，KDD 出现了，数据挖掘走进了历史舞台。也正是在这个时期，出现了些成熟的算法，能够“学习”数据间关系，相关领域的专家能够从中推测出各种数据关系的实际意义。

20 世纪 90 年代，数据仓库决策与支持阶段。OLAP 与数据仓库技术的突飞猛进使得多层次的数据回溯与动态处理成为现实，人们可以用数据来获取知识，对经营进行决策，零售公司和金融团体使用数据挖掘分析数据和观察趋势以扩大客源，预测利率的波动，股票价格以及顾客需求。

21 世纪至今，真正的数据挖掘的时代。现在是大数据的时代，因为信息化的发展非常快，数据也在不断更新，相应的数据研究也越来越难。我们需要对这些大数据进行处理，从中提取出有价值的信息。随着信息技术的发展，数据挖掘已经越来越成熟，成为一门交叉学科。一般来说，数据挖掘结合了数据库，人工智能，模式识别，神经网络，机器学习，统计，高性能计算，数据可视化，空间数据分析和信息检索等很多方面的知识[76]。

13.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为数据挖掘领域全球学者分布情况：



图 13-2 数据挖掘领域全球人才部分

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量遥遥领先且主要分布于其东西海岸；欧洲、亚洲也有较多的人才分布；其他诸如非洲、南美洲等地区的学者非常稀少；可视化领域的人才分布与各地区的科技、经济实力情况大体一致。

此外，在性别比例方面，数据挖掘领域中男性学者占比 89.4%，女性学者占比 10.6%，男性学者占比远高于女性学者。

数据挖掘领域学者的 h-index 分布如下所示，大部分学者的 h-index 分布在中间区域，其中 h-index 在 20-30 区间的人数最多，有 683 人，占比 33.9%，小于 20 区间的人数最少，共 138 人。

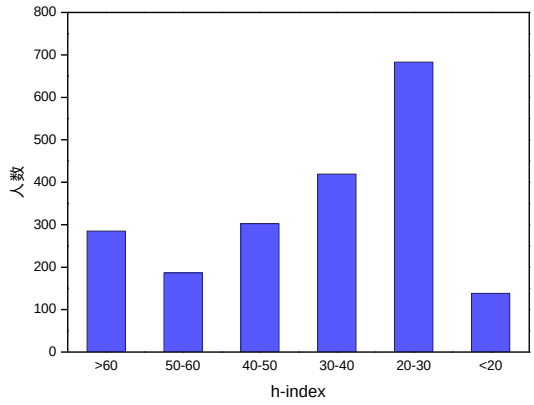


图 13-3 数据挖掘领域学者 h-index 分布

- 中国人才分布



图 13-4 数据挖掘领域中国学者分布

我国专家学者在数据挖掘领域的分布如上图所示。通过下图我们可以发现，京津地区在本领域的人才数量最多，其次是珠三角和长三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，

通过观察中国周边国家的学者数量情况，特别是与日本、东南亚等亚洲国家相比，中国在数据挖掘领域学者数量较多且有一定的优势。

中国与其他国家在数据挖掘领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 13-1 数据挖掘领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	475	35338	74	986
中国-新加坡	49	3841	78	106
中国-澳大利亚	34	1266	37	65
中国-加拿大	32	2938	92	65
中国-英国	23	515	22	41
中国-德国	12	177	15	21
中国-印度	9	638	71	25
中国-瑞士	7	528	75	20
中国-荷兰	6	101	17	10
中国-沙特阿拉伯	5	183	37	8

从上表数据可以看出，中美合作的论文数、引用数、学者数遥遥领先，表明中美间在数据挖掘领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席；中国与加拿大合作的论文数虽然不是最多，但是拥有最高的平均引用数说明在合作质量上中加合作达到了较高的水平。

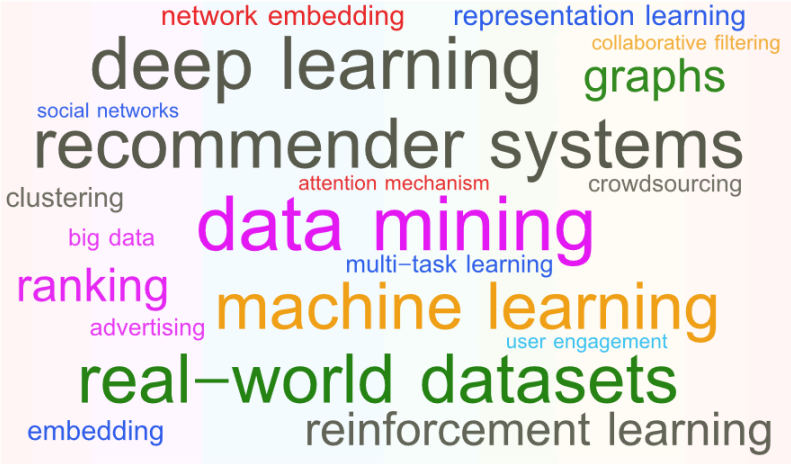
13.4 论文解读

本节对本领域的高水平学术会议论文进行挖掘，解读这些会议在 2018-2019 年的部分代表性工作。会议具体包括：

ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

ACM International Conference on Web Search and Data Mining

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，数据挖掘（data mining）、深度学习（deep learning）、推荐系统（recommender systems）是本领域中最热的关键词。



论文题目：*Graph Convolutional Neural Networks for Web-Scale Recommender Systems*

中文题目：图卷积神经网络应用于网络规模推荐系统

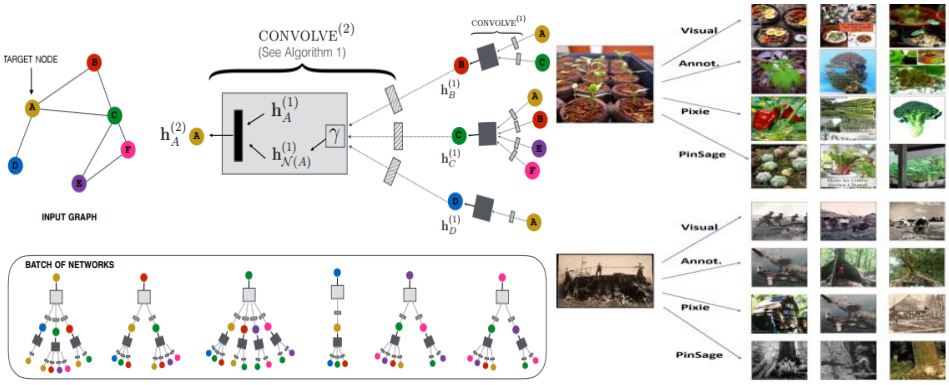
论文作者：Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L. Hamilton, and Jure Leskovec.

论文出处：In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining （KDD '18）.

论文地址：https://dl.acm.org/citation.cfm?doid=3219819.3219890

研究问题：

图结构数据作为深层神经网络最新研究成果应用于推荐系统基准的最新的指标，传统的深度学习网络主要针对图片，语音等欧氏空间内规则型数据，但是现实中存在很多不是欧氏数据，它们的结构不规则，难以用卷积神经网络对其进行结构信息的聚合，故将其扩展到有数十亿用户的网络级推荐系统是一个巨大的挑战。



研究方法：

本研究提出了一个数据有效格拉夫卷积网络（GDN）算法 PinSage，并结合高效的随机游动和局部图卷积来生成有两个图结构和节点（项目）特征的嵌入模型。该方法的关键计算工作是局部图卷积的概念。为了产生用于一个节点（一个项目）的嵌入，我们应用其聚集来自所述节点的本地图形附近的特征的信息（例如，视觉，文本特征）（上图所示）的多个卷积模块。每个模块学习如何聚合信息从一个小图附近，并通过堆叠多个这样的模块，我们的方法可以得到有关本地网络拓扑信息。重要的是，这些局部卷积模块的参数在所有节点共享，使得独立于输入图形大小的我们的方法的参数的复杂性。与之前 GCN 方法相比，该方法提高了模型的鲁棒性和收敛性；用 2 个深度卷积来概述提出的方法，左边将目标节点输入图形，经过底部计算目标节点的两层神经网络，每个目标的神经网络不同，但它们都有相同的共享参数集（即右边的函数和算法，矩形框和细框表示紧密连接的多层神经网络），输出与之关联的目标节点。

研究结果：

本文评估了 PinSage 在两个任务中生成的嵌入:推荐相关的 pin 和在用户的 home/news feed 中推荐 pin。为了推荐相关的引脚，我们在嵌入空间中选择查询引脚的 K 个近邻。我们使用离线排名和受控用户研究来评估这个相关 pin 推荐任务的性能。对于 homefeed 推荐任务，我们选择嵌入空间中距离用户最近的固定项 pin。我们使用 A / B 测试来评估完全部署的生产系统在此任务上的性能，以度量对用户参与的总体影响。在这种测试情况下 PinSage 是表现最好的，60%的情况下 PinSage 的推荐结果是优于 Pixie 的。

论文题目：*Unbiased Learning-to-Rank with Biased Feedback*

中文题目：基于有偏差的反馈数据进行无偏差的 LTR

论文作者：Joachims, Thorsten and Swaminathan, Adith and Schnabel, Tobias.

论文出处：In Proceeding WSDM '17 Proceedings of the Tenth ACM International Conference on Web Search and Data Mining.

论文地址：https://dl.acm.org/citation.cfm?id=3018699

研究问题：

隐藏的反馈数据（如点击，停留时间等）是人类交互系统的数据的丰富来源。虽然隐藏的反馈数据具有很多优点（如采集费用低，以用户为中心且实时性强），但如何从有偏差的反馈数据中，训练无偏差的排序模型，是这些数据有效利用的一个主要障碍。例如，搜索排名位置偏差会强烈影响检索结果的点击次数，往往只对排名较前的数据进行点击，后面的即使与检索内容相关，也不会被点击，用户的反馈就受到现在系统的影响，导致检索无法提升到全局最优的情况。

研究方法：

本文使用一种可证明的无偏估计器，用来评估有偏差的反馈数据评估排序性能，并且提出一种针对 LTR 的倾向加权经验风险最小化的方法，得到了一个倾向加权排序支持向量机（倾向加权是用来纠正选择偏差时，丢弃查询没有点击在学习排序，此工作的关键是认识到逆倾向得分可以使用的更有力,排名位置偏差,相信偏差,语境效果,文档等使用较多的点击模型来估计每个点击的倾向而不是查询获得点击的倾向），将其用于隐式反馈中进行判别学习，其中点击模型扮演倾向估计器的角色。与应用点击模型来消除点击数据偏差的大多数传统方法相比，这个方法允许对排序函数进行训练。

$$
\begin{aligned}
\hat{w} \;=\; & \operatorname{argmin}_{w,\xi} \; \frac{1}{2} w \cdot w + \frac{C}{n} \sum_{j=1}^{n} \frac{1}{q_j} \sum_{y \in Y_j} \xi_{jy} \\
s.t. \quad & \forall y \in Y_1 \backslash \{y_1\} : w \cdot [\phi(\boldsymbol{x}_1, y_1) - \phi(\boldsymbol{x}_1, y)] \geq 1 - \xi_{1y} \\
& \vdots \\
& \forall y \in Y_n \backslash \{y_n\} : w \cdot [\phi(\boldsymbol{x}_n, y_n) - \phi(\boldsymbol{x}_n, y)] \geq 1 - \xi_{ny} \\
& \forall j \forall y : \xi_{jy} \geq 0.
\end{aligned}
$$

$$
\begin{aligned}
\operatorname{rank}(y_i|\boldsymbol{y}) - 1 &= \sum_{y \neq y_i} \mathbb{1}_{w \cdot [\phi(\boldsymbol{x}_i, y) - \phi(\boldsymbol{x}_i, y_i)] > 0} \\
&\leq \sum_{y \neq y_i} \max(1 - w \cdot [\phi(\boldsymbol{x}_i, y_i) - \phi(\boldsymbol{x}_i, y)], 0) \\
&\leq \sum_{y \neq y_i} \xi_{iy}.
\end{aligned}
$$

本文提出的模型通过一个松弛公式有效地求解这类二次规划问题，并且我们使用支持向量机 rank1 适当修改基于逆倾向得分（IPS）局部信息特征的权值 1/qj。生成的代码将应用于在线推理模型。通过 Position-based Click Model with Click Noise（PCMCN）进行排序文档的建模，用以方便预测点击概率（前系统选择各个文档的概率）。

研究结果：

通过提供了广泛的经验证据 https://www.joachims.org/svm light/svm rank.html，证明了使用倾向加权排序支持向量机实例化经验风险最小化方法，并通过广泛的经验证据表明，本文提出的学习方法对选择偏差、噪声和模型错误规范具有鲁棒性。此外，我们在一个实时搜索引擎上的实验表明，该方法在学习过程中没有任何启发或手动干预的情况下，可以显著地提高检索效果。

题目：*Personalized Top-N Sequential Recommendation via Convolutional Sequence Embedding*

中文题目：基于卷积序列嵌入的个性化 Top-N 序列推荐

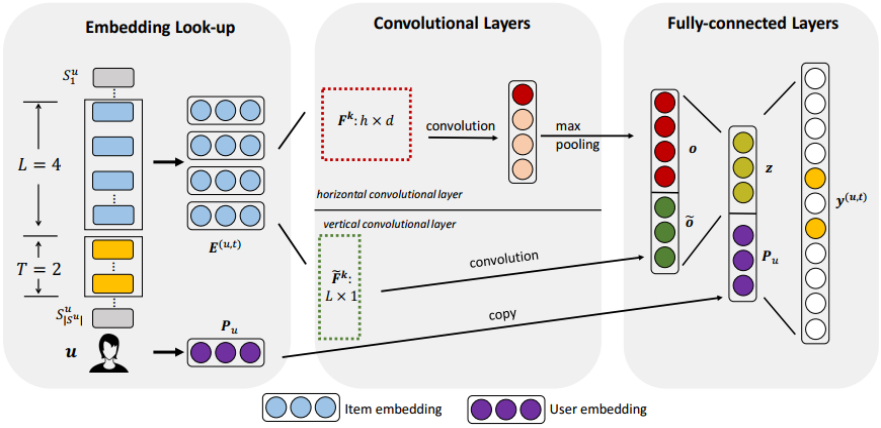
论文作者：Tang, Jiaxi，Wang, Ke.

论文出处：In Proceeding WSDM '18 Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining.

论文地址：https://arxiv.org/abs/1809.07426

研究问题：

推荐系统渐渐成为许多应用软件的核心技术，目前往往只考虑拥有一般偏好的推荐系统在销售 iPhone 后将错过推荐手机配件的机会，因为购买手机配件并不是一个长期的用户行为，现有的 Top-N 序列模型不能对联合级别的顺序模式进行建模并且没有考虑先后互换的跨越行为，Top -N 顺序推荐将每个用户建模为过去交互的项目序列，并旨在预测用户在“不久的将来”可能交互的排名前 n 的项目。交互顺序意味着顺序模式扮演了一个重要的角色，其中序列中最近的项目对下一个项目有更大的影响。鉴于此，本文提出一种卷积序列嵌入推荐模型（简称 Caser）来解决 top-N 序列推荐问题。



研究方法：

本文提出一个卷积序列嵌入推荐模型（Caser）作为一个解决方案来解决这一需求。我们的想法是在时间和潜在空间中，把一系列最近的项目嵌入到一个“图像”中，将顺序模式作为图像的局部特征并使用卷积滤波器学习这些特征，其使用水平和垂直卷积过滤器来捕获点级、联合级和跳过行为的顺序模式，再对用户的一般偏好和顺序模式进行建模，并在一个统一的框架中概括了几种现有的先进方法。故此方法可以提供一个统一的、可构建（可性）的网络结构，用于获取一般偏好和顺序模式。Caser 网络结构图如上图所示，矩形框表示用户序列中的条目 S_1^u，•....$S_{|Su|}^u$，内嵌有圆圈的矩形框表示某个向量，如用户嵌入 Pu。虚线矩形框是具有不同尺寸的卷积滤波器。卷积层中的红色圆圈表示每个卷积结果的最大值。我们使用之前的 4 个行为($L = 4$)来预测这个用户将在接下来的 2 个步骤（$T = 2$）中与哪些项目交互。

研究结果：

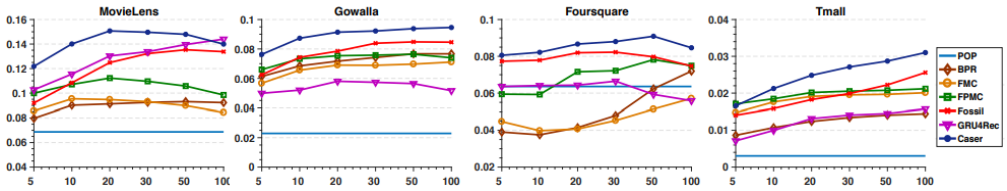
通过公共数据集上的实验表明，Caser 优于最先进的序列推荐方法，作为一种新颖的 top-N 序列推荐解决方案，Caser 将最近的动作建模为时间和潜在维度之间的“图像”，并应用卷积滤波器学习序列模式。Caser 提供了一个统一和灵活的网络结构捕捉序列推荐的许多重要的功能（实现点级和联合级顺序模式、跳过行为和用户一般偏好）。

Table 1: Statistics of the datasets

Datasets	Sequential Intensity	#users	#items	avg. actions per user	Sparsity
MovieLens	0.3265	6.0k	3.4k	165.50	95.16%
Gowalla	0.0748	13.1k	14.0k	40.74	99.71%
Foursquare	0.0378	10.1k	23.4k	30.16	99.87%
Tmall	0.0104	23.8k	12.2k	13.93	99.89%

Table 2: Performance comparison on the four data sets.

Dataset	Metric	POP	BPR	FMC	FPMC	Fossil	GRU4Rec	Caser	Improv.
MovieLens	Prec@1	0.1280	0.1478	0.1748	0.2022	0.2306	**0.2515**	0.2502	-0.5%
	Prec@5	0.1113	0.1288	0.1505	0.1659	0.2000	0.2146	**0.2175**	1.4%
	Prec@10	0.1011	0.1193	0.1317	0.1460	0.1806	0.1916	**0.1991**	4.0%
	Recall@1	0.0050	0.0070	0.0104	0.0118	0.0144	**0.0153**	0.0148	-3.3%
	Recall@5	0.0213	0.0312	0.0432	0.0468	0.0602	0.0629	**0.0632**	0.5%
	Recall@10	0.0375	0.0560	0.0722	0.0777	0.1061	0.1093	**0.1121**	2.6%
	MAP	0.0687	0.0913	0.0949	0.1053	0.1354	0.1440	**0.1507**	4.7%
Gowalla	Prec@1	0.0517	0.1640	0.1532	0.1555	0.1736	0.1050	**0.1961**	13.0%
	Prec@5	0.0362	0.0983	0.0876	0.0936	0.1045	0.0721	**0.1129**	8.0%
	Prec@10	0.0281	0.0726	0.0657	0.0698	0.0782	0.0571	**0.0833**	6.5%
	Recall@1	0.0064	0.0250	0.0234	0.0256	0.0277	0.0155	**0.0310**	11.9%
	Recall@5	0.0257	0.0743	0.0648	0.0722	0.0793	0.0529	**0.0845**	6.6%
	Recall@10	0.0402	0.1077	0.0950	0.1059	0.1166	0.0826	**0.1223**	4.9%
	MAP	0.0229	0.0767	0.0711	0.0764	0.0848	0.0580	**0.0928**	9.4%
Foursquare	Prec@1	0.1090	0.1233	0.0875	0.1081	0.1191	0.1018	**0.1351**	13.4%
	Prec@5	0.0477	0.0543	0.0445	0.0555	0.0580	0.0475	**0.0619**	6.7%
	Prec@10	0.0304	0.0348	0.0309	0.0385	0.0399	0.0331	**0.0425**	6.5%
	Recall@1	0.0376	0.0445	0.0305	0.0440	0.0497	0.0369	**0.0565**	13.7%
	Recall@5	0.0800	0.0888	0.0689	0.0959	0.0948	0.0770	**0.1035**	7.9%
	Recall@10	0.0954	0.1061	0.0911	0.1200	0.1187	0.1011	**0.1291**	7.6%
	MAP	0.0636	0.0719	0.0571	0.0782	0.0823	0.0643	**0.0909**	10.4%
Tmall	Prec@1	0.0010	0.0111	0.0197	0.0210	0.0280	0.0139	**0.0312**	11.4%
	Prec@5	0.0009	0.0081	0.0114	0.0120	0.0149	0.0090	**0.0179**	20.1%
	Prec@10	0.0007	0.0063	0.0084	0.0090	0.0104	0.0070	**0.0132**	26.9%
	Recall@1	0.0004	0.0046	0.0079	0.0082	0.0117	0.0056	**0.0130**	11.1%
	Recall@5	0.0019	0.0169	0.0226	0.0245	0.0306	0.0180	**0.0366**	19.6%
	Recall@10	0.0026	0.0260	0.0333	0.0364	0.0425	0.0278	**0.0534**	25.6%
	MAP	0.0030	0.0145	0.0197	0.0212	0.0256	0.0164	**0.0310**	21.1%



论文题目：*Theoretical Impediments to Machine Learning With Seven Sparks from the Causal Revolution*

中文题目：机器学习的理论局限性与因果革命的七个火花

论文作者：Judea Pearl.

论文出处：Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining

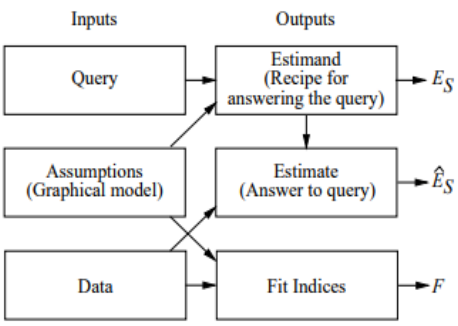
论文地址：https://dl.acm.org/citation.cfm?id=3176182

研究问题：

当前的机器学习系统几乎完全以一种统计的或无模型的模式运行，即通过输入进行参数优化以提高系统的性能，这对机器学习系统的能力和性能造成了严重的理论限制。这样的系统不能对干预和回顾进行推理，因此这种系统不能作为强人工智能的基础。为了实现人类级别的智能，学习系统应该需要真实的模型的指

导，就如同在因果推理任务中的模型一样。为证明这些模型的重要作用，文章概述了七个当前机器学习系统无法实现、但可以使用因果建模工具完成的任务。

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?



研究方法：

图形和结构模型的进步使得假设在计算上易于管理，从而使模型驱动的推理成为一个更有前途的方向，可以在此基础上建立强人工智能，为描述机器学习系统所面临的障碍，即左图所示，使用一个从关联（看正在发生什么）到干涉（做出干预措施）、最后到反事实（想象，回顾）的三层结构来体现因果推理中的推论，并且在因果关系的层次结构中，只有在 i 级或更高级别的信息可用时，才能回答 i 级的问题。哈佛大学教授 Garry King 认为，“在过去的几十年里，人们已经学到了比以往所有历史记录中所学到的所有内容的总和还要多的关于因果推断的知识”，文章称其为“因果革命”，相应的数学框架称为结构因果模型（Structural Causal Models, SCM）。右图为以推理机的形式表述结构因果模型，其包括询问，假设和数据 3 个输入和预估计（回答问题的秘密方法），估计（问题的答案）及拟合指数（用来测量数据与模型所传达的假设之间的兼容性。）三个输出。接下来，文章介绍了 SCM 框架的 7 项最重要的特性，即需要结构因果模型来完成的七个重点任务：1.编码因果假设的透明度和可测试性；2.do-calculus 和混杂控制；3.反事实算法；4.调节分析和直接、间接影响评估；5.外部效度和样本选择偏差；6.缺失数据；7.因果关系发现。

研究结果：

为了达到人类的智能水平，机器学习需要一个现实模型的指导，类似于在因果推理任务中使用的模型，本文已经描述了其中的一些认知任务 ，并展示了它们是如何在结构因果模型的框架内运行且体现了基于模型的方法对认知任务是

至关重要的。本文提出了结构因果模型的七个任务，这七个任务是目前的机器学习系统无法达到的，使用因果建模工具和科学的数据可以达到帮助实现人类水平的强人工智能。

论文题目：*Network Density of States*

中文题目：状态网络密度

论文作者：Kun Dong, Austin R. Benson, and David Bindel.

论文出处：In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'19).

论文地址：https://dl.acm.org/citation.cfm?id=3330891

研究问题：

本论文旨在解决真实世界图形光谱内部存在的难以计算和解释的问题。与几何学不同，通过频谱密度的整体分布来研究图形在很大程度上限于简单的随机图形模型。真实世界图谱的内部仍未开发，难以计算和解释。本文旨在深入研究了现实世界图谱密度的核心，并展示频谱密度的估计如何促进许多常见中心度度量的计算，使用频谱密度来估计关于图结构的有意义的信息。

研究方法：

态密度在理解凝聚态物理学中的电子能带结构中起着重要作用，以往已经有文献提出了几种方法来估计光谱密度。本文采用了其中两种方法：核多项式方法（KPM），其中涉及 DOS / LDOS 的多项式展开，以及通过 Lanczos 迭代实现的高斯正交（GQL）。这些方法由 Cohen-Steiner 等人在很早以前提出，但尚未应用在网络环境中。本文在此基础上，提出了一种用于 LDOS 的新的直接嵌套剖析方法（ND），以及针对图形的特定修改方法以改进 KPM 和 GQL 方法的收敛性。之后，本文由提出了主题滤波方法（Motif Filtering），来分离频谱中尖峰成分，从而使用更少的切比雪夫矩来获得更准确的近似值。如图所示

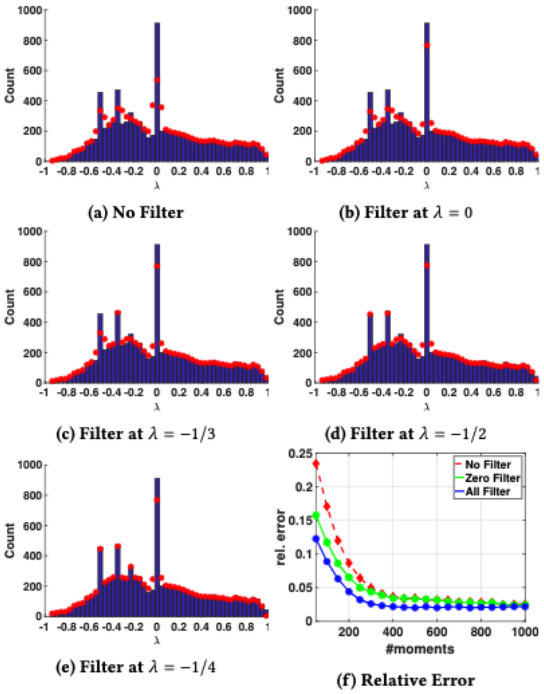


Figure 3: The improvement in accuracy of the spectral histogram approximation on the normalized adjacency matrix for the High Energy Physics Theory (HepTh) Collaboration Network, as we sweep through spectrum and filter out motifs. The graph has 8638 nodes and 24816 edges. Blue bars are the real spectrum, and red points are the approximated heights. (3a-3e) use 100 moments and 20 probe vectors. (3f) shows the relative L_1 error of the spectral histogram when using no filter, filter at $\lambda = 0$, and all filters.

研究结果：

本文使频谱密度的计算成为分析大型实际网络的实用工具，具体方法借鉴了凝聚态物理学中的方法，但是通过对图形图案的特殊处理保留了独特的光谱指纹，这些调整可改善网络分析设置中的性能。本文证明了光谱密度对于图中的小变化都是稳定的，并提供了对方法中近似误差的分析。实验通过仅使用一个计算节点来处理具有数千万个节点和数十亿条边的图形来验证了方法的效率。

论文题目：*Real-time Personalization using Embeddings for Search Ranking at Airbnb*

中文题目：利用嵌入式表示的 Airbnb 实时个性化搜索排序

论文作者：Mihajlo Grbovic, Haibin Cheng

论文出处：In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'18).

论文地址：https://dl.acm.org/citation.cfm?id=3219885

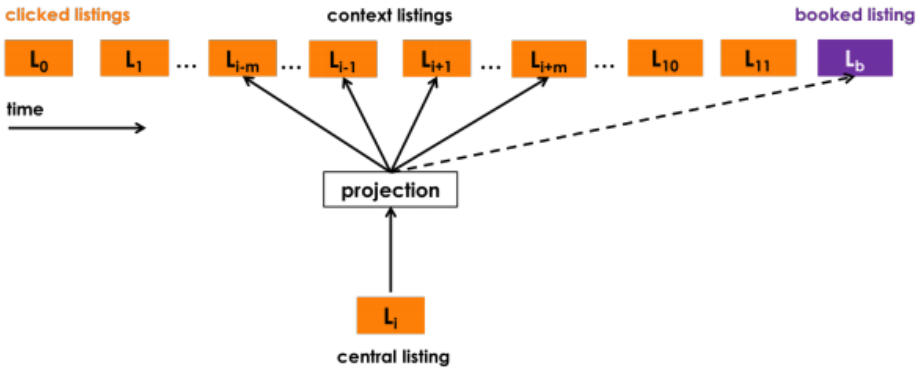
研究问题：

Airbnb 是全世界最大的房屋短租网站，提供了一个连接房主和租客的中介平台，因此也需要一个高水准的搜索与推荐服务来支持其使用。系统中房主和租客的交互方式包括：租客点击某个房屋商品、租客预订某个房屋商品、房主（可能）拒绝某个预订请求。基于上述几种交互数据，本文提出了一个实时的排序模型来捕捉用户的短时兴趣以及长期兴趣，从而达到一个更好的搜索效果。

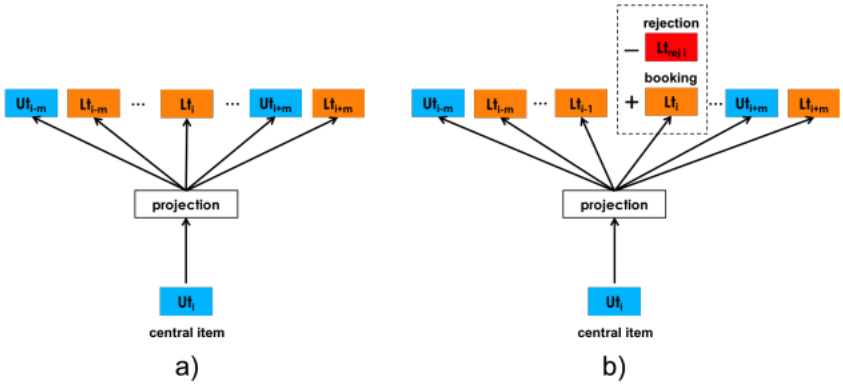
研究方法：

文章认为，用户的搜索目标分为短时兴趣和长时期兴趣两部分，其中短时兴趣代表用户因为当前的需求所导向的商品倾向，而长时期兴趣则表示用户选择商品的习惯，为了更好地捕获这两部分信息，本文通过两种方式生成了不同的嵌入式表示来进行分别建模。

通过用户点击数据生成 listing 的嵌入表示，旨在进行 listing 的相似推荐，以及对用户会话内的实时个性化推荐，主要亮点是在训练时增加了一个确定的全局上下文窗口。



通过预定信息生成 user-type 和 listing-type 的嵌入表示，目的是捕捉不同 user-type 的长期喜好。由于预定信息过于稀疏，Airbnb 对同属性的 user 和 listing 进行了聚合，另一个亮点是提出利用用户的负反馈（即房主的拒绝行为）来修正嵌入表示的训练。



研究结果：

本文通过两部分的实验表明了其有效性。

其一是提出了一种离线的推荐系统测试方法，用本文训练出的嵌入表示来对商品进行排序，用这个排序结果与实际用户的点击和预订进行对比来实现离线评测。从这个角度看，本文提出的嵌入表示方法显著优于已有的方法。

其二是给出了一些应用了本文嵌入表示方法得出的相似商品及用户习惯的案例分析，并提供了工具，通过案例观察可以发现，本文给出的相似商品不仅体现在价格等已知条目中，也体现了商品的风格信息，这表明即使不利用图片，本文给出的方法也可以挖掘出相似风格的商品。

总的来说，本文是一篇实用性很强的文章，不仅从实际角度出发，对用户的短期兴趣和长期偏好进行了建模，而且针对推荐系统嵌入表示训练中的冷启动问题等都进行了很有效的处理，并提出了一些针对操作系统的离线测试手段，是很有效的将理论应用到工程中的工作。

13.5 数据挖掘进展

近几年，我们已经迎来了大数据时代，各大互联网企业每天都在产生数以亿计的数据。各类数据往往都隐含着一些有价值的信息，如果人们手动地进行数据分析，往往需要耗费大量的时间。同时，大量未经处理的数据可能会被人们所忽视。数据挖掘就是想自动地从大规模的数据中挖掘出有意义的知识或者模式。这里，我们将数据挖掘领域近期的主要发展归为两大类：复杂数据挖掘、分布式数据挖掘。

复杂数据包括序列数据、图数据等。在序列数据挖掘中，基于注意力（Attention）机制的 Transformer 模型表现出了巨大的潜力，在机器翻译等任务上取得了非常好的效果。随后，BERT 模型使用双向 Transformer 通过预训练方式在各种自然语言处理的任务上都达到了当时最好的结果。在图数据挖掘研究中，网络表示学习仍然是近年来非常热门的话题。从 DeepWalk 算法开始，基于随机游走的算法在无监督的表示学习任务中表现良好。NetMF 算法将几种基于随机游走的算法统一写成了矩阵分解的形式，给网络表示学习算法提供了理论基础。图卷积神经网络是另一种处理图数据的有效方法，借鉴了图谱论中的图卷积并使用图的拉普拉斯矩阵，在半监督的节点分类任务和图分类任务中都表现出很好的效果。除此之外，异构网络的表示与挖掘也逐渐被大家所关注。

分布式数据挖掘已成为数据挖掘领域非常有前途的方向。随着数据挖掘计算成本的增加和数据隐私保护的问题，分布式数据挖掘开始备受关注。分布式数据挖掘利用分布式站点的资源来降低计算成本并增强数据保密性。由于分布式数据挖掘采用了不同的计算方式，传统的数据挖掘技术很难直接应用于分布式数据挖掘。目前，数据安全与数据隐私开始被大家所关注。2018 年 5 月，通用数据保护条例（GDPR）在欧盟正式生效，这也使得基于隐私保护的分布式数据挖掘方法逐渐被研究者所重视。

数据挖掘已经被广泛地应用于各类实际问题，包括金融数据分析、推荐系统等。数据挖掘相关研究需要结合实际问题，注重与机器学习、统计学科等的交叉，从大数据中挖掘出有价值的信息。

14 信息检索与推荐

14.1 信息检索与推荐概念

- 信息检索

R.Baeza-Yates 教授在其著作《现代信息检索中》中指出，信息检索（Information Retrieval，IR）是计算机科学的一大领域，主要研究如何为用户访问他们感兴趣的信息提供各种便利的手段，即：信息检索涉及对文档、网页、联机目录、结构化和半结构化记录及多媒体对象等信息的表示、存储、组织和访问，信息的表示和组织必须便于用户访问他们感兴趣的信息[77]。

在范围上，信息检索的发展已经远超出了其早期目标，即对文档进行索引并从中寻找有用的文档。如今，信息检索的研究包括用户建模、Web 搜索、文本分析、系统构架、用户界面、数据可视化、过滤和语言处理等技术。

信息检索的主要环节包括信息内容分析与编码、组成有序的信息集合以及用户提问处理和检索输出。其中信息提问与信息集合的匹配、选择是整个环节中的重要部分。当用户向系统输入查询时，信息检索过程开始，接着用户查询与数据库信息进行匹配。返回的结果可能是匹配或不匹配查询，而且结果通常被排名。大多数信息检索系统对数据库中的每个对象与查询匹配的程度计算数值分数，并根据此值进行排名，然后向用户显示排名靠前的对象，信息检索框架如下图所示。

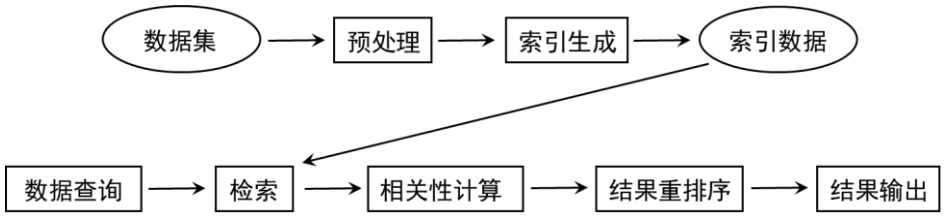


图 14-1 信息检索系统流程

- 推荐系统

推荐系统（Recommendation System，RS）是指信息过滤技术，从海量项目（项目是推荐系统所推荐内容的统称，包括商品、新闻、微博、音乐等产品及服

务）中找到用户感兴趣的部分并将其推荐给用户，这在用户没有明确需求或者项目数量过于巨大、凌乱时，能很好地为用户服务，解决信息过载问题[78]。

如下图所示，一般推荐系统模型流程通常由 3 个重要的模块组成：用户特征收集模块，用户行为建模与分析模块，推荐与排序模块。推荐系统通过用户特征收集模块收集用户的历史行为，并使用用户行为建模和分析模块构建合适的数学模型分析用户偏好，计算项目相似度等，最后通过推荐与排序模块计算用户感兴趣的项目，并将项目排序后推荐给用户[79]。

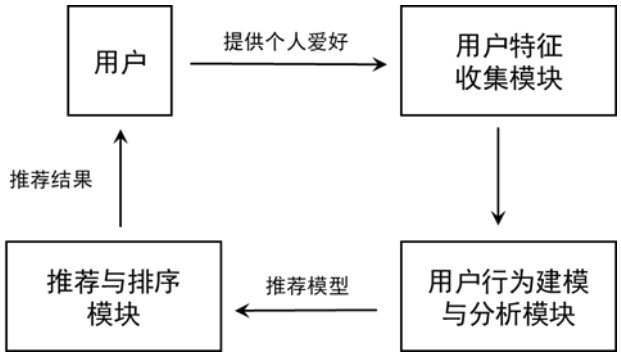


图 14-2 推荐系统模型流程

- 联系与区别

信息的检索与推荐都是用户获取信息的手段，无论是在互联网上，还是在线下的生活场景里，这两种方式都大量并存，两者之间的关系是互补的：搜索引擎需要用户主动提供准确的关键词来寻找信息，因此不能解决用户的很多其他需求，比如当用户无法找到准确描述自己需求的关键词时，搜索引擎就无能为力了。和搜索引擎一样，推荐系统也是一种帮助用户快速发现有用信息的工具。与搜索引擎不同的是，推荐系统不需要用户提供明确的需求，而是通过分析用户的历史行为给用户的兴趣建模，从而主动给用户推荐能够满足他们兴趣和需求的信息。因此，从某种意义上说，推荐系统和搜索引擎对于用户来说是两个互补的工具。搜索引擎满足了用户有明确目的时的主动查找需求，而推荐系统能够在用户没有明确目的的时候帮助他们发现感兴趣的新内容。在实际生活中也有很多运用。

同时，信息的检索与推荐也有着一定的区别，可以分为以下几个方面：

首先是主动与被动的不同。搜索是一个非常主动的行动，用户的需求也十分明确，在搜索引擎提供的结果里，用户也能通过浏览和点击来明确的判断是否满足了用户需求。然而，推荐系统接受信息是被动的，需求也都是模糊而不明确的。

其次是个性化程度的高低。搜索引擎虽然也可以有一定程度的个性化，但是整体上个性化运作的空间是比较小的，因为当需求非常明确时，找到结果的好坏通常没有太多个性化的差异。但是推荐系统在个性化方面的运作空间要大很多，虽然推荐的种类有很多，但是个性化对于推荐系统是非常重要，以至于在很多时候大家索性就把推荐系统称为“个性化推荐”甚至“智能推荐”。

再次就是需求时间不同。在设计搜索排序算法里，需要想尽办法让最好的结果排在最前面，往往搜索引擎的前三条结果聚集了绝大多数的用户点击。简单来说，“好”的搜索算法是需要让用户获取信息的效率更高、停留时间更短。但是推荐恰恰相反，推荐算法和被推荐的内容往往是紧密结合在一起的，用户获取推荐结果的过程可以是持续的、长期的，衡量推荐系统是否足够好，往往要依据是否能让用户停留更多的时间，对用户兴趣的挖掘越深入，越“懂”用户，那么推荐的成功率越高，用户也越愿意留在产品里。

最后是评价方法不同。搜索引擎通常基于 Cranfield 评价体系，整体上是将优质结果尽可能排到搜索结果的最前面，让用户以最少的点击次数、最快的速度找到内容是评价的核心。而推荐系统的评价要宽泛很多，既可以用诸如 MAP（Mean Average Precision）的常见量化方法评价，也可以从业务角度进行侧面评价[80]。

14.2 信息检索和推荐技术发展历史

- **信息检索**

信息检索的目的是获取所需信息，而这要基于比较完善的检索技术，用户需求的变化和信息技术的进步对信息检索的发展有着重要的影响。根据技术的演化，我们将信息检索发展历程分为三个阶段：

（1）数字图书馆 / 文档电子化时代

1954 年，Vannevar Bush（范内瓦·布什）在“Atlantic Monthly”7 月号发表了一篇名为“As We May Think”的文章，这篇文章影响了几代的计算机科学家。文章提到：“未来人们能够实现对海量图书资源（1M）进行快速的访问”。概括出了信息检索在数字图书馆时代的特征，即对文档全文内容的快速检索。

范内瓦·布什在担任美国科学研究与发展办公室主任期间推进了美国军队研究机构与高校研究机构的合作，正是当时在这种合作关系中发挥最重要影响的三所大学（哈佛大学、麻省理工学院、加州大学伯克利分校）与后来成立的美国国防部高等研究计划署（ARPA）合作开发出了互联网的雏形：ARPANET。



图 14-3 范内瓦·布什（1890-1974）

1957 年，Luhn 在论文“*A Statistical Approach to Mechanized Encoding and Searching of Literary Information*”里提到“*…a writer chooses that level of subject specificity and that combination of words which he feels will convey the most meaning.*”，这是一种以单词作为索引单元的文档检索方法。

20 世纪 60 年代，Gerard Salton 创造了信息检索系统 SMART（Salton’s Magic Automatic Retrieval of Text），推进了信息检索相关研究的水平提升。SMART 系统并非搜索引擎，但它具备搜索引擎所具有的文本索引、查询处理、结果排序等功能。

20 世纪 60 年代后期另外两个研究领域需要提及。第一个是 Julie Beth Lovins 于 1968 年在麻省理工学院开发的词干算法（Stemming Algorithm）；另一个研究涉及评估指标，例如 William Cooper 在 1968 年提出的“Cooper”，这个度量标准目前已在多个应用程序中大量使用。

在数字图书馆时代，信息检索技术主要应用于封闭数据集合、单机模式或专网内的主机-终点模式，在商业应用方面，则是提供软件/解决方案，专网内的查询服务。

（2）早期互联网时代

随着信息技术的爆炸式发展，信息检索的发展发生了质的飞跃。Tim Berners-Lee（蒂姆·伯纳斯·李）基于尚未被商用的互联网提出了万维网（Web）的原型建议。1991 年 8 月，蒂姆·伯纳斯·李在一台 NeXT 电脑上建立了第一个网站 http://nxoc01.cern.ch/。他一直坚持将公开和开放作为万维网的灵魂。



图 14-4 蒂姆·伯纳斯·李和他的 NeXT 电脑

从事检索业务的公司随着互联网的发展而快速崛起，如雅虎、百度等。在众多公司中，谷歌被公认为全球最大的搜索引擎公司，其业务包括互联网搜索、云计算、广告技术，开发并提供大量基于互联网的产品与服务。随着互联网的发展，面对众多杂乱无章的信息，如何对数以亿计的相关网页进行排序成为搜索引擎算法的核心问题，为此谷歌开发出了著名的 PageRank 算法。

PageRank 的主要原理是用链接数量作为搜索排序的一个因子。在互联网上，如果一个网页被很多其他网页所链接，说明它受到普遍的承认和信赖，那么它的排名就高，这就是 PageRank 的核心思想。PageRank 算法将互联网中大多数的网页通过基于链接来计算网页质量的方式进行排名，为搜索引擎用户提供较好的基于链接查询的搜索结果，同时该算法能够进行离线分析处理，大大缩短了搜索引擎用户的服务响应时间，实属计算机科学史上一项伟大成就，它以及其简明的逻辑，发明了迄今为止在搜索引擎领域还相当有代表性的算法，解决了数以亿计的

网页质量评估问题，抛开它难以估量的商业价值不谈，就说其学术方面，这种依靠数据自身结构的学习方法，也依然还在当前很多信息检索领域启发着我们。

业界主要表现为第一代搜索引擎和第二代搜索引擎的出现，国外有 AltaVista、Excite、WebCrawler 和 Yahoo！，国内有应用于国防和安全领域的“天罗”和面向公众提供服务的天网。第二代搜索引擎的代表是 1998 年成立的 Google 和 2000 年 1 月创建的中文搜索引擎——百度。在百度之后，多家中文搜索引擎相继出现，例如中搜、搜狗、搜搜和有道。

这个时期信息检索的应用形态的特征是开放的、大规模的、实时的、多媒体的，尤其巨型搜索引擎采集到的公开数据和用户访问日志等非公开数据深刻地影响着这一时期信息检索领域的创新模式。

（3）Web 2.0 时代

在 Web 2.0 时代，用户对 Web 有更深入的参与需求，这就对信息检索提出了更高的要求。信息搜索的发展开始更加关注用户需求，以实现内容与行为的精准 Web 搜索。

这个时期的信息检索实现了内容数据与社会各侧面的电子化数据（万维网、社交网、物联网、地理信息等）的全面融合；尤其是对社交网络数据的采集和大数据处理技术出现了社会化趋势。

- 信息推荐

上个世纪最后二十年以来，互联网的发展和普及为人们提供了一个全新的信息存储、加工、传递和使用的载体，网络信息也迅速成为了社会成员获取知识和信息的主要渠道之一。

一般认为推荐系统的研究始于 1994 年明尼苏达大学，Group Lens 研究组推出了 Group Lens 系统，该工作不仅首次提出了协同过滤的思想，并且为推荐问题建立了一个形式化的模型，为随后几十年推荐系统的发展带来了巨大影响。

之后，推荐系统的相关技术得到了进一步发展和重视。1995 年 3 月，卡耐基·梅隆大学的 Robert Armstrong 等人在美国人工智能协会提出了个性化导航系统 Web Watcher；斯坦福大学的 Marko Balabanovic 等人在同一会议上推出了个

性化推荐系统 LIRA；1997 年，AT&T 实验室提出了基于协同过滤的个性化推荐系统 PHOAKS 和 Referral Web；2000 年，NEC 研究院的 Kurt 等人为搜索引擎 Cite Seer 增加了个性化推荐功能；2003 年，Google 开创了 AdWords 盈利模式，通过用户搜索的关键词来提供相关的广告。2007 年开始，Google 为 AdWords 添加了个性化元素，不仅仅关注单词搜索的关键词，而且对用户一段时间内的推荐历史进行记录和分析，据此了解用户的喜好和需求，更为精确地呈现相关的广告内容。

信息推荐系统的演变始终伴随着网络的发展，第一代信息推荐系统使用传统网站从以下三个来源收集信息：来自购买或使用过的产品的基础内容数据；用户记录中收集的人口统计数据；以及从用户的项目偏好中收集的基于记忆的数据。第二代推荐系统通过收集社交信息，例如朋友、关注者、粉丝等。第三代推荐系统使用网上集成设备提供的信息。

14.3 人才概况

- 全球人才分布

学者地图用于描述特定领域学者的分布情况，对于进行学者调查、分析各地区竞争力现况尤为重要，下图为信息检索与推荐领域全球学者分布情况：



图 14-5 信息检索与推荐领域全球人才分布

地图根据学者当前就职机构地理位置进行绘制，其中颜色越深表示学者越集中。从该地图可以看出，美国的人才数量优势明显且主要分布于其东西海岸；欧

洲、亚洲也有较多的人才分布；其他诸如非洲、南美洲等地区的学者非常稀少；信息检索与推荐领域的人才分布与各地区的科技、经济实力情况大体一致。此外，在性别比例方面，信息检索与推荐领域中男性学者占比 90.6%，女性学者占比 9.4%，男性学者占比远高于女性学者。

信息检索与推荐领域学者的 h-index 分布如下图所示，大部分学者的 h-index 分布在中低区域，其中 h-index 小于 20 区间的人数最多，有 870 人，占比 42.8%，50-60 区间的人数最少，有 82 人。

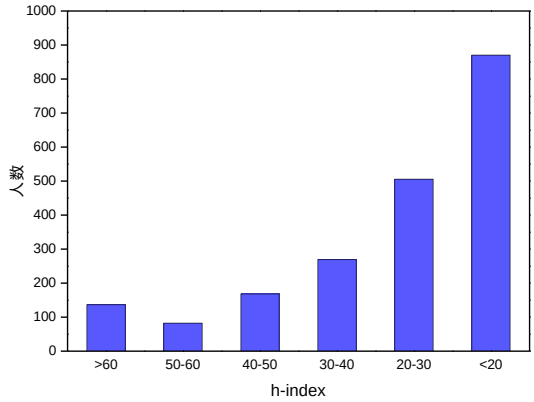


图 14-6 信息检索与推荐领域学者 h-index 分布

- 中国人才分布



图 14-7 信息检索与推荐领域中国学者分布

我国专家学者在信息检索与推荐领域的分布如上图所示，从中可以发现京津地区在本领域的人才数量最多，其次是长三角和珠三角地区，相比之下，内陆地区的人才较为匮乏，这种分布与区位因素和经济水平情况不无关系。同时，通过观察中国周边国家的学者数量情况，特别是与日韩、东南亚等地相比，中国在信息检索与推荐领域学者数量较多且优势较大。

中国与其他国家在信息检索与推荐领域的合作情况可以根据 AMiner 数据平台分析得到，通过统计论文中作者的单位信息，将作者映射到各个国家中，进而统计中国与各国之间合作论文的数量，并按照合作论文发表数量从高到低进行了排序，如下表所示。

表 14-1 信息检索与推荐领域中国与各国合作论文情况

合作国家	论文数	引用数	平均引用数	学者数
中国-美国	204	6858	34	443
中国-新加坡	49	1067	22	84
中国-英国	33	1057	32	44
中国-澳大利亚	32	746	23	63
中国-荷兰	22	502	23	20
中国-加拿大	19	514	27	42
中国-日本	7	63	9	17
中国-印度	4	58	15	10
中国-希腊	4	47	12	7
中国-德国	3	34	11	11

从上表数据可以看出，中美合作的论文数、引用数、平均引用数以及学者数遥遥领先，表明中美间在信息检索与推荐领域合作之密切；此外，中国与欧洲的合作非常广泛，前 10 名合作关系里中欧合作共占 4 席。

14.4 论文解读

本节对本领域的高水平学术会议及期刊论文进行挖掘，解读这些会议和期刊在 2018-2019 年的部分代表性工作。这些会议和期刊包括：

International ACM SIGIR Conference on Research and Development in Information Retrieval

ACM Transactions on Information Systems

ACM Recommender Systems

我们对本领域论文的关键词进行分析，统计出词频 Top20 的关键词，生成本领域研究热点的词云图，如下图所示。其中，推荐（recommendation）、检索（retrieval）、排序学习（learning to rank）是本领域中最热的关键词。



论文题目：*Adversarial Personalized Ranking for Recommendation*

中文题目：对抗式个性化推荐排名

论文作者：Xiangnan He, Zhankui He, Xiaoyu Du anTat-Seng Chua.

论文出处：The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

论文地址：https://arxiv.org/pdf/1808.03908.pdf

研究问题：

贝叶斯个性化排名（Bayesian Personalized Ranking, BPR）是一种成对学习的排序方法，用于优化个性化排序的推荐模型。它以内隐反馈学习为目标，假定观察到的交互比未观察到的交互排在更高的位置。矩阵因子分解（Matrix Factorization, MF）是最基本也是最有效的推荐模型。MF 将每个用户和项表示为嵌入向量，通过嵌入向量之间的内积来估计用户对某一项的偏好程度。在信息检索领域，贝叶斯个性化排名训练的矩阵分解模型（MF-BPR）学习一个与训练数据相适应的复杂函数，不能很好地泛化，且其鲁棒性较差，易受参数的对抗性扰

动。因此本文提出了一种新的个性化排名训练方法-对抗的个性化排名（Adversarial Personalized Ranking, APR）。

研究方法：

以 BPR 为基础，APR 中引入一个额外的目标函数，对其进行优化来量化推荐模型在参数扰动下的损失，使推荐模型既适合于个性化排序，又具有对抗性扰动的鲁棒性。

$$L_{BPR}(\mathcal{D}|\Theta) = \sum_{(u,i,j)\in\mathcal{D}} -\ln\sigma(\hat{y}_{ui}(\Theta) - \hat{y}_{uj}(\Theta)) + \lambda_\Theta||\Theta||^2,$$

$$L_{APR}(\mathcal{D}|\Theta) = L_{BPR}(\mathcal{D}|\Theta) + \lambda L_{BPR}(\mathcal{D}|\Theta + \Delta_{adv}),$$

$$\text{where} \quad \Delta_{adv} = \arg\max_{\Delta,||\Delta||\leq\epsilon} L_{BPR}(\mathcal{D}|\hat{\Theta} + \Delta),$$

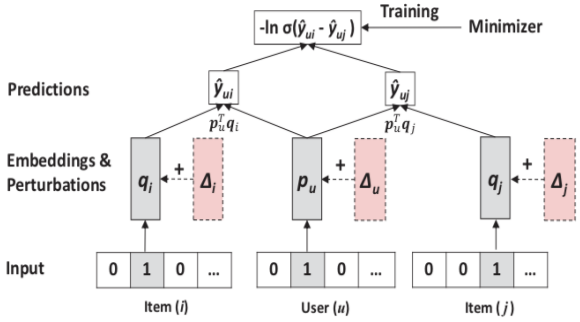
上式和下式分别是 BPR 和 APR 的目标函数，Δ_{adv} 是对抗性扰动，旨在最大化 BPR 目标函数的扰动。APR 可以看作是在玩一个极小极大的游戏，在这个游戏中，优化扰动使 BPR 损失最大化，并且在对抗扰动的情况下训练模型使 BPR 损失和附加损失最小化。APR 指定了一个与模型无关的通用学习框架，只要底层模型是可微的，就可以在 APR 框架下使用反向传播或者基于梯度的优化算法训练模型。

具体地，由于 APR 的目标函数中含有非线性函数，且训练实例数目庞大，故使用随机梯度下降法（Stochastic Gradient Descent, SGD）对 APR 进行优化。SGD 的思想是随机选择一个训练实例，并只针对单个实例更新模型参数，因此如何根据一个随机采样实例（u, i, j）优化模型参数是关键。本文提出的求解框架包括对抗性扰动构建和模型参数学习两步，具体步骤详见算法 1。

Algorithm 1: SGD learning algorithm for APR.

Input: Training data $\mathcal{D}$, adversarial noise level ϵ, adversarial regularizer λ, L_2 regularizer λ_Θ, learning rate η;
Output: Model parameters Θ;

```
Initialize Θ from BPR ;
while Stopping criteria is not met do
    Randomly draw (u, i, j) from D ;
    // Constructing adversarial perturbations
    Δ_adv ← Equation (8) ;
    // Updating model parameters
    Θ ← Equation (11) ;
end
return Θ
```



为了说明 APR 是如何工作的，本文提出了一个基于 MF 的推荐解决方案。首先用 BPR 训练 MF，然后在 APR 框架下进一步优化它，因此把这种方法称为对抗性矩阵分解（AMF）。AMF 如上图所示。由于 MF 的参数是用户和项的嵌入向量，故对嵌入向量加以对抗性扰动，再将算法 1 应用到 AMF 中，这需要对 AMF 进行小批量训练，直到 AMF 达到收敛状态或性能开始下降。

研究结果：

本文在 Yelp、Pinterest 和 Gowalla 三个公共数据集上进行了大量的实验，这三个数据集分别代表不同的应用场景。定量分析和定性分析都证明了对个性化排名进行对抗性训练的有效性和合理性。AMF 优于 MF-BPR，归一化折现累积增益（NDCG）和命中率（HR）平均提高了 11%，它也优于最近提出的推荐模型，成为最前沿的推荐模型。

论文题目：*Neural Compatibility Modeling with Attentive Knowledge Distillation*

中文题目：基于注意力知识蒸馏的神经兼容性建模

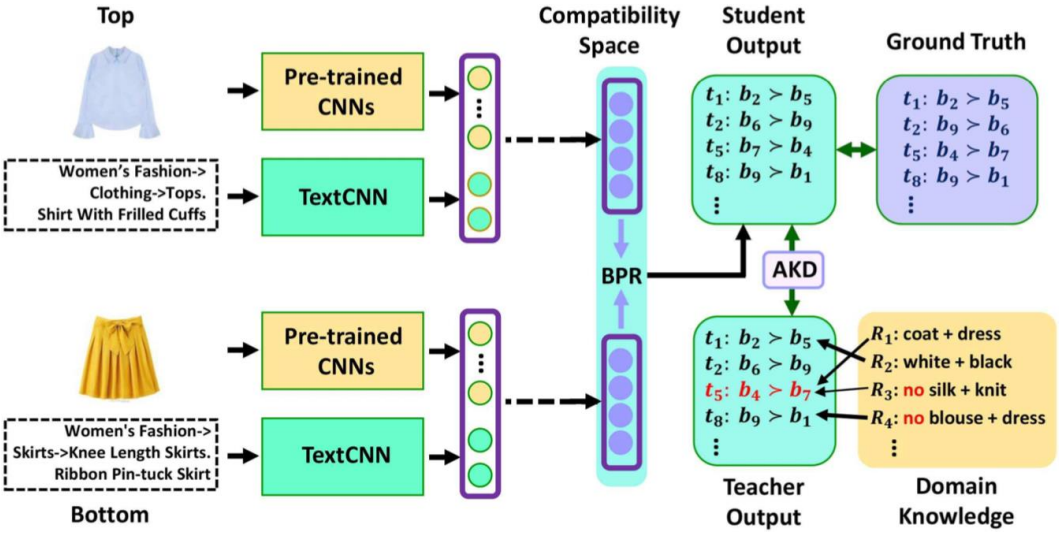
论文作者：Xuemeng Song, Fuli Feng, Xianjing Han, Xin Yang, Wei Liu and Liqiang Nie.

论文出处：The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

论文地址：https://arxiv.org/pdf/1805.00313.pdf

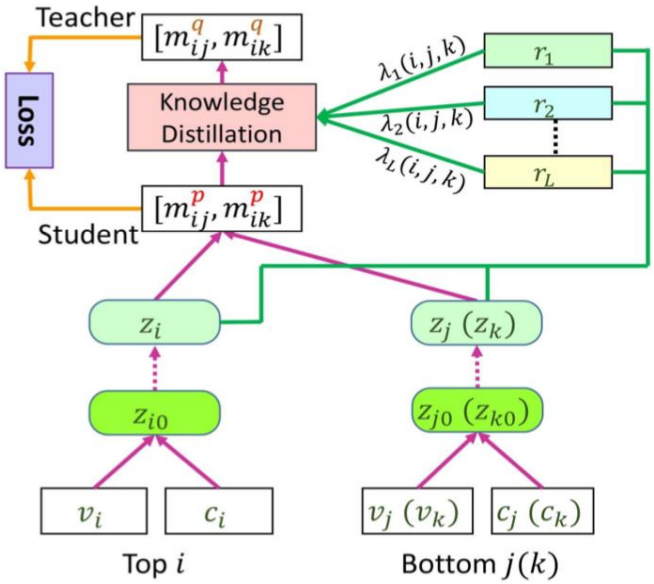
研究问题：

服装搭配与人们的日常生活息息相关，现有研究大多依赖深度神经网络来提取时尚单品的有效表征来解决服装搭配问题。但作为纯数据驱动方法的神经网络不仅具有较差的可解释性，而且也忽视了搭配领域知识。即使从数据驱动和知识驱动的角度对服装单品之间的兼容性进行全面建模也面临许多挑战，如搭配领域知识是无结构且模糊的，如何将搭配规则无缝地编码到纯数据驱动的学习框架中以及对于不同的样本知识规则可能表现出不同的置信度，从而提供不同的搭配指导。为此，本文提出了一种基于注意力知识蒸馏的神经兼容性建模方法（AKD-DBPR）。



研究方法：

AKD-DBPR 能够从特定数据样本和一般领域知识中学习，采用教师-学生模式来整合领域知识（教师）并提高神经网络（学生）的性能。其基本思想类似于人类教育，教师知道几个专业规则，因此教师可以用自己对特定问题的解决方案指导学生。



学生网络 p 作为一种纯数据驱动的神经网络，其目标是学习一个隐含的兼容空间，用双路径神经网络将来自异构空间的时尚单品统一起来。为了对不同模式之间的兼容性和语义关系全面建模，学生网络通过在视觉和语境表示的连接向量上添加隐含层，无缝地集成时尚单品的视觉和语境模式。此外，为了更好地描述时尚单品之间的相对兼容性，构建基于贝叶斯个性化排名（BPR）框架的学生网

络来研究互补单品之间的配对偏好。同时，用一组灵活的结构化逻辑规则对领域知识进行编码，并利用正则化器将这些知识规则编码到教师网络 q 中。但不同的规则对于不同的样本可能有不同的置信水平，因此引入注意力机制来分配规则置信度，进一步用于指导学生网络的训练。上图是注意力知识蒸馏的流程。v 和 c 分别表示单品的视觉和上下文语境，$m_{i,j}$表示上装 i 和下装 j 的兼容性，r 表示规则。最后，鼓励学生网络达到良好的兼容性建模性能，而且能很好地模拟规则正则化的教师网络。

研究结果：

在真实数据集上进行的大量实验证明了 AKD-DBPR 在服装搭配领域具有良好性能且能应用到互补时尚单品检索的实践中，除此之外也证明了引入注意机制有助于克服人为定义的模糊规则的局限性。

论文题目：*Improving Sequential Recommendation with Knowledge-Enhanced Memory Networks*

中文题目：基于知识增强记忆网络的序列推荐

论文作者：Jin Huang, Wayne Xin Zhao, Hong-Jian Dou, Ji-Rong Wen and Edward Y. Chang.

论文出处：The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

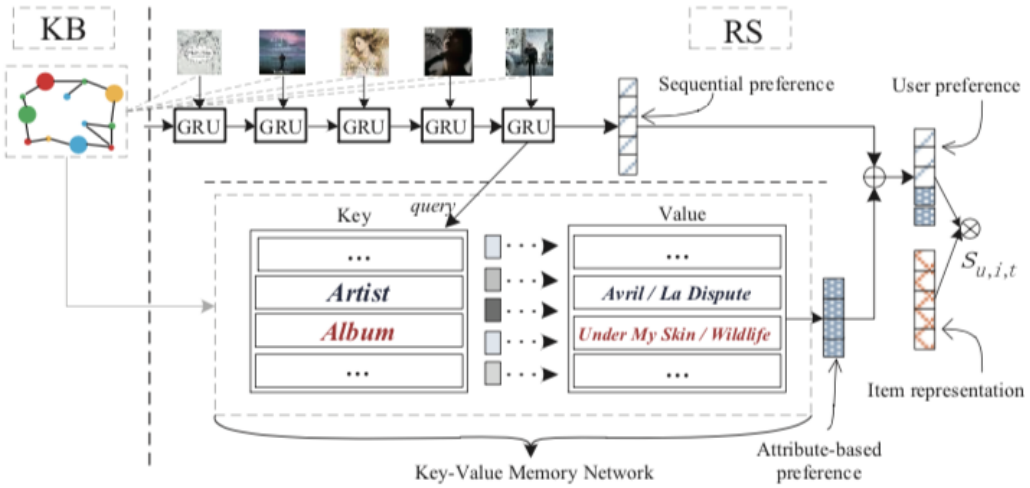
论文地址：https://sci-hub.tw/10.1145/3209978.3210017

研究问题：

推荐系统可以为用户推荐其感兴趣的内容并给出个性化的建议。基于 RNN 的网络可以将历史交互记录编码为隐藏状态向量，但是它很难从交互序列中捕获细粒度的用户偏好，而且隐向量表示的可解释性也较差。为了以一种可解释的方式增强细粒度用户偏好建模的能力，本文提出一种知识增强的序列推荐（KSR）模型。

研究方法：

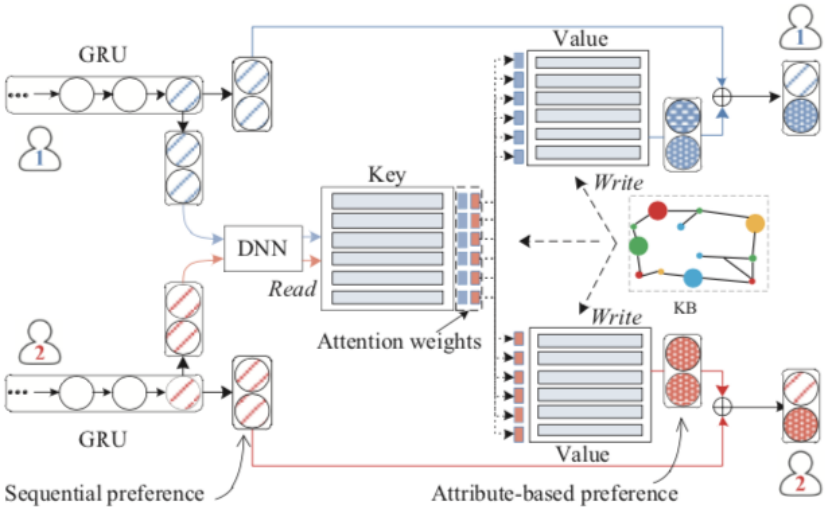
KSR 模型将基于 RNN 的网络（GRU）与键值记忆网络（KV-MNs）相结合来增强推荐系统的特征捕获能力与解释性。GRU 部分用于捕捉用户序列偏好特征，而 KV-MNs 用于捕捉基于属性的用户偏好特征。该模型如下图所示。



给定一组用户 u 的交互序列$\{i_1, \ldots, i_t\}^1$，在表示好 GRU 模型的隐藏层和用户 u 的序列偏好向量之后，对每个对象使用基于贝叶斯后验优化的个性化排序算法（BPR）进行预训练，再通过下式计算排名得分对候选对象排名，将得分最高的对象推荐给目标用户。

$$s_{u,i,t} = g(u,i,t) = \boldsymbol{h}_t^{u\top} \cdot \boldsymbol{q}_i,$$

KSR 模型的整体工作机制如上图所示。基于 RNN 的知识增强序列推荐模型的短期记忆能力有限，不适合长期存储知识库信息，因此通过 KV-MNs 来整合知识库知识。KV-MNs 将键向量设置为从知识数据中学习的嵌入关系，对应于实体属性。此外，对于给定的键向量设置一个特定于用户的值向量来存储相应属性的用户偏好特征。通过这种方式，外部知识库知识被有效地整合到 KV-MNs 中。准备好 KV-MNs 后，将其与基于 RNN 的序列推荐模型集成。在每次推荐时使用来自 RNNs 的序列偏好作为查询来读取特定于用户的 KV-MNs 值向量相关内容。值向量被组合到具有注意力权值的基于属性的偏好表示中，基于属性的偏好表示与序列偏好表示相结合形成用户偏好的最终表示。



研究结果：

本文在四个推荐系统数据集上进行实验，结果表明 KSR 模型在每个数据集的推荐性能都比表现最好的基准推荐模型有所提升，能够生成更高质量的序列推荐。此外，还定量分析了 KSR 模型的可解释性，结果表明该模型具有高度的可解释性。

论文题目：*Equity of Attention: Amortizing Individual Fairness in Rankings*

中文题目：注意力的公平性：在排名中平摊个体公平

论文作者：Asia J. Biega, Krishna P. Gummadi and Gerhard Weikum

论文出处：The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval - SIGIR '18

论文地址：https://arxiv.org/pdf/1805.01788.pdf

研究问题：

从招聘网站到共享经济平台，人员和对象的排名是选拔、配对和推荐系统的核心。由于排名位置会影响被排名对象受到的关注度，排名中的偏差会导致机会和资源的不公平分配。因此，本文提出了新的措施和机制来量化和减轻所有排名的不公平性。一个对象得到的关注应该与其相关性成正比并且注意力在排名中需公平分配。所提出的方法关注的是个体对象层面的公平，并将群体公平作为一个特例纳入其中，还设计了一种新的机制来确保排名中注意力平摊的公平性。

研究方法：

对象的关注受位置偏差的影响较大，即多个具有相似相关性的对象并没有得到相同的排名位置和相近的关注度。本文定义的考虑位置偏差的注意力公平（equity of attention）认为一个序列排名中每个对象获得的累积注意力（A）与其累积相关性（R）成比例。注意力公平是个体层次的，当个体的注意力达到公平时也会在群体层次上达到公平。注意力公平定义如下：

$$\frac{\sum_{l=1}^{m} a_{i1}^{l}}{\sum_{l=1}^{m} r_{i1}^{l}} = \frac{\sum_{l=1}^{m} a_{i2}^{l}}{\sum_{l=1}^{m} r_{i2}^{l}}, \forall u_{i1}, u_{i2}$$

下式将不公平度量为累积注意力和累积相关性之间的距离（$\rho^1, \ldots, \rho^m$是一个序列排名）：

$$unfairness(\rho^1, \ldots, \rho^m) = \sum_{i=1}^{n} |A_i - R_i| = \sum_{i=1}^{n} \left| \sum_{j=1}^{m} a_i^j - \sum_{j=1}^{m} r_i^j \right|$$

当相关程度较低的对象排名高于相关程度较高的对象排名时，为了满足公平标准而对排名进行过滤可能会导致质量损失。因此使用 IR 评估的度量指标来量化排名质量。解决公平性的方法是在原有排名的基础上重新排名，可将原始排名ρ作为参考来评估重新排名ρ^*的质量。即通过量化新排名ρ^*与原始排名 ρ 的差异来量化排名质量，如下式所示：

$$NDCG\text{-}quality@k(\rho, \rho^*) = \frac{DCG@k(\rho^*)}{DCG@k(\rho)}$$

为了提高公平性，对基于相关性的排名加以扰动，这会导致排名质量的下降，为解决此问题，可通过在质量约束下最小化不公平（即对可接受的排名质量设置下限）来对二者进行权衡。

将此方法转化为线上优化问题，排名摊销需要以线上方式完成，在不了解未来查询负载的情况下，就要对当前排名进行重新排序，从而在当前排名质量受到约束的情况下，将排名中累积注意力和相关性分布的不公平性降到最低。此线上优化问题可以通过整型线性规划（ILP）来解决。假设要在一系列排名中重新排列第 l 位，引入n^2个决策变量$X_{i,j}$，如果对象分配到排名位置 j，设置决策变量的

值为 1，否则为 0。对每个对象u_i，累积注意力和相关性分别初始化为$A_i^0 = 0$，$R_i^0 = 0$，ILP 定义如下：

$$\begin{aligned}
\text{minimize} \quad & \sum_{i=1}^{n}\sum_{j=1}^{n}|A_i^{l-1} + w_j - (R_i^{l-1} + r_i^l)| \cdot X_{i,j} \\
\text{subject to} \quad & \sum_{j=1}^{k}\sum_{i=1}^{n}\frac{2^{r_i^l} - 1}{log_2(j+1)}X_{i,j} \geq \theta \cdot IDCG@k \\
& X_{i,j} \in \{0,1\},\ \forall_{i,j} \\
& \sum_i X_{i,j} = 1,\ \forall_j \\
& \sum_j X_{i,j} = 1,\ \forall_i
\end{aligned}$$

第一项约束限制了排名质量损失，其他约束确保解是对象到排名位置的双向映射。当且仅当对象 u_i 映射到位置 j 时，$A_i^{l-1} + w_j$和$R_i^{l-1} + r_i^l$这两项分别更新累积注意力和相关性。还需注意的是 ILP 运行在一个巨大的组合空间中，所以又设计了过滤器来精简 ILP 的组合空间。

研究结果：

实验验证在两类数据集上进行，一是通过三个人工合成数据集来分析模型在不同相关性分布下（均匀分布、线性分布和指数分布）的性能，二是分析模型在真实场景中的性能，数据集采用世界不同地域三个城市的 Airbnb 公寓排名和基于 StackExchange 查询日志和文档集合构建的数据集。实验结果表明了提高注意力的公平性的重要性，在共享经济或排名影响人们经济生活的市场平台中尤为重要，而且这是可以在不牺牲排名质量的情况下做到的。

论文题目：*Impact of Item Consumption on Assessment of Recommendations in User Studies*

中文题目：关于用户是否消费过物品对推荐系统的评价影响研究

论文作者：Benedikt Loepp, Tim Donkers, Timm Kleemann, Jürgen Ziegler

论文出处：ACM Recommender Systems

论文地址：https://doi.org/10.1145/3240323.3240375

研究问题：

在对推荐系统的用户研究中，参与者通常不能消费推荐的商品。尽管如此，他们还是被要求通过问卷的方式来评估推荐质量和与用户体验相关的其他方面。然而，如果没有听过推荐的歌曲或看过推荐的电影，这可能是一个很容易出错的任务，可能会限制这些研究结果的有效性。在这篇文章中，作者调查了实际消费的推荐项目的影响。

研究方法：

作者提出了两项在不同领域进行的用户研究，研究表明在某些情况下，推荐评估和问卷调查结果存在差异。显然，在不允许用户消费项目的情况下，并不总能充分衡量出用户体验。另一方面，根据领域和提供的信息，参与者有时似乎很好地估计了推荐的实际价值。

一项是在音乐领域，用听歌做测试（记为 S），一项是在电影领域（记为 M）。通过对两组 40 名平均年龄为二十几岁的男女大致均衡的参与者，分别进行了音乐领域和电影领域的对照实验。控制的变量条件是消费前（前）和消费后（后）进行问卷调查，以及仅在消费后（后）进行问卷调查。

研究结果：

实验结果认为，我们有必要对推荐系统的问卷调查结果有所保留（即保持怀疑态度）。在某些情况下，如果不使用推荐的物品，参与者无法充分评估推荐系统的各个方面，特别是与用户体验相关的方面。

举例来说，比如音乐领域的参与者通常倾向于低估歌曲，并且在只能从包含描述性信息做判断，而不能够听歌时，对被要求在推荐列表中进行选择时满意度较低。听歌对满意度会产生正向影响，导致相关问卷项目的得分明显更高。比如在对选择满意度的研究中表明人们会倾向于高估过去事件的影响，并且在某些情况下，几周后满意度会下降，因此有必要调查这些结果的稳定性。 主观系统方面，如感知的召回质量的评价与是否消费过无关。对于电影而言，这在更多方面似乎是正确的，尤其是在可以使用高质量的文字说明的情况下，与具有抽象情感内容的音乐相比，电影的推荐更生动具体。确实一些为人熟知的歌曲也有电影的

这种效果，这会带来一些推荐的偏差。还有就是，短电影在这方面是完全未知的，被人耳熟能详的歌曲数量也不多，所以也没有产生出相关性。但对物品的熟悉程度及其对评估推荐的影响还是应该进行更详细的研究。另外，需要更多地考虑广泛使用问卷的一般影响，因为有意识地思考决策并不总是有益的（对于决策结果来说，结果越直观、简单越好）。

总的来说，定量结果表明，实际经验是否足以被替代，在很大程度上取决于推荐的领域，以及推荐系统提供的信息的类型和数量。本文通过研究定性评论反映了这一点。因此，作者建议避免不同设置之间的比较（注意控制变量），并在要求参与者在不消费的情况下对推荐进行评级时，要对用户实验更加注意才行（注重实践和对结果的把控）。还有，现实世界中提供的物品评分不会抑制评分，例如在看电影或烹饪食谱之前的评分，应该与看过后以及吃过后的评分谨慎的有机结合。只有当提供了足够信息的时候，参与者的反应才可能是可靠的，在这种情况下，可能就不需要（用户实际）消费过这个物品了。因为他们已经在心里形成了一个模型，考虑了自己的喜好，使这个建议像他已经真的买了或用了一样可靠（也就是自己说服自己）。

下图是作者针对条件（S1, S2/M1, M2）和时间点（事前，事后）交互作用的用户研究的混合模型结果。正差异结果表明用户消费后的结果会更好。

Table 1: Results of our mixed models for both user studies for interaction of condition (S1, S2/M1, M2) and point in time (Pre, Post). Positive differences indicate better results after consumption (*Choice Diff.*, *Effort* and *Doubts* are reversed accordingly).

Study 1	Interaction	S1-Pre vs. S1-Post			S1-Pre vs. S2-Post		
	Sig.	Est. Diff.	Std. Err.	Sig.	Est. Diff.	Std. Err.	Sig.
Perceived Rec. Quality [11]	.390	0.38	0.28	.183	0.15	0.29	.611
Mean Recommendation Rating	**.009***	**0.59**	**0.18**	**.004***	0.30	0.24	.226
Choice Satisfaction [11]	**.000***	**0.71**	**0.21**	**.003***	**1.29**	**0.28**	**.000***
Choice Difficulty [11]	**.001***	**1.14**	**0.29**	**.001***	0.55	0.38	.156
Effort [11]	.415	0.21	0.16	.196	0.10	0.23	.664
Effectiveness [11]	**.000***	**0.81**	**0.19**	**.000***	**1.08**	**0.33**	**.002***
Diversity [11]	.056	-0.38	0.26	.151	0.42	0.31	.184
Novelty [15]	.288	-0.19	0.13	.144	0.11	0.30	.731
Information Sufficiency [15]	**.000***	**1.48**	**0.38**	**.000***	**1.67**	**0.38**	**.000***
Transparency [15]	.104	0.48	0.22	.051	0.61	0.38	.113
Confidence and Trust [15]	**.017***	**0.54**	**0.20**	**.014***	**0.64**	**0.26**	**.020***
Doubts	**.000***	**2.19**	**0.33**	**.000***	**1.71**	**0.38**	**.000***
Overall Satisfaction [15]	**.005***	**0.62**	**0.20**	**.005***	**0.89**	**0.31**	**.007***

Study 2	Interaction	M1-Pre vs. M1-Post			M1-Pre vs. M2-Post		
	Sig.	Est. Diff.	Std. Err.	Sig.	Est. Diff.	Std. Err.	Sig.
Perceived Rec. Quality [11]	.467	-0.14	0.17	.411	-0.27	0.27	.328
Mean Recommendation Rating	.771	-0.08	0.14	.578	-0.11	0.21	.574
Choice Satisfaction [11]	**.020***	-0.19	0.25	.450	0.03	0.35	.937
Choice Difficulty [11]	.968	0.05	0.31	.877	-0.05	0.37	.905
Effort [11]	**.012***	-0.07	0.08	.383	**-0.47**	**0.15**	**.003***
Effectiveness [11]	.479	-0.14	0.22	.520	-0.41	0.34	.229
Diversity [11]	.117	0.24	0.19	.224	-0.37	0.34	.288
Novelty [15]	.218	0.14	0.09	.106	0.14	0.20	.472
Information Sufficiency [15]	**.041***	-0.33	0.23	.149	-0.37	0.32	.250
Transparency [15]	.763	-0.14	0.21	.499	-0.16	0.36	.658
Confidence and Trust [15]	.787	0.04	0.16	.826	-0.18	0.28	.527
Doubts	.680	-0.14	0.27	.605	-0.29	0.35	.407
Overall Satisfaction [15]	.442	-0.14	0.22	.525	-0.36	0.30	.235

论文题目：*Should I Follow the Crowd? A Probabilistic Analysis of the Effectiveness of Popularity in Recommender Systems*

中文题目：我应该跟随潮流（人群）吗？推荐系统中流行度有效性的概率分析

论文作者：Rocío Cañamares，Pablo Castells

论文出处：International ACM SIGIR Conference on Research and Development in Information Retrieval

论文地址：https://dl.acm.org/citation.cfm?id=3210014

研究问题：

近年来，在推荐系统的评价中使用 IR 方法已经成为一种普遍的做法。然而，人们发现 IR 指标对推荐热门物品的奖励算法存在强烈的偏差（bias），这与目前最先进的推荐算法存在的偏差是相同的。最近也有一些研究证实并测量了这种偏差，并提出了避免这种偏差的方法。

可是，最根本的问题仍悬而未决——一个物品的流行度（popularity）的偏差是否是我们应当避免的；无论这个偏差是一个对于推荐来说的好信号，还是可能因为受到实验偏差而导致的坏信号。

研究方法：

论文方法是根据关键随机变量之间的依赖关系，识别和建模可以确定答案的条件，包括物品评级、发现和相关性等。作者发现了保证流行有效或完全相反的条件，以及度量值反映真实有效性或在质量上偏离真实有效性的条件。

作者构建了一个众包数据集，其中没有公共可用数据显示的常见偏差，在这个数据集中，作者说明了在一个常见的有偏差的离线实验设置中所能测量的准确性与通过无偏差的观察所能测量的实际准确性之间的矛盾。

研究结论：

本文作者通过研究证实了普遍流行的有效性趋势，并用公式证明与解释了原因。同时作者还发现，在许多情况下表观的准确度可能会产生误导（即与真实准确度不匹配）。这大多是因为推荐的物品出现与用户兴趣相差较大的情况。作者发现，通常的实验（即观察到的准确度为度量值）可能对平均评级（average rating）及其个性化衍生品（personalized derivatives）相当不公平。与目前文献中观察到的结果相反，平均评级在大多数情况下，相对于正评级（positive ratings），在真实的准确率上可能更好、更安全，鲁棒性更强。

作者进一步发现，当涉及到流行度的有效性或平均评分及其度量时，喜欢的物品被评分的次数多少并不重要，重要的是评级是否依赖相关性，依赖是完全的还是部分的；当发现主要依赖于相关性，或者几乎不依赖于相关性时，流行度和

平均评分才是真正准确的理想条件。平均评级似乎比流行度对相关性独立性的影响更大，因此在有高偏差的情况下（如营销驱动），它比流行度更可取。

论文题目：*Variance Reduction in Gradient Exploration for Online Learning to Rank*

中文题目：一种基于方差缩减梯度搜索的在线学习排序方法

论文作者：Huazheng Wang, Sonwoo Kim, Eric McCord-Snook, Qingyun Wu, Hongning Wang

论文出处：International ACM SIGIR Conference on Research and Development in Information Retrieval

论文地址：https://dl.acm.org/citation.cfm?id=3331264

研究问题：

在线学习排名（OL2R）算法从用户的即时反馈中学习，算法的关键是对梯度的无偏估计，通常是通过从整个参数空间均匀采样来实现的。然而，这导致了梯度估计的高方差，会使模型更新时的效果不佳，特别是在参数空间的维数较大的时侯。本文旨在降低 OL2R 算法中梯度估计的方差。

研究方法：

在交叉测试之后，作者将选择的更新方向（如目标方向 winning direction）投影到当前查询下被检索（examined）文档的特征向量所跨越的空间（简称为“文档空间 document space”）。作者的主要观点是，交叉测试的结果完全由用户对所检索文档的相关性评估控制。因此，该测试引入的真实梯度只反映在构建的文档空间中，为了减少方差，我们可以安全地删除与文档空间正交的梯度分量。作者证明了这个投影梯度仍然是一个真实梯度的无偏估计，并且证明了这种低方差梯度估计能够显著减少 regret。

在本文中，作者提出并发展了文档空间投影（Document Space Projection，DSP）方法来减少梯度估计中的方差，提高在线学习的排序性能。DSP 的核心思想是认识到交叉测试只揭示了真实梯度在被检文档的跨空间上的投影。包含任何超出此空间的模型更新只会引入噪声。因此，作者将选择的模型更新方向投射回文档空间以减少方差。作者同时证明了 DSP 保持着一个无偏的梯度估计，并且

通过减少方差可以显著提高 DBGD 类型算法的 regret 界。通过大量的实验，发现 DSP 能够在方差减少和整体性能方面，特别是在排序特征数量较大的情况下，对几种最先进的 OL2R 模型提供统计上显著的改进。下表是在线和离线 NDCG@10，经过一万次查询，每种算法的文档空间投影的标准差和相对改进。

Table 2: Online NDCG@10, standard deviation and relative improvement of document space projection of each algorithm after 10,000 queries.

Click Model	Algorithm	MQ2007	MQ2008	MSLR-WEB10K	NP2003	Yahoo
Perfect	DBGD	679.3 (21.6)	847.1 (38.4)	532.2 (15.3)	1130.2 (43.3)	1165.5 (22.6)
	DBGD-DSP	689.1 (19.5)(+1.44%)	858.0 (39.2)(+1.29%)	553.6 (13.1)(+4.02%)	1198.8 (40.0) (+6.07%)	1198.8 (31.5)(+2.86%)
	MGD	689.1 (34.6)	859.4 (38.1)	558.3 (7.0)	1192.9 (44.6)	1201.9 (36.3)
	MGD-DSP	**757.3 (16.2)(+9.90%)**	**919.5 (42.2)(+6.99%)**	626.4 (9.6)(+12.20%)	1335.3 (39.1)(+11.94%)	**1309.4 (10.6) (+8.94%)**
	NSGD	684.4 (20.5)	867.5 (40.3)	589.5 (14.2)	1274.9 (47.4)	1162.3 (12.9)
	NSGD-DSP	732.5 (20.0)(+7.03%)	904.3 (38.6)(+4.24%)	**635.6 (12.8)(+7.82%)**	**1368.5 (41.1)(+7.34%)**	1270.1 (25)(+9.27%)
Navigational	DBGD	646.1 (23.4)	817.9 (45.5)	517.5 (20.9)	1062.3 (55.4)	1133.3 (40.8)
	DBGD-DSP	664.9 (26.9)(+2.91%)	830.3 (44.1)(+1.52%)	543.1 (14.8)(+4.95%)	1140.1 (52.5)(+7.32%)	1199.4 (34.6)(+5.83%)
	MGD	632.7 (35.5)	827.5 (35.5)	538.2 (7.2)	1115.4 (44.6)	1171.3 (20.4)
	MGD-DSP	**694.5 (15.7)(+9.77%)**	**882.3 (40.0)(+6.62%)**	586.9 (9.5)(+9.05%)	**1300.9 (39.6)(+16.63%)**	**1290.2 (15.3) (+10.15%)**
	NSGD	660.1 (24.5)	849.1 (36.6)	562.1 (18.8)	1211.1 (66.5)	1186.2 (36.8)
	NSGD-DSP	724.6 (24.5)(+9.77%)	895.8 (34.2)(+5.50%)	**608.3 (12.1) (+8.22%)**	1296.2 (24.3) (+7.03%)	1283.4 (7.2)(+8.19%)
Informational	DBGD	583.4 (46.0)	763.9 (55.1)	472.4 (34.6)	849.8 (144.5)	1107.3 (46.6)
	DBGD-DSP	620.1 (40.8)(+6.29%)	782.4 (51.8) (+2.42%)	522.1 (18.6) (+10.52%)	992.5 (81.1)(+16.79%)	1158.5 (22.6)(+4.62%)
	MGD	621.2 (38.2)	817.5 (45.3)	538.3 (10.8)	1107.9 (46.2)	1146.6 (37.5)
	MGD-DSP	**671.4 (18.9)(+8.08%)**	**865.9 (37.7)(+5.92%)**	580.5 (10.4)(+7.84%)	**1274.5 (12.9)(+15.04%)**	**1268.1 (16.4)(+10.60%)**
	NSGD	629.7 (25.3)	814.9 (37.1)	532.9 (15.2)	1123.5 (59.8)	1110.5 (20.9)
	NSGD-DSP	703.6 (29.2)(+11.74%)	871.3 (48.3)(+6.92%)	**597.9 (14.1)(+12.20%)**	1222.8 (43.8)(+9.03%)	1204.7 (9.6)(+8.48%)

Table 3: Offline NDCG@10, standard deviation and relative improvement of document space projection of each algorithm after 10,000 queries.

Click Model	Algorithm	MQ2007	MQ2008	MSLR-WEB10K	NP2003	Yahoo
Perfect	DBGD	0.484 (0.023)	0.683 (0.023)	0.331 (0.009)	0.737 (0.056)	0.688 (0.011)
	DBGD-DSP	0.480 (0.020) (-0.83%)	0.685 (0.024) (+0.29%)	0.333 (0.011) (+0.6%)	0.738 (0.039) (+0.14%)	0.681 (0.013) (-1.02%)
	MGD	0.495 (0.022)	0.691 (0.020)	0.334 (0.003)	0.746 (0.048)	0.715 (0.002)
	MGD-DSP	**0.501 (0.021)(+1.21%)**	**0.695 (0.022)(+0.58%)**	**0.409 (0.006)(+22.46%)**	0.748 (0.055)(+0.27%)	**0.725 (0.003)(+1.40%)**
	NSGD	0.488 (0.019)	0.689 (0.024)	0.397 (0.012)	0.743 (0.050)	0.691 (0.005)
	NSGD-DSP	0.491 (0.022)(+0.61%)	0.691 (0.029)(+0.29%)	**0.398 (0.008) (+0.25%)**	**0.750 (0.042) (+0.94%)**	0.717 (0.004)(+3.76%)
Navigational	DBGD	0.463 (0.028)	0.667 (0.021)	0.320 (0.012)	0.728 (0.054)	0.663 (0.020)
	DBGD-DSP	0.465 (0.024)(+0.43%)	0.668 (0.023)(+0.15%)	0.327 (0.011)(+2.19%)	0.734 (0.052)(+0.82%)	0.656 (0.013)(-1.06%)
	MGD	0.426 (0.019)	0.664 (0.016)	0.321 (0.003)	0.740 (0.048)	0.703 (0.010)
	MGD-DSP	0.467 (0.021)(+9.62%)	**0.684 (0.017)(+3.01%)**	0.331 (0.005)(+3.12%)	0.744 (0.051)(+0.54%)	**0.714 (0.006)(+1.56%)**
	NSGD	0.473 (0.022)	0.676 (0.024)	**0.389 (0.013)**	0.732 (0.053)	0.686 (0.008)
	NSGD-DSP	**0.478 (0.020)(+1.06%)**	**0.683 (0.026)(+1.04%)**	0.376 (0.014)(-3.34%)	**0.788 (0.006)(+7.65%)**	**0.711 (0.001)(+3.64%)**
Informational	DBGD	0.410 (0.034)	0.641 (0.031)	0.294 (0.022)	0.699 (0.063)	0.623 (0.017)
	DBGD-DSP	0.427 (0.027)(+4.15%)	0.632 (0.031)(-1.4%)	0.309 (0.011)(+32.65%)	0.692 (0.062)(-1.00%)	0.63 (0.030)(1.12%)
	MGD	0.406 (0.020)	0.651 (0.020)	0.317 (0.003)	0.726 (0.050)	0.668 (0.044)
	MGD-DSP	0.444 (0.025)(+0.44%)	0.669 (0.018)(+0.67%)	0.325 (0.004)(+0.33%)	0.738 (0.054)(+0.74%)	**0.701 (0.005)(+4.94%)**
	NSGD	**0.469 (0.018)**	**0.674 (0.023)**	**0.360 (0.013)**	0.733 (0.056)	0.663 (0.015)
	NSGD-DSP	0.466 (0.019)(-0.64%)	0.668 (0.026)(-0.89%)	0.340 (0.018)(-5.56%)	**0.789 (0.013)(+7.64%)**	0.685 (0.004)(+3.32%)

研究结果：

本文通过与几种最先进的 OL2R 算法的大量实验比较，验证了作者提出的方法在减少方差和提高整体排名性能方面的有效性（如表 2 和 3 所示）。但是观察到在模拟用户点击反馈时，DSP 在不同点击模式下的性能有所不同。未来，作者计划合并不同的点击模式解决方案，以更精确地构建文档空间。在交叉测试之前，研究如何进行基于文档空间的探索性方向生成也是有意义的。探索性方向预选有

望进一步加速梯度探索，提高在线学习过程中的用户满意度，不过同时也必须确保它是无偏差的。

14.5 信息检索与推荐进展

随着互联网中数字信息数量的增长，商品、书籍、新文章、歌、电影、研究文件等日常基础性事物，其数量和种类填满了多个数据仓库和数据库。蕴含着智能推荐系统和强大的搜索引擎的在线商店、在线音乐、在线视频和图片库等已成为人们快速寻找信息的主要方式。此类系统的流行程度和有用性在于它们能够便捷地显示几乎无限的物品信息。比如，Amazon、Netflix 等推荐系统尝试了解用户兴趣，并向用户推荐他们感兴趣的商品。尽管这些系统由于使用场景而各不相同，但其寻找用户感兴趣商品的核心机制都是用户兴趣与商品匹配的机制。

为了提高信息检索与推荐系统中算法模型的准确性和可解释性，研究人员近年来主要关注无偏的在线排序学习模型，以及利用知识信息增强推荐系统的表现和可解释性等方面的研究。其中，无偏的在线排序学习模型是指自动利用大规模用户点击数据训练搜索结果的排序模型。用户点击数据是现代搜索引擎的重要数据来源，具有成本低廉，并且对以用户为中心的检索应用程序（如搜索排名）特别有用等优点。为了充分利用用户点击数据开发一个无偏的学习排名系统，研究人员试图消除用户偏见对排名模型训练的影响。近年来，一种基于反事实学习和图形模型的无偏学习排名框架引起了人们的广泛关注。该框架侧重于使用反事实学习直接训练带有偏倚点击数据的排名模型。这个无偏的学习排名框架对待点击偏差作为一个反事实的影响和去偏用户反馈加权每点击与他们的反向倾向加权。它使用倾向性模型来量化点击的偏差，并没有明确地估计查询文档与培训数据的相关性。研究人员从理论上证明，在正确的偏差估计下，在该框架下使用点击数据训练的排序模型将收敛于使用真实相关信号训练的排序模型。

信息检索与推荐系统可以为用户推荐其感兴趣的内容并给出个性化的建议。而现在的推荐系统大都着眼于被推荐对象的序列建模，而忽略了它们细粒度的特征。为了解决以上问题，研究人员提出了多任务可解释推荐模型（Multi-Task Explainable Recommendation, MTER）和知识增强的序列推荐模型（Knowledge-enhanced Sequential Recommender, KSP）。其中，MTER 模型是一个用于可解释

推荐任务的多任务学习方法，通过联合张量分解将用户、产品、特征和观点短语映射到同一向量空间，来从用户评论中提取产品细粒度的个性化特征。KSR 模型提出了利用结合知识库的记忆网络来增强推荐系统的特征捕获能力与解释性，解决序列化推荐系统不具有解释性，且无法获取用户细粒度特征的不足。MTER 和 KSR 模型通过对推荐结果的解释，分析被推荐对象的特征，可以让用户可以对使用哪些推荐结果做出更明智，更准确的决策，从而提高他们的满意度。

近年来，信息检索与推荐领域比较流行的开源平台主要包括基于深度学习的检索模型（MatchZoo）、基于 tensorflow 的 learning to rank 模型（TF-Ranking）和 microsoft recommenders。其中，MatchZoo 是由中国科学院计算技术研究所网络数据科学与技术重点实验室近期发布的深度文本匹配开源项目。MatchZoo 是一个 Python 环境下基于 TensorFlow 开发的开源文本匹配工具，使用了 Keras 中的神经网络层，并有数据预处理，模型构建，训练与评测三大模块组成，旨在让大家更加直观地了解深度文本匹配模型的设计、更加便利地比较不同模型的性能差异、更加快捷地开发新型的深度匹配模型。MatchZoo 提供了基准数据集（TREC MQ 系列数据、WiKiQA 数据等）进行开发与测试，整合了当前最流行的深度文本匹配的方法（包括 DRMM, MatchPyramid, DUET, MVLSTM, aNMM, ARC-I, ARC-II, DSSM, CDSSM 等算法的统一实现），旨在为信息检索、数据挖掘、自然语言处理、机器学习等领域内的研究与开发人员提供便利，可以应用到的任务场景包括文本检索，自动问答，复述问题，对话系统等等。

TF-Ranking 是一个可扩展的基于 tensorflow 的用于排序的库，由 google 于 2018 年提出。TF-Ranking 提供了一个统一的框架，其中包括一套最先进的学习排序算法，并支持成对或列表损失函数、多项评分、排序度量优化和无偏学习排序。TF-Ranking 速度很快并且易于使用，可以创建高质量的排序模型。 统一的框架使机器学习的研究人员、实践者和爱好者能够在一个库中评估和选择一系列不同的排序模型。此外，这个开源库不仅提供了合理的默认模型，还可以让用户能够开发自己的定制模型，且提供了灵活的 API，用户可以在其中定义和插入自己定制的损失函数、评分函数和指标。Microsoft Recommenders 是微软云计算和人工智能开发团队与微软亚洲研究院团队深度合作，基于多年来各类大型企业级客户的项目经验以及最新学术研究成果，搭建的完整推荐系统的最新实操技巧开

源项目。该项目有效解决了定制和搭建企业级推荐系统中的几个难点，包括如何将学术研究成果或开源社区提供的范例适用于企业级应用、如何集成信息检索与推荐领域的学习指导资源俩协助从业人员深入理解并实际搭建完整推荐系统、如何选择最优算法以应对具体应用场景等。

为了协助信息检索与推荐领域的算法模型的训练和优化，微软公司提供了一个大规模支持机器阅读理解和问答系统等多种领域研究的数据集，简称 MS MACRO。该数据集从必应（bing）的搜索查询记录中取样，每个问题都有人工生成的答案和完全人工重写的答案。此外，数据集包含从通过 bing 检索的 web 文档中提取的百万个密码，这些密码提供了管理自然语言答案所需的信息。使用这个数据集，本文提出三个不同层次的难度不同的任务：（i）根据一组上下文段落预测一个问题是否可以回答，然后像人类一样提取和合成答案（ii）基于根据问题和段落语境信息可以被理解的上下文段落，来生成格式良好的答案（如果可能），最后（iii）根据给定的一个问题，对检索得到的段落进行排序。数据集的大小和问题来自真实用户搜索查询的事实，该数据集的规模和真实世界的性质使它对基准测试机器阅读理解和问答模型具有吸引力。

15 结束语

本报告依托于清华大学科技情报大数据挖掘与服务平台 AMiner 完成编制，对人工智能各领域进行了详细介绍。在介绍各领域概念及发展情况等内容的基础上，报告着重介绍了各领域人才分布情况，并对代表性论文进行了深度解读。

在人才情况方面，总的来看，美国的人才数量遥遥领先且主要分布在其东西海岸，独成第一梯队，凸显了其在人工智能领域的人才优势；欧洲及亚洲也有较多的人才分布，欧洲的人才主要集中在中西部，亚洲的人才主要集中于我国东部及日韩地区；其他诸如非洲、南美洲等地区的学者非常稀少。人才分布与各地区的科技、经济实力情况大体一致。对于我国而言，人才数量在大部分领域领跑第二梯队，但与位居首位的美国相比，中国高影响力学者数量明显不足，顶尖学者相对匮乏，中美之间还存在较大的赶超空间。我国高影响力学者主要集中在京津冀、东部、南部沿海地区及港台地区，相较而言，中西部地区人才分布较少，这种分布与国内区位因素和经济水平情况不无关系。从国际合作关系可以看出，在各领域中，中美合作论文数量均居首位，引用数及学者数也有明显优势，中美在人工智能领域的积极合作，有利于我国在人工智能前沿领域的学习发展；同时，我国也重视与欧洲、新加坡、日本等国的合作，博采众长。

在论文解读方面，报告对各领域的高水平学术会议及期刊论文进行挖掘，这些论文既包括近年高引论文、会议最佳论文，又有专家推荐的代表性工作。报告对各领域近年的代表性论文进行了详尽解读，解读面向前沿热点研究问题，深入探讨研究方法，展现最新研究成果。同时，报告对各领域论文的关键词进行了挖掘分析，统计出高频关键词，通过词云图反映研究焦点。

当前，人工智能已经成为引领新一轮科技革命和产业变革的战略性技术，我国在人工智能领域的科学技术研究和产业发展，起步稍晚于以美国为代表的发达国家，但是，在最近十余年的人工智能爆发发展期我国抓住了机遇，进入了快速发展阶段。在这个阶段，能够推动技术突破和创造性应用的高端人才对人工智能的发展起着至关重要的作用。本报告对人工智能各领域技术与人才情况等内容的挖掘分析，希望能对推动我国实施人工智能发展起到借鉴参考作用。

参考文献

[1] Samuel A L. Some studies in machine learning using the game of checkers[J]. IBM Journal of research and development, 1959, 3(3): 210-229.

[2] 机器学习简介（一）机器学习简史; http://www.studyai.com/article/ad75a319.

[3] 黄佳. 基于 OPENCV 的计算机视觉技术研究[D]. 华东理工大学, 2013.

[4] 计算机视觉入门系列（一）综; https://blog.csdn.net/wangss9566/article/details/54618507.

[5] 计算机视觉简介; https://blog.csdn.net/xiangz_csdn/article/details/78628521.

[6] 人工智能之知识图谱; https://www.aminer.cn/research_report/5c3d5a8709e961951592a49d?download=true&pathname=knowledgegraph.pdf.

[7] 自然语言处理研究报告; https://www.aminer.cn/research_report/5c35cdc55a237876dd7f127e?download=true&pathname=nlp.pdf.

[8] 高维深. 基于 HMM/ANN 混合模型的非特定人语音识别研究[D]. 电子科技大学, 2013.

[9] 陈硕. 深度学习神经网络在语音识别中的应用研究[D]. 华南理工大学,2013.

[10] 白琳. 基于语音识别的机器人控制技术研究[D]. 西南石油大学,2014.

[11] 计算机图形学研究报告; https://www.aminer.cn/research_report/5c2edcae81ecb9818a800700?download=true&pathname=cg.pdf.

[12] Doersch C, Gupta A, Efros A A. Unsupervised Visual Representation Learning by Context Prediction[C] // 2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, 2015.

[13] Noroozi M, Favaro P. Unsupervised Learning of Visual Representations by Solving Jigsaw Puzzles[J]. 2016.

[14] Pathak D, Krahenbuhl P, Donahue J, et al. Context Encoders: Feature Learning by Inpainting[J]. 2016.

[15] Richard Zhang, Phillip Isola, Alexei A Efros. Colorful Image Colorization[C] // European Conference on Computer Vision. Springer, Cham, 2016.

[16] Dosovitskiy A, Fischer P, Springenberg J, et al. Discriminative Unsupervised Feature Learning with Exemplar Convolutional Neural Networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015:1-1.

[17] Wu Z, Xiong Y, Yu S, et al. Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination[J]. 2018.

[18] Wu Z , Xiong Y , Yu S , et al. Unsupervised Feature Learning via Non-Parametric Instance-level Discrimination[J]. 2018.

[19] Zhuang C, Zhai A L, Yamins D. Local aggregation for unsupervised learning of visual embeddings[C] // Proceedings of the IEEE International Conference on Computer Vision. 2019: 6002-6012.

[20] He K, Fan H, Wu Y, et al. Momentum Contrast for Unsupervised Visual Representation Learning[J]. arXiv preprint arXiv:1911.05722, 2019.

[21] Bachman P, Hjelm R D, Buchwalter W. Learning representations by maximizing mutual information across views[J]. arXiv preprint arXiv:1906.00910, 2019.

[22] Oord A, Li Y, Vinyals O. Representation learning with contrastive predictive coding[J]. arXiv preprint arXiv:1807.03748, 2018.

[23] Tian Y, Krishnan D, Isola P. Contrastive Multiview Coding[J]. arXiv preprint arXiv:1906.05849, 2019.

[24] Kirillov A, He K, Girshick R, et al. Panoptic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 9404-9413.

[25] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C] // European conference on computer vision. Springer, Cham, 2014: 740-755.

[26] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3213-3223.

[27] Zhou B, Zhao H, Puig X, et al. Scene parsing through ade20k dataset[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 633-641.

[28] Neuhold G, Ollmann T, Rota Bulo S, et al. The mapillary vistas dataset for semantic understanding of street scenes[C] // Proceedings of the IEEE International Conference on Computer Vision. 2017: 4990-4999.

[29] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C] // Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[30] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2881-2890.

[31] de Geus D, Meletis P, Dubbelman G. Panoptic segmentation with a joint semantic and instance segmentation network[J]. arXiv preprint arXiv:1809.02110, 2018.

[32] Kirillov A, Girshick R, He K, et al. Panoptic feature pyramid networks[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 6399-6408.

[33] Li J, Raventos A, Bhargava A, et al. Learning to fuse things and stuff[J]. arXiv preprint arXiv:1812.01192, 2018.

[34] Li Y, Chen X, Zhu Z, et al. Attention-guided unified network for panoptic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 7026-7035.

[35] Liu H, Peng C, Yu C, et al. An end-to-end network for panoptic segmentation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 6172-6181.

[36] Xiong Y, Liao R, Zhao H, et al. Upsnet: A unified panoptic segmentation network[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 8818-8826.

[37] Hinton G E, Srivastava N, Krizhevsky A, et al. Improving neural networks by preventing co-adaptation of feature detectors[J]. arXiv preprint arXiv:1207.0580, 2012.

[38] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C] // Advances in neural information processing systems. 2012: 1097-1105.

[39] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, and Quoc V. Le. Mnasnet: Platform-aware neural architecture search for mobile. CoRR, abs/1807.11626, 2018.

[40] Bowen Baker, Otkrist Gupta, Nikhil Naik, and Ramesh Raskar. Designing neural network architectures using reinforcement learning. In5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings, 2017.

[41] Ye-Hoon Kim, Bhargava Reddy, Sojung Yun, and Chanwon Seo. Nemo : Neuro-evolution with multi objective optimization of deep neural network for speed and accuracy. In AutoML Workshop at ICML 2017, 2017.

[42] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions. In 6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Workshop Track Proceedings, 2018.

[43] Ekin Dogus Cubuk, Barret Zoph, Dandelion Man´e, Vijay Vasudevan, and Quoc V. Le. Autoaugment: Learning augmentation policies from data. CoRR, abs/1805.09501, 2018a.

[44] Zoph B, Vasudevan V, Shlens J, et al. Learning transferable architectures for scalable image recognition[C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8697-8710.

[45] Zoph B, Le Q V. Neural architecture search with reinforcement learning[J]. arXiv preprint arXiv:1611.01578, 2016.

[46] Zhong Z, Yan J, Wu W, et al. Practical block-wise neural network architecture generation[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 2423-2432.

[47] Real E, Moore S, Selle A, et al. Large-scale evolution of image classifiers[C] // Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017: 2902-2911.

[48] Hanxiao Liu, Karen Simonyan, Yiming Yang. DARTS: Differentiable Architecture Search. ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings, 2018.

[49] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C] // Advances in neural information processing systems. 2014: 2672-2680.

[50] Mirza M, Osindero S. Conditional generative adversarial nets. arXiv preprint arXiv: 1411.1784, 2014.

[51] Qi G J. Loss-sensitive generative adversarial networks on Lipschitz densities. arXiv preprint arXiv: 1701.06264, 2017.

[52] Miyato T, Kataoka T, Koyama M, Yoshida Y. Spectral normalization for generative adversarial networks. arXiv preprint arXiv: 1802.05957, 2018

[53] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv: 1511.06434, 2015.

[54] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks[J]. arXiv preprint arXiv:1805.08318, 2018.

[55] Karras T, Aila T, Laine S, Lehtinen J. Progressive Growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv: 1710.10196, 2017.

[56] 王昕. 关于媒体和多媒体的概念[J]. 现代电信科技, 1998(10): 44.

[57] 多媒体技术; https://wiki.mbalib.com/wiki/%E5%A4%9A%E5%AA%92%E4%BD%93%E6%8A%80%E6%9C%AF.

[58] 多媒体技术的发展简史; http://blog.sina.com.cn/s/blog_4cf2499101000al9.html.

[59] 袁保宗,阮秋琦,王延江,刘汝杰,唐晓芳.新一代（第四代）人机交互的概念框架特征及关键技术[J].电子学报,2003（S1）: 1945-1954.

[60] 徐菁.一种 CAD/CAM 平台虚拟交互模块的研究与开发[D].沈阳航空工业学院, 2009.

[61] 李兆堃. 基于 Kinect 体感技术的人机交互环境[J]. 数字技术与应用,2013（09）: 65-66.

[62] Minsky M.Society of mind[M]. Simon and Schuster, 1988.

[63] 任福继, 孙晓. 智能机器人的现状及发展[J]. 科技导报,2015,33（21）: 32-38.

[64] 张俊, 吴绍辉. 数据库技术的研究现状及发展趋势[J]. 工矿自动化, 2011, 37（07）: 34-36.

[65] 数据库的发展历史和当前主流技术和产品; http://www.omegaxyz.com/2018/03/26/brief_history_of_database/.

[66] 陈为, 沈则潜, 陶煜波等. 数据可视化[M]. 北京:电子工业出版社, 2013:2-124.

[67] Liu S, Cui W, Wu Y, et al. A survey on information visualization: recent advances and challenges[J]. The Visual Computer, 2014, 30(12):1373-1393.

[68] Satyanarayan A, Moritz D, Wongsuphasawat K, et al. Vega-lite: A grammar of interactive graphics[J]. IEEE transactions on visualization and computer graphics, 2016, 23(1): 341-350.

[69] Li J K, Ma K L. P5: Portable Progressive Parallel Processing Pipelines for Interactive Data Analysis and Visualization[J]. IEEE transactions on visualization and computer graphics, 2019.

[70] Segel E, Heer J. Narrative visualization: Telling stories with data[J]. IEEE transactions on visualization and computer graphics, 2010, 16(6): 1139-1148.

[71] Chen S, Li J, Andrienko G, et al. Supporting Story Synthesis: Bridging the Gap between Visual Analytics and Storytelling[J]. IEEE transactions on visualization and computer graphics, 2018.

[72] Wang Y, Sun Z, Zhang H, et al. DataShot: Automatic Generation of Fact Sheets from Tabular Data[J]. IEEE transactions on visualization and computer graphics, 2019.

[73] Cui W, Zhang X, Wang Y, et al. Text-to-Viz: Automatic Generation of Infographics from Proportion-Related Natural Language Statements[J]. IEEE transactions on visualization and computer graphics, 2019.

[74] Strobelt H, Gehrmann S, Pfister H, et al. Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks[J]. IEEE transactions on visualization and computer graphics, 2017, 24(1): 667-676.

[75] Han J.Data Mining: Concepts and Techniques[M]. Morgan Kaufmann Publishers Inc.2005.

[76] 琚会婧.数据挖掘技术在客户关系管理（CRM）中的应用研究[D]. 华北理工大学, 2019.

[77] Ricardo Baeza-yates, Berthier Ribeiro-Neto 著. 黄宣菁, 张奇, 邱锡鹏译. 《现代信息检索》. 机械工业出版社. 2012 年 10 月第 1 版.

[78] 刘士琛.面向推荐系统的关键问题研究及应用[D]. 中国科学技术大学, 2014.

[79] 李川.实时个性化推荐系统的设计与实现[D]. 北京邮电大学, 2015.

[80] 详细分析推荐系统和搜索引擎的差异. https://blog.csdn.net/cserchen/article/details/50422553

附录

附录 1 知识图谱/知识工程知识树

知识图谱/知识工程的知识树共包括 10 个二级分类和 212 个三级分类。 图中带“< >”的节点表示关系，没有标”< >”的标明的节点关系是上下位关系。		
一级分类	二级分类	三级分类
Knowledge Engineering	<is_kind _of>	knowledge technology 知识技术
		semantic (web) technology 语义技术
		web science 万维科学
		information science 情报科学
	<multidiscipline_of>	cognitive science 认知科学
		semantic web 语义网
		artificial intelligence 人工智能
		computer science 计算机科学
		natural language processing 自然语言处理
		information processing 信息处理
		social machine 社交机器
	<using_techniques>	unified modeling language 统一建模语言
		pattern recognition 模式识别
		information processing 信息处理
		clustering 聚类
		clistering algorithms 聚类算法
		data visualization 数据可视化
		data mining 数据挖掘
		quality management 质量管理
		design methodology 设计方法论
		feature extraction 特征提取
		feature space 特征空间
		feature selection 特征选择
		human centered computing 人机交互技术
		support vector machine 支持向量机
		statistical model 统计模型
		service oriented architecture 面向服务的体系结构
		markov chains 马尔可夫链
		social network analysis 社会网络分析
		decision models 决策模型
		data management 数据管理
		human interaction 人机交互
		decission tree 决策树
		genetic algorithm 遗传算法
		machine learning 机器学习

		information retrieval 信息检索
		semantic similarity 语义相似度
		semantic relatedness 语义相关性
		semantic computing 语义计算
		semantic analysis 语义分析
		graph theroy 图论
		explanation based learning 解释学习
		data integrity 数据完整性
		text anaylsis 文本分析
		text mining 文本挖掘
		bootstrapping method 拔靴法
		reinforcement learning 强化学习
		human computer interaction 人机交互
		transfer learning 迁移学习
		domain experts 领域专家
		situation aware 情境感知
		graphical user interface 图形用户界面
		predictive model 预测模型
	<aims_at>	better understanding 内涵理解
		computational intelligence 智能计算
		knowledge based system 知识系统
		knowledge base 知识库
		RDF repository 资源描述框架存储库
		knowledge management 知识管理
		knowledge management systems 知识管理系统
		decision support system 决策支持系统
		decision models 决策模型
		decision maker 决策者
		adaptive systems 自适应系统
		recommender systems 推荐系统
		multiagent systems 多智能体系统
		multi agnet systems 多智能体系统
		automomous systems 自动系统
		automomous agent 自动代理
		semantic search 语义检索
		question answering system 问答系统
		human robot interaction 人机交互
		intelligent assistant 智能辅助
		knowledege reuse 知识再利用
		knowledge sharing 知识共享
		expert system 专家系统
		intelligent systems 智能系统
	<using_information_sources>	social networks 社会网络

		web resource 网络资源
		world wide web 万维网
		distributed databases 分布式数据库
		big data 大数据
		information sources 信息源
		xml database 可扩展标志语言数据库
		heterogeneous database 异构数据库
		heterogeneous data source 异构数据源
		multimedia 多媒体
		wireless networks 无线网络
		relational database 关系数据库
	<appliations_in>	electric commerce 电子商务
		disaster management 灾害管理
		computational biology 计算生物学
		biomedical domain 生物医学领域
		health care 卫生保健
		scientific domain 科学域
		education 教育
		open government data 政府公开数据
		life science 生命科学
		gene expression data 基因表达数据
	knowledge representation	data model 数据模型
		concept modelling 概念模型
		concept model 概念模型
		conceptual model 概念模型
		semantic model 语义模型
		knowledge model 知识模型
		structured data 结构化数据
		formal specification 形式描述
		formal meaning prepresentation 形式意义表示
		formal semantics 形式语义
		commonsense knowledge 常识
		world knowledge 世界知识
		web of data 数据网
		background knowledge 背景知识
		domain knowledge 领域知识
		semantic network 语义网络
		ontology 本体论
		rough set 粗糙集
		rough set theroy 粗糙集理论
		concept map 概念图
		fuzzy sets 模糊集合
		rule based 基于规则

		rule based system 基于规则系统
		heuristic rule 启发式规则
		object oriented 面向对象
		semantic workflow 语义工作流
		first order logic 一阶逻辑
		logic programming 逻辑编程
		frame based system 框为本的系统
		fuzzy logic 模糊逻辑
		fuzzy systems 模糊系统
		formal logic 形式逻辑
		decision rule 决策规则
		temporal logic 时态逻辑
		dynamic logic 动态逻辑
		domain specific language 领域专用语言
		resource description framework 资源描述框架
		ontology language 本体语言
		web ontology lanuage 网络本体语言
		semantic web rule language 语义网规则语言
		owl 2
		collabrative ontology engineering 联合本体工程
		ontology engineering 本体工程
		ontology development 本体开发
		collabrative ontology development 联合本体开发
		ontology extraction 本体抽取
		ontology evolution 本体演化
		ontology versioning 本体版本
	knowledge acquisition	knowledge extraction 知识提取
		knowledge capture 知识获取
		knowledge construction 知识建构
		knowledge building 知识建构
		information extraction 信息提取
		entity resolution 实体解析
		entity recognition 实体识别
		entity disambiguation 实体消歧
		semantic annotation 语义标注
		taxonmy induction 感应规范
		concept clustering 概念聚类
		concept formation 概念形成
		concept learning 概念学习
		attribute value taxonomy 属性分类规范
		event detection 事件检测
		event identificaton 事件识别
		event extraction 事件抽取

		relation extraction 关系抽取
		semantic relation learning 语义关系学习
		relational learning 关系学习
		inference rule 推理规则
		rule learning 规则学习
	knowledge reasoning	case based reasoning 实例推理
		logical implication 逻辑蕴涵
		inference mechanisms 推理机制
		knowledge verification 知识验证
		semantic interpretation 语义解释
		uncertainty reasoning 不精确推理
		causual models 因果模型
		nonmonotonic reasoning 非单调推理
		spatial reasoning 空间推理
		temporal reasoning 时序推理
		abductive reasoning 溯因推理
		defult reasoning 默认推理
	knowledge integration	knowledge fusion 知识融合
		semantic integration 语义集成
		data fusion 数据融合
		inconsistent ontology 本体不一致
		heterogenous ontology 异构本体
		ontology interoperablity 互用性本体
		ontology mapping 本体映射
		ontology alignment 本体映射
		ontology matching 本体匹配
		schema mapping 模式映射
		schema matching 模式匹配
		matching function 匹配函数
		instance matching 实例匹配
		date linking 日期链接
		date interlinking 日期互联
		record linkage 记录链接
		thesaurus alignment 同义对齐
	knowledge storage	triple store 三元组存储
		RDF database 资源描述框架数据库
		RDF storge 资源描述框架存储
		graph database 图数据库
		exhaustive indexing 完整索引
		query language 查询语言
		conjunctive quries 合取查询
		RDF query 资源描述框架查询
		graph query 图查询

		query rewrite 查询重写
		distributed query 分布式查询
		subgraph matching 子图匹配
		graph partitioning 图划分
		data partitioning 数据划分

附录 2 Data Mining 知识图谱（共包含二级节点 15 个，三级节点 93 个）

领域	二级分类	三级分类
data mining (数据挖掘)	time series analysis(时间序列分析)	data streams(数据流)
		time series data(时间序列数据)
		real time(实时)
		time series(时间序列)
		complex dynamical networks(复杂动态网络)
		dynamic system(动态系统)
		nonlinear dynamics(非线性动力学)
		system dynamics(系统动力学)
		time frequency analysis(时频分析)
	association rule(关联规则)	rule induction (规则归纳)
		rule learning (规则学习)
		sequential pattern(序列模式)
		frequent itemsets(频繁项目集)
		pattern mining(模式挖掘)
		pattern matching(模式匹配)
		pattern classification(模式分类)
		frequent pattern(频繁模式)
	algorithm(算法)	algorithm design and analysis(算法设计与分析)
		upper bound(上界)
		prediction algorithms(预测算法)
		efficient algorithm(有效算法)
		computational modeling(计算模型)
		predictive models(预测模型)
		reinforcement learning(强化学习)
		neural networks(神经网络)
		computational complexity(计算复杂性)
		probabilistic logic(概率逻辑)
		structural risk minimization (结构风险最小化)
		constrained least squares (约束最小二乘)
		incremental learning(增量学习)
		pruning technique(修剪技术)
		matrix decomposition(矩阵分解)
		generative model(生成模型)

		hidden markov models(隐马尔可夫模型)
	big data(大数据)	dynamic databases(动态数据库)
		heterogeneous data(异构数据)
		text data(文本数据)
		data models(数据模型)
		sensor data(传感器数据)
		data warehouses(数据仓库)
		query processing(查询处理)
		data structure(数据结构)
		data analysis(数据分析)
		data privacy(数据隐私)
		personal data(个人数据)
		cloud computing(云计算)
		user behavior(用户行为)
		parallel processing(并行处理)
		graph data(图形数据)
		data intensive computing(数据密集型计算)
		data stream(数据流)
		distributed databases(分布式数据库)
		data handling(数据处理)
		data center(数据中心)
		data management(数据管理)
		data warehouse(数据仓库)
		data security(数据安全)
		data warehousing(数据仓库)
		privacy preservation(隐私保护)
		database management systems(数据库管理系统)
		data generation(数据生成)
	web mining(网络挖掘)	web search (网络检索)
		information retrieval(信息检索)
		link analysis (链接分析)
		image retrieval (图像检索)
		utility mining(效用挖掘)
		relevance feedback (相关反馈)
		recommender systems(推荐系统)
		mobile computing(移动计算)
		loclation based services(基于位置的服务)
		web pages(web 页面)
		collaborative filtering(协同过滤)
		social network(社交网络)
		social interaction(社交互动)
		social media(社交媒体)
		information filtering(信息过滤)

		social network analysis(社交网络分析)
		graph theory(图论)
		sentiment analysis(情感分析)
		opinion mining(意见挖掘)
		semantic web(语义网)
		social web(社交网页)
		online social network(在线社交网络)
		world wide web(万维网)
		web 2.0(网络 2.0)
		linked data(关联数据)
		social tagging system(社交标签系统)
		user generated content(用户生成内容)
		social tagging(社交标签)
		tag recommendation(标签推荐)
		link prediction(链接预测)
		web usage mining(web 使用挖掘)
		online community(网络社区)
		interaction network(交互网络)
		web forum(web 论坛)
	knowledge discovery(知识发现)	
	knowledge management(知识管理)	project management(项目管理)
		information technology(信息技术)
		information system(信息系统)
		database management(数据库管理)
		customer relationship management(客户关系管理)
		management system(管理系统)
	data management(数据管理)	data integration(数据整合)
		data compression(数据压缩)
		data point(数据点)
		spatial database(空间数据库)
		time series data(时间序列数据)
		range query(范围查询)
	text mining(文本挖掘)	text analysis(文本分析)
		text classification (文本分类)
		information retrieval(信息检索)
		natural language processing(自然语言处理)
		language model(语言模型)
		retrieval models(检索模型)
		feature selection(特征选择)
		text mining technique(文本挖掘技术)
		information retrieval models(信息检索模型)
		text data(文本数据)

		topic model(主题模型)
		recommender system(推荐系统)
		opinion mining(意见挖掘)
		feature extraction(特征提取)
		event detection(事件检测)
		information filtering(信息过滤)
		opinion analysis(舆情分析)
		sentiment analysis(情感分析)
		social media(社交媒体)
		disastrous event(灾难性事件)
		text summarization(文本摘要)
		query language(查询语言)
		query expansion(查询扩展)
		language modeling approach(语言模型方法)
		machine translation(机器翻译)
		biomedical text(生物医学文本)
	image mining(图像挖掘)	image reconstruction(图像重建)
		image segmentation(图像分割)
		image classification(图像分类)
		object recognition(目标识别)
	information network(信息网络)	information network mining(信息网络挖掘)
		heterogeneous information network(异构信息网络)
		graph theory(图论)
		online social networks(在线社交网络)
		recommender system(推荐系统)
		graph mining(图挖掘)
		location based service(基于位置的服务)
		network analysis(网络分析)
		link prediction(链接预测)
		graph data(图数据)
		factor graph(因子图)
		complex network(复杂网络)
		network topology(网络拓扑)
		homogeneous network(同构网络)
		information network analysis(信息网络分析)
		graph classification(图分类)
		graph clustering(图聚类)
		graph structure(图结构)
		random walk(随机游走)
		biological network(生物网络)
		computer networks(计算机网络)
		information integration(信息集成)
		graph database(图数据库)

		large graph(大图)
		heterogeneous network(异构网络)
		entity recognition(实体识别)
	graph mining(图挖掘)	large graph(大图)
		graph classification(图分类)
		random graph(随机图)
		directed graph(有向图)
		undirected graph(无向图)
	health care(卫生保健)	electronic health records(电子健康档案)
		gene expression(基因表达)
		biomedical research(生物医学研究)
		adverse drugs reactions(药物不良反应)
		genome wide association study(全基因组关联分析)
		patient care(病人医疗护理)
		computational biology(计算生物学)
		biological sciences(生物科学)
		medical research(医学研究)
	visualisation(可视化)	information visualization(信息可视化)
		data visualization(数据可视化)
		visual analytics(可视化分析)
		data visualisation(数据可视化)
		data analysis(数据分析)
		network visualization(网络可视化)
		visualization technique(可视化技术)
		visual content(视觉内容)
		visualization tool(可视化工具)
		interactive visualization(交互式可视化)
		graph visualization(图形可视化)
		graphical user interfaces(图形用户界面)
		computer animation(计算机动画)
		visual representation(视觉表征)
	information system(信息系统)	
	fuzzy data mining(模糊数据挖掘)	fuzzy set theory(模糊集合论)
		fuzzy set(模糊集)
		fuzzy clustering (模糊聚类)
	expert systems(专家系统)	knowledge management(知识管理)
		knowledge representation(知识表达)
		knowledge discovery(知识发现)
	similarity(相似性)	kernel operator (核算子)
		similarity relationship (相似关系)
		nearest neighbor (近邻)
		dissimilarity (相异性)

		citation matching (引文匹配)
		similarity search(相似搜索)
		similar kernel function(相似核函数)
		earth mover's distance(EMD 距离)
		kernel function(核函数)
		search problems(搜索问题)
		string matching(串匹配)
		similarity measure(相似性度量)
		keyword search(关键字检索)
		semantic similarity(语义相似度)
	data structure(数据结构)	data hierarchy (数据层次)
		complex data(复杂数据)
	unsupervised learning(无监督学习)	clustering (聚类)
		document clustering (文档聚类)
		hierarchical clustering (层次聚类)
		image clustering (图像聚类)
		data clustering (数据聚类)
		fuzzy clustering (模糊聚类)
		collaborative filtering (协同过滤)
		nonnegative matrix factorization (非负矩阵分解)
		cluster-based retrieval (聚类检索)
		fuzzy clustering (模糊聚类)
		clustering algorithms(聚类算法)
		outlier detection(孤立点检测)
		topic modeling(主题模型)
		subspace clustering(子空间聚类)
		pattern recognition(模式识别)
		mixture of gaussians(混合高斯模型)
		gaussian processes(高斯过程)
		density estimation(密度估计)
		dimensionality reduction(降维)
		dimension reduction(降维)
		maximum likelihood estimation(最大似然估计)
		matrix decomposition(矩阵分解)
		nonnegative matrix factorization(非负矩阵分解)
		sparse representation(稀疏表示)
		sparse matrices(稀疏矩阵)
		probability distribution(概率分布)
		probabilistic model(概率模型)
		hidden markov model(隐马尔可夫模型)
	supervised learning(有监督学习)	classification (分类)
		feature selection (特征选择)
		neural networks (神经网络)

		inductive learning (归纳学习)
		markov processes(马尔可夫过程)
		belief propagation(置信传播)
		decision tree(决策树)
		support vector machines(支持向量机)
		semi supervised learning(半监督学习)
		action recognition(行为识别)
		pattern recognition(模式识别)
		statistical analysis(统计分析)
		sparse coding(稀疏编码)
		object detection(目标检测)
		object recognition(目标识别)
		probabilistic logic(概率逻辑)
		regression(回归)
		manifold learning(流形学习)
		linear programming(线性规划)
		convex programming(凸规划)
		active learning(主动学习)
		random forest(随机森林)
		inference mechanisms(推理机制)
		bayes methods(贝叶斯方法)
		neural network(神经网络)
		classification algorithms(分类算法)
		bayesian methods(bayes 方法)
		random processes(随机过程)
		deep learning(深度学习)
		feature extraction(特征提取)
		recurrent neural network(递归神经网络)
		restricted boltzmann machines(受限玻尔兹曼机)
		hidden markov model(隐马尔可夫模型)
		boltzmann machine(玻尔兹曼机)
		bayesian inference(贝叶斯推断)
		convolutional neural networks(卷积神经网络)
		conditional random field(条件随机场模型)
		generative model(生成模型)
		probability distribution(概率分布)
		probabilistic model(概率模型)
		deep belief network(深度信念网络)
		logistic regression(logistic 回归)
	network analysis(网络分析)	social network(社交网络)
		social media(社交媒体)
		graph theory(图论)
		sensor networks(传感器网络)

		network analysis(网络分析)
		information diffusion(信息扩散)
		community detection(社区发现)
		network structure(网络结构)
		link prediction(链接预测)
		dynamic network(动态网络)
		network formation(组网)
		social learning(社会学习)
		social science(社会科学)
		information cascades(信息追随)
		communication networks(通讯网络)
		social influence(社会影响)
		complex network(复杂网络)
		network theory(网络理论)
		social interaction(社交互动)
		shortest path(最短路径)
		social behavior(社交行为)
		social life networks(社交生活网络)
	Decision analysis(决策分析)	decision support systems (决策支持系统)
		decision making (决策)
		data envelopment analysis (数据包络分析)
		information resource management (信息资源管理)



学术头条

学术君

关注“学术头条”并回复“caai”下载电子版