专题综述

视觉关系检测研究进展

南京大学　任桐炜 孙旭 于凡 武港山

一、引言

视觉关系检测旨在以<主语，谓语，宾语>形式的三元组来表示图像和视频中的物体及其交互关系[1]。图 1 和图 2 分别展示了面向图像和面向视频的视觉关系检测示例。对于给定的源图像，视觉关系检测方法会生成一定数量的三元组，以及每个主语、宾语所对应的物体的边界框(示例中用相同颜色标识主语/宾语和其对应物体边界框)；对于给定的源视频，视觉关系检测方法则需要为每个视觉关系三元组中的主语和宾语生成对应的物体边界框轨迹。

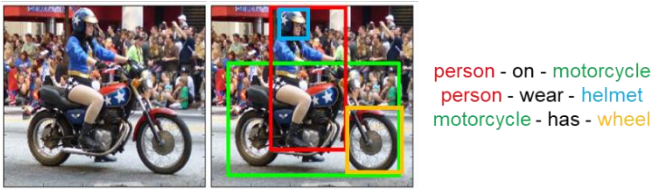


图 1 面向图像的视觉关系检测示例[1]

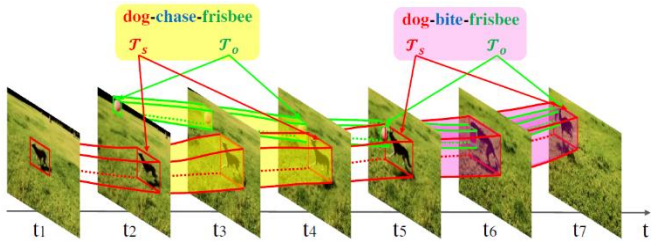


图 2 面向视频的视觉关系检测示例[2]

相比于物体检测、动作识别等技术，视觉关系检测提供了更加丰富的图像和视频语义描述，更好地连接视觉内容和自然语言描述，为图像和视频的语义结构化提供了新的探索途径，能够为内容检索、描述生成、视觉问答等任务提供更加有力的支撑，因而在近年来广受研究者的关注。

作为将计算机视觉与自然语言相结合的复杂语义理解任务，视觉关系检测有着独特的挑战。首先，视觉关系的类别由主语、谓语、宾语三者的类别组合而成。假设主语和宾语的类别空间为 O ,谓语的类别空间为 P，理论上视觉关系的类别空间规模为$|O|^2|P|$，因此视觉关系的类别数量会随物体和谓语类别数量的增加呈爆炸式增长。其次，海量的视觉关系类别在现实世界数据中仅有一少部分会以较高的频率出现，其余的大多数类别则只会在特定上下文中出现，这导致数据集的分布呈现严重的长尾现象。极不平衡的数据分布为视觉关系的分类带来了巨大的挑战，尾部数据的稀疏分布则对模型的小样本学习能力提出了要求。此外，少量视觉关系类别由于很罕见，难以获得训练数据，但这类关系在实际场景中仍有可能出现，因此零样本学习也是视觉关系检测会遇到的关键问题。

二、面向图像的视觉关系检测研究现状

2.1. 代表性方法

视觉关系检测研究起源于面向图像的视觉关系检测。由于面向图像的物体检测已经形成了很好的技术积累，早期的图像视觉关系检测大多采用独立的物体检测器，得到图像中的物体类别和位置后，再输入关系预测网络预测物体对之间的关系三元组。例如，Lu 等人[1]采用深度网络对每个物体对的联合区域提取特征；并且考虑到视觉关系中主语、谓语和宾语在语义上具有很高的

关联性，引入了语言模型使得相近的对象有相近的关系分布。Zhang 等人[3]通过将对象的视觉特征映射到低维的关系空间中，然后用对象间的转移向量来表示对象之间的关系。

为了提升视觉关系预测效果，端到端的视觉关系预测模型被提出。这些模型将在预测视觉关系时利用物体检测中所提取的视觉特征和空间特征，而非仅使用检测结果。例如，Zheng 等人[4]采用多线索注意力机制，更多关注有意义的视觉关系，并结合视觉特征、空间特征和语言特征，从而让不同的视觉关系更好地利用关联度更高的特征。由于视觉关系可以构成图结构，部分研究者采用图模型来引入物体和关系信息作为上下文，使得视觉关系的预测更加准确和完善。例如，Hu 等人[5]在视觉特征、空间特征和语言特征的基础上，采用交互图对物体及其关系建模，并通过信息传递的方式引入上下文更新物体和关系的特征。

由于视觉关系数据集的标注工作量远大于物体检测等任务的数据集，很难为视觉关系检测任务构建出大规模并具有很好一致性、完备性的数据集。为此，研究者从不同方向开展努力。为了解决训练数据规模的问题，Peyre 等人[10]将弱监督学习引入视觉关系检测，利用对整幅图像标注的视觉关系三元组即可作为训练数据，降低了训练数据标注的负担。同时，为了解决数据集中标注完备性、一致性较低的问题，Sun 等人[11]提出了层次式视觉关系检测，鼓励在无法产生准确的关系预测时，产生语义更加抽象但置信度更高的三元组来提高预测结果的正确率，以解决现有视觉关系检测中由于仅关注召回率而引入过多错误结果的问题。

2.2. 数据集

目前，有 2 个代表性的面向图像的视觉关系检测数据集：VRD 数据集[1]和 VG 数据集[6]。其中，VRD 数据集包含了 5,000 张图像、100 类物体、70 类视觉关系谓语和 37,993 个视觉关系。VG 数据集包含了 108,077 张图像、33,877 类物体、42,374 类视觉关系谓语和 230 万个视觉关系。由于这两个数据集存在较为明显的数据不均衡问题，导致所训练的模型容易依赖数据偏向，更多地预测占比大的视觉关系谓语。因此，Liang 等人基于 VG 数据集提出了 VrR-VG 数据集[7]，包含了更有视觉相关性的视觉关系谓语，并调整了每类关系的比例。此外，部分数据集专注于人和物体之间的交互关系，如 HICO-DET 数据集[8]和 HCVRD 数据集[9]。

三、面向视频的视觉关系检测研究现状

与面向图像的视觉关系检测相比，面向视频的视觉关系检测除了输入输出形式上的差异外，具有其自身的难点[2]。首先，面向视频的视觉关系检测使用边界框序列而非单个边界框来表示主语/宾语所对应的物体，需要准确检测出每个物体出现和消失的视频帧以及期间其在每帧上对应的边界框。其次，视频中包含了动态的视觉关系类型，如物体之间相对位置的变化、运动速度的相对快慢等，增加了视觉关系中谓语预测的复杂性。最后，视频中同一对主语和宾语所对应物体之间的关系可能会随时间推移而变化，因此需要准确判定每个视觉关系实例所持续的时间范围。

3.1. 代表性方法

针对面向视频的视觉关系检测特点，研究者们提出一些有效解决方案。例如，Shang 等人[2]将完整视频切分成固定长且有重叠的短视频片段，利用视频内容在短时间内能够保持相对稳定的特性，在每个视频片段上进行物体轨迹抽取和短期视觉关系识别，最后将从连续视频片段中检测到的短期视觉关系实例合并产生完整视频视觉关系实例。Tsai 等人[12]构建了基于条件随机场的门限时空能量图模型，同时从时间维度和空间维度进行关系推理。Qian 等人[13]将连续视频片段中检测到的物体连接为 3D 时空图，利用图卷积网络融合时间和空间上下文进行视觉关系识别，并采用孪生网络匹配合并相邻视频片段中检测到的短期视觉关系实例以提升性能。

3.2. 数据集

第一个面向视频的视觉关系检测数据集 ImageNet-VidVRD[2]基于 ILSVRC2016-VID 数据集构建，包含了 1,000 个视频，对其进行了稠密物体轨迹和视觉关系的标注，涉及 35 个物体类别、132 个谓语类别、3,219 个视觉关系类别，平均每个视频包含 9.5 个视觉关系实例。

由于 ImageNet-VidVRD 数据集中的视频场景相对简单且视频长度较短，Shang 等人从社交媒体收集了 10,000 个用户生成、总时长 84 小时的日常视频，对其进行了稠密物体轨迹和视觉关系的标注，构建了更具普适意义的 VidOR 数据集[14]。VidOR 数据集涉及 80 个物体类别、50 个谓语类别、6,553 个视觉关系类别，平均每个视频包含 38 个视觉关系实例。与 ImageNet-VidVRD 数据集相比，VidOR 数据集中的视频内容复杂多样，存在较为明显的物体遮挡、镜头抖动等现象，使得视觉关系检测更具挑战性也更贴近真实应用场景。

四、技术挑战赛

现有的视觉关系检测相关的技术挑战赛主要有 2 个：以人为中心的视觉关系竞赛(Person In Context，PIC)和视频视觉关系理解竞赛(Video Relation Understanding，VRU)，分别为面向图像和面向视频的视觉关系检测。

PIC 竞赛于 2018 年首次举办，依托 ECCV 会议，要求参赛者分割出图像中的人和物体并预测以人为主语的视觉关系，由南京大学获得冠军。第 2 届 PIC 竞赛于 2019 年依托 ICCV 会议举办，分为关系分割和人-物交互检测两个任务，均由天津大学获得冠军。

VRU 竞赛于 2019 年首次举办，依托 ACM MM 会议，分为视频物体检测和视觉关系检测两个任务，分别由深兰科技和南京大学取得第一名。第 2 届 VRU 竞赛仍然依托 ACM MM 会议，目前正在进行中。

五、未来发展方向

视觉关系检测为理解图像和视频语义提供了新的途径，但目前的研究还较为初步，仍有不少问题有待解决。例如，图像和视频中会包含大量的视觉关系，但仅有一小部分能够反映图像和视频的内容主旨。研究如何有效检测出能够反映内容主旨的视觉关系，有助于更好地支撑描述生成等上层应用。再如，当前的视觉关系检测数据集在规模上与真实世界中的视觉关系存在着很大的差距，所涉及到的物体类别、谓语类别数量依然十分有限，使得已有的视觉关系检测模型不足以产生足够丰富且准确的视觉关系。构建面向真实场景的更大规模、更丰富类型的视觉关系检测数据集将带来更多全新的挑战，包括更大容量的学习模型、多模态特征的融合、更大规模数据的收集、感知与逻辑推理的结合以及使用有限的数据进行知识迁移等。

责任编委 崔海楠

参考文献

[1] Lu C, Krishna R, Bernstein M, et al. Visual Relationship Detection with Language Priors[C]. European Conference on Computer Vision. 2016: 852-869.

[2] Shang X, Ren T, Guo J, et al. Video Visual Relation Detection[C]. ACM International Conference on Multimedia. 2017: 1300-1308.

[3] Zhang H, Kyaw Z, Chang S F, et al. Visual Translation Embedding Network for Visual Relation Detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5532-5540.

[4] Zheng S, Chen S, Jin Q. Visual Relation Detection with Multi-Level Attention[C]. ACM International Conference on Multimedia. 2019: 121-129.

[5] Hu Y, Chen S, Chen X, et al. Neural Message Passing for Visual Relationship Detection[C]. ICML Workshop on Learning and Reasoning with Graph-Structured Representations, 2019.

[6] Krishna R, Zhu Y, Groth O, et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations[J]. International Journal of Computer Vision, 2017, 123(1): 32-73.

[7] Liang Y, Bai Y, Zhang W, et al. VrR-VG: Refocusing Visually-Relevant Relationships[C]. IEEE International Conference on Computer Vision. 2019: 10403-10412.

[8] Chao Y W, Liu Y, Liu X, et al. Learning to Detect Human-Object Interactions[C]. IEEE Winter Conference on Applications of Computer Vision. 2018: 381-389.

[9] Zhuang B, Wu Q, Shen C, et al. HCVRD: a Benchmark for Large-scale Human-centered Visual Relationship Detection[C]. AAAI Conference on Artificial Intelligence. 2018: 7631-7638.

[10] Peyre J, Sivic J, Laptev I, et al. Weakly-supervised learning of visual relations[C]. IEEE International Conference on Computer Vision. 2017.

[11] Sun X, Zi Y, Ren T, et al. Hierarchical Visual Relationship Detection[C]. ACM International Conference on Multimedia. 2019: 94-102.

[12] Tsai Y H, Divvala S, Morency L P, et al. Video Relationship Reasoning Using Gated Spatio-temporal Energy Graph[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2019: 10424-10433.

[13] Qian X, Zhuang Y, Li Y, et al. Video Relation Detection with Spatio-temporal Graph[C]. ACM International Conference on Multimedia. 2019: 84-93.

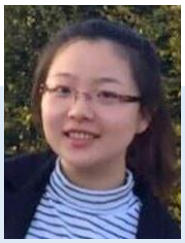
[14] Shang X, Di D, Xiao J, et al. Annotating Objects and Relations in User-generated Videos[C]. ACM International Conference on Multimedia Retrieval. 2019: 279-287.

任桐炜

南京大学副教授。主要研究方向为视觉媒体分析。
Email: rentw@nju.edu.cn

孙旭

南京大学硕士研究生。主要研究方向为视觉关系检测。
Email: sunx@smail.nju.edu.cn



于凡

南京大学硕士研究生。主要研究方向为视觉关系检测。
Email: yf@smail.nju.edu.cn

武港山

南京大学教授。主要研究方向为多媒体，计算机视觉，模式识别。
Email: gswu@nju.edu.cn

热点追踪

基于稀疏量化的神经网络加速与压缩

中国科学院自动化研究所 王培松 程健

图 1 本工作受邀在 NeurIPS 2019 进行展示

深度学习在越来越多的领域取得飞跃性进展，然而模型存储和计算复杂度却越来越高，严重妨碍了深度学习模型在实际场景中的大规模部署。因此，以模型压缩和加速为代表的深度学习计算优化技术成为最近几年学术界和工业界最为关注的焦点之一，对深度学习的进一步落地具有重要意义。

目前主流的神经网络加速与压缩方法包括网络剪枝、网络量化和知识蒸馏等。针对任意一种技术，均有大量的相关工作，例如网络剪枝，可以实现 90%以上的稀疏度而不会对精度产生影响；例如网络量化，可以实现 2~4 比特无损压缩，甚至对于某些网络和任务可以实现 1 比特压缩。一方面，单纯依赖某一种压缩手段，都很难实现更高比率的加速和压缩；另一方面，这些不同的优化技术具有一定的互补性，如何综合运用这些技术手段，以达到更高压缩率和加速率，成为网络压缩领域的热点研究问题之一。

MicroNet Chanllenge 竞赛是由 Google、Facebook 和 OpenAI 等机构在 NeurIPS 2019 上共同主办的，旨在通过优化神经网络架构和计算，达到模型高精度、计算效率和硬件资源占用等方面的平衡，实现软硬协同优化发展，启发新一代硬件架构设计和神经网络架构设计等。此次竞赛吸引了 MIT、加州大学、KAIST、华盛顿大学、京都大学、浙大和北航等国内外著名前沿科研院校，以及 ARM、IBM、高通和 Xilinx 等国际一流芯片公司参加。我们团队参加了此次比赛的图像分类任务，并获得 ImageNet 和 CIFAR-100 两个赛道的双料冠军。下面简单介绍我们提出的综合利用稀疏、量化以及蒸馏的网络压缩方案。

首先，对于网络剪枝，不同的层具有不同的敏感性，因此需要使用不同的稀疏率进行剪枝。我们提出一种简单的鲁棒性分析方法，对每一层权值进行不同稀疏度剪枝，并观察对精度的影响，从而确定每一层的稀疏度。图 2 展示了 MixNet-S 网络每一层的敏感度，可以看出

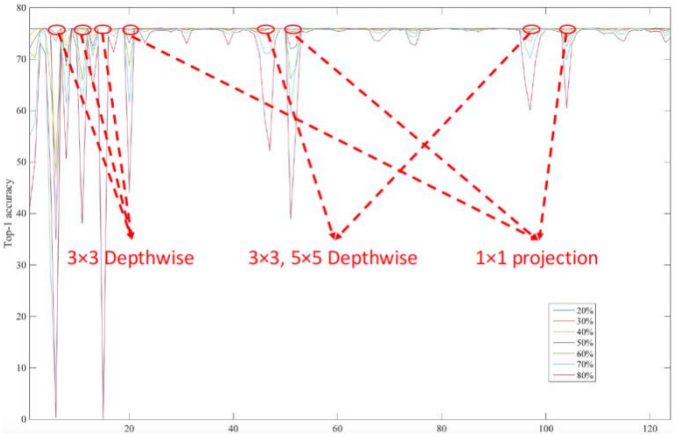


图 2 网络不同层鲁棒性分析

Student	Teacher	Teacher Acc.	Top-1 Accuracy
MixNet-s-pruned	-	-	74.4 %
MixNet-s-pruned	MixNet-s	75.9 %	74.6 %
MixNet-s-pruned	MixNet-m	77.2 %	74.9 %
MixNet-s-pruned	**MixNet-l**	**78.9 %**	**75.0 %**
MixNet-s-pruned	SENet154	81.3 %	74.7 %

表 1 不同性能教师网络对知识蒸馏的影响

对于不同层，网络的敏感度差异非常大，因此对于这些比较敏感的层需要使用更低的压缩率进行压缩。

对于量化，我们采用了两阶段量化方式，以实现激活量化和权值量化的解耦。具体而言，我们首先对激活进行量化，而保持权值为全精度浮点数。此时，网络仅仅学习中间层的低比特特征表达，浮点数权值能够保障学习到更好的中间层特征。在低比特特征学习之后，再对权值进行量化，通过激活量化和权值量化的解耦，降低网络量化的难度，从而达到更高的精度。另一方面，与网络剪枝类似，不同的层对于量化的敏感度不一样，我们采用了混合精度量化方案，即对于不同的层选取不同的位宽，达到精度和压缩率的平衡。

知识蒸馏使用一个高精度的教师网络去监督一个低精度的学生网络，往往可以大幅度提升学生网络的性能。然而，教师网络的选择对于学生网络的影响非常大。我们通过大量实验发现，通过使用更高性能的教师网络作为监督，并不能保证学生网络的性能更高。表 1 展示了不同性能的教师网络去监督 MixNet-S 网络的结果，可以看出，教师网络的性能过高或者偏低都不利于学生网络的学习。

通过综合运用以上网络稀疏化技术、网络量化技术以及知识蒸馏技术，可以将深度学习算法模型进行轻量化和计算提速，大幅度降低算法模型对算力、功耗以及内存的需求。最终，相比于主办方提供的基准模型，我们在 ImageNet 任务上取得了 20.2 倍的压缩率和 12.5 倍的加速比，在 CIFAR-100 任务上取得了 732.6 倍的压缩率和 356.5 倍的加速比，并获得上述两个任务的第一名。

责任编委 王金甲

王培松

中国科学院自动化研究所助理研究员，主要研究方向为机器学习、计算机视觉、神经网络高效计算。

Email: peisong.wang@nlpr.ia.ac.cn

程健

中国科学院自动化研究所研究员，人工智能和先进计算联合实验室主任，自动化所南京人工智能芯片研究院常务副院长，主要研究方向为机器学习、计算机视觉、芯片设计。

Email: jcheng@nlpr.ia.ac.cn

热点追踪

TCNN：利用轨迹片段进行视频目标检测

悉尼大学 欧阳万里

近年来图像目标检测准确率有了很大提升。除了更强力的卷积神经网络架构，新型检测算法框架也扮演了十分重要的角色。但是在视频目标检测领域，相关框架性研究仍然十分少，视频时序和上下文信息仍然没有被很好地利用起来。TCNN(Tubelets with Convolutional Neural Network)视频目标检测框架的提出，能够较好地将视频的时序和上下文信息利用起来，极大地提升了将单图目标检测算法应用在视频中的准确率。

该框架提出利用轨迹片段(Tubelets)信息来提升视频目标检测的准确率。主要模块包括多上下文抑制(Multi-Context Suppression)及运动导向的帧间传播(Motion-Guided Propagation)、轨迹片段二次估计(Tubelet Re-scoring)等。为了得到更好的结果，该方法还使用了多模型组合增强的技术。

多上下文抑制模块考虑了整个视频检测目标的置信度，综合排序后筛去低置信度目标，有效地降低了误检率。运动导向的帧间传播模块使用了光流信息估计一个检测目标在相邻若干帧图像中的运动位置变化，并将该检测目标向相邻帧传播，以此降低漏检率，提高召回率。轨迹片段二次估计利用了视觉目标跟踪的思想，将高置信度的目标跟踪形成轨迹片段，并在轨迹片段内重新估计所有目标的置信度，从而得到更可靠的估计。相比此前的视频目标检测的尝试，该框架充分分析了视频目标检测中的问题，系统地提出了如何利用视频时序和上下文信息的解决方法，为许多后来工作提供了重要的启示和借鉴。

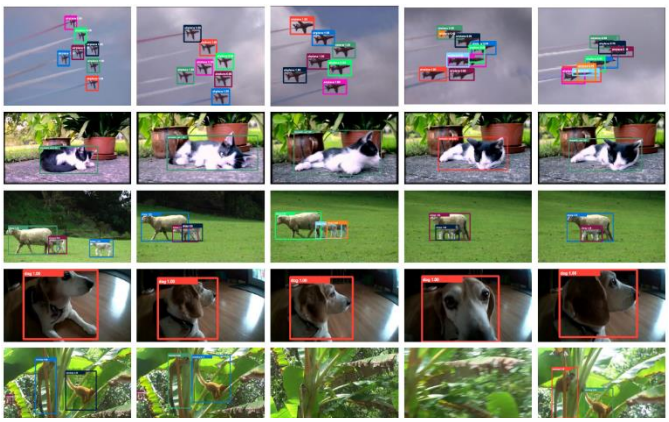
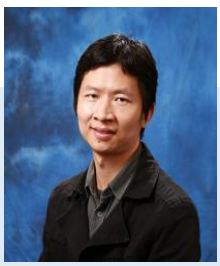


图 1 TCNN 算法效果示例

该算法框架在 2015 年的 ImageNet 大规模视觉识别竞赛中取得视频目标检测赛道冠军，相关论文投稿于 TCSVT 期刊。2020 年，论文第一作者康恺博士被授予 IEEE CAS 杰出青年作者奖。该奖项每年在电路与系统协会五个期刊发表的所有论文中选择一篇。

责任编委 储珺



欧阳万里

悉尼大学高级讲师，IEEE 高级会员，主要研究方向为基于深度学习的物体检测和跟踪。
Email: wanli.ouyang@sydney.edu.au

顶会观察

ICLR 2020

燕山大学　　王金甲

国际表征学习大会(International Conference on Learning Representations， ICLR)是人工智能(AI)分支表征学习，特别是深度学习领域的顶级学术会议。ICLR是由深度学习三巨头中的Yoshua Bengio和Yann LeCun 主导创办。ICLR 是世界上增长最快的一个人工智能会议。ICLR 与会者背景广泛，从学术界和工业界的研究人员，企业家和工程师，到研究生和博士后。

2020 年 4 月 27 日至 30 日第八届 ICRL 原定在埃塞俄比亚首都亚的斯亚贝巴举行，由于 COVID-19 新冠病毒疫情的全球肆虐，成为史上首个完全线上虚拟会议。

一、国际表征学习会议的亮点

由于新冠病毒疫情，非洲大陆史上第一次承办的大型人工智能顶会被迫采用完全数字虚拟会议，这成为了本次会议最大的一个亮点。虚拟会议网站的开发人员将每一篇论文的摘要、演讲、论文、代码、评论全部集中在一起。ICLR 会议接收论文的作者需要预先录制好自己的视频、音频或文本。每篇论文分两次播放，且每篇论文都有一个讨论聊天功能，可以通过文本进行交互。会议提前召集了 500 名志愿者对系统的各个方面进行了压力测试。虚拟会议减少了研究人员差旅费、注册费和时间成本，1300 多名演讲者吸引了来自近 90 个国家的 5600 多名参与者，参加人数暴增了一倍。根据大会官方统计数据，线上演讲视频观看次数突破 10 万+，相关 Zoom 讨论会议多达 1400 余次。大会主席康奈尔大学技术学院 Alexander Sasha Rush 居功甚伟。会议线上 Q&A 环节也十分精彩。全部会议内容免费开放。

二、论文录用情况

ICLR 2020 共收到了 2594 篇论文投稿，最终有 687 篇被接收，录取率为 26.5%，低于去年 31.4%。接收论文包括 48 篇 oral，108 篇 spotlight 和 531 篇 poster。每篇 oral 的演讲时间 10 分钟以上，每篇 spotlight 的演讲时间 4 分钟。

据统计，412 篇被接收论文中有华人学者参与，占比 60.0%。2566 位全部作者中共有 655 位华人学者，占比 25.5%。入选 ICLR 2020 论文最多的学者是 UC Berkeley 的副教授 Sergey Levine，共有 13 篇论文。入选论文较多的华人学者中有三位最为瞩目，分别是清华大学计算机系朱军教授，佐治亚理工学院计算科学与工程系终身副教授宋乐和北京大学的王立威教授。

根据清华大学计算机系研发的 AMiner 学术数据库统计显示，谷歌入选论文 80 余篇，其中 12 篇 Oral，7 篇满分论文。在入选高分论文中，如华为、腾讯、快手、字节跳动等国内企业，清华大学、北京大学、南京大学、哈尔滨工业大学、西安电子科技大学等国内高校都榜上有名。可以看到，无论是华人学者，还是国内企业、高校和研究所在人工智能领域发挥越来越大的作用。

ICLR2020 会议论文关键词包括深度学习(Deep Learning)、强化学习(Reinforcement Learning)、表征学习(Representation Learning)、图神经网络(Graph Neural Network)、生成模型(Generative Model)等。深度学习、强化学习持续占领最受欢迎关键词王冠，图神经网络则新贵上位。

三、主题演讲

会议一共邀请了 8 位全球高影响力的学者作主题演讲，包括 FAIR 首席科学家 Devi Parikh、“卷积网络之父” Yann LeCun、深度学习先驱 Yoshua Bengio、美国三院院士 Michael I. Jordan 等。这些特邀报告都是提前录制的，并当天开放，每个报告有 Q&A 讨论环节。

Devi Parikh 讨论了计算机视觉和自然语言处理交叉 AI 系统和挑战，解释了为什么视觉和语言的交集问题是令人兴奋的。Yann LeCun 看好自监督学习，“我们作为人类学到的大多数知识以及动物学到的大多数知识都是自监督模式，而不是强化模式。” 自监督学习算法不依赖注释，而是通过揭示数据各部分之间的关系从数据生成标签。Yann LeCun 认为不确定性是阻碍自监督学习成功的主要障碍，自监督表示学习的未来在于正规化的隐变量能量模型。Yoshua Bengio 总结了深度学习先验和意识处理的思想、因果结构和注意力机制的最近进展，目的是为了获得更好的分布外泛化能力。Michael I. Jordan 认为机器学习(ML)的决策将是未来的重点，从动力、统计和经济角度给出了机器学习的决策进展。

四、满分论文和新颖想法的论文

ICLR 会议采用 Open Review 的评审制度。所有提交论文都公开论文姓名等信息，并且接受所有同行的评审以及提问(open peer review)。会议出现了多达 34 篇满分论文，涉及的内容有神经结构搜索、生成对抗网络、多智能体学习、神经机器翻译、语言生成、信息瓶颈、稀疏表示、字典学习、图神经网络、强化学习、元学习、贝叶斯深度学习、无监督学习、对抗学习、最优化、推理、因果发现和自动微分框架等。可以看出这些满分论文的关键词也体现出了深度学习的主要趋势。这里分享几篇满分论文的精彩报告，分别来自北京大学[1]、清华大学[2]、图宾根大学[3]、斯坦福大学[4]、麻省理工学院[5]、西安电子科技大学[6]、南京大学[7]、DeepMind[8]和谷歌[9]等，论文内容分别涉及复杂系统框架的设计[1]、门控可学习 ISTA 稀疏编码[2]、自动微分框架的反向传播算法[3]、用不动点概念来描述 ReLU 函数[4]、生成对抗网络和隐空间插值用于模型泛化[5]，贝叶斯深度学习的神经语言模型[6]、生成网络的神经机器翻译[7]、元学习与扭曲梯度下降[8]，多智能体学习广义训练方法[9]等。

本次会议涌现出了许多令人感兴趣的论文，这里提及其中的几篇论文。Yoshua Bengio 等人提出了元学习因果结构；DeepMind 和哈佛大学研究了神经网络控制虚拟小白鼠；斯坦福大学和谷歌给出了 ELECTRA 预训练模型实现更高效的自然语言处理；谷歌发布了元数据集，学会从少样本中学习的数据集；DeepMind 和麻省理工学院等联合发布 CLEVRER 数据集，推动视频理解中的因果逻辑推理；谷歌和 UC 伯克利推出了单个 GPU 上可运行更高效的 Transformer 模型 Reformer；斯坦福大学和哈佛大学给出了预训练图神经网络模型。

五、研讨会

本次会议共有 15 个研讨会(workshops)[10-11]，涉及的主题有 AI 时代基础科学、AI 在癌症护理克服全球差异的能力、地球科学中的 AI、AI 和认知科学、深度神经网络和微分方程、重新思考 ML 的安全和隐私、ML 应对气候变化、现实生活中 ML、神经结构搜索、非洲自然语言处理、负担起的医疗 AI、不断变化的环境和任务中的强化学习、农业中的计算机视觉、用于决策的因果学习，以及发展中国家受限/低资源场景下的 ML。

六、总结与展望

AAAI 2020 实现了半线上会议，ICLR 2020 实现了完全线上虚拟会议。ICLR2020 成为了在线举行的最大人工智能学术研讨会，成为了完全线上虚拟会议的开端。ICASSP 2020 和计算机视觉顶会 CVPR 2020 也实现了完全线上虚拟会议。ECCV 2020、ICML 2020 和 NeurIPS2020 也宣称将转为完全线上虚拟会议。或许线上虚拟会议将是未来学术交流形态发展的必然趋势？

责任编委 任桐炜

参考文献

[1] Yilun Xu, Shengjia Zhao, Jiaming Song, Russell Stewart, Stefano Ermon. "A Theory of Usable Information under Computational Constraints", ICLR2020, https://openreview.net/forum?id=r1eBeyHFDH.

[2] Kailun Wu, Yiwen Guo, Ziang Li, Changshui Zhang. " Sparse Coding with Gated Learned ISTA ", ICLR2020, https://openreview.net/forum?id=BygPO2VKPH.

[3] .Felix Dangel, Frederik Kunstner, Philipp Hennig. "BackPACK: Packing more into Backprop", ICLR2020, https://openreview.net/forum?id=BJlrF24twB&noteId=EHmfrSLVu.

[4] Vaggos Chatziafratis、Sai Ganesh Nagarajan、Ioannis Panageas、Xiao Wang. "Depth-Width Trade-offs for ReLU Networks via Sharkovsky's Theorem", ICLR2020, https://openreview.net/forum?id=BJe55gBtvH.

[5] Ali Jahanian 、 Lucy Chai 、 Phillip Isola. "On the "steerability" of generative adversarial networks", ICLR2020, https://openreview.net/forum?id=HylsTT4FvB.

[6] Dandan Guo, Bo Chen, Ruiying Lu, and Mingyuan Zhou. "Recurrent Hierarchical Topic-Guided Neural Language Models", ICLR2020, https://openreview.net/forum?id=Byl1W1rtvH.

[7] Zaixiang Zheng 、 Hao Zhou 、 Shujian Huang， "Mirror-Generative Neural Machine Translation", ICLR2020, https://openreview.net/forum?id=HkxQRTNYPH&noteId=HmFOnYDNdx.

[8] Sebastian Flennerhag, Andrei A. Rusu, Razvan Pascanu, Francesco Visin, Hujun Yin, Raia Hadsell. " Meta-Learning with Warped Gradient Descent ", ICLR2020, https://openreview.net/forum?id=rkeiQlBFPB&noteId=zwVZBPeDT.

[9] Paul Muller、Shayegan Omidshafiei、Mark Rowland. "A Generalized Training Approach for Multiagent Learning", ICLR2020, https://openreview.net/forum?id=Bkl5kxrKDr&noteId=XEgYShJzuK.

[10] https://iclr.cc/virtual_2020/index.html.

[11] https://www.aminer.cn/conf/iclr2020/.

王金甲

燕山大学信息科学与工程学院教授，主要研究方向为信号处理和模式识别。
Email: wjj@ysu.edu.cn