中科院自动化所高君宇：图神经网络在视频分类中的应用【附PPT与视频资料】

[专知](javascript:void(0);)

**专知**

微信号 Quan\_Zhuanzhi

功能介绍 专知，为人工智能从业者服务，提供专业可信的人工智能知识与技术服务，让认知协作更快更好！

昨天

【导读】中国科学院自动化研究所高君宇讲述图神经网络在视频分析的应用

**作者介绍**

高君宇，中国科学院自动化研究所博士生，导师为徐常胜研究员。研究方向为基于深度学习的视频理解与应用。在IEEE Transaction on Image Processing（TIP）、CVPR、AAAI、ACM MM等CCF推荐的A类期刊、会议中发表多篇一作论文。获得了国家奖学金、中国科学院大学三好学生、三好学生标兵、百度奖学金、必和必拓奖学金、Rokid奖学金等。



高君宇

**导读**

当前，有监督的行为分类方法取得了显著的进展和很好的效果，但是这些方法依赖于大量的标注样本，而标注这些数据是极为耗时耗力的。因此，零样本视频分类的方法应运而生。目前，通过自动挖掘潜在概念（如行为、属性等）进行零样本视频分类的方法获得了极大的成功。但是，大多数现有方法只利用了视频的视觉信息而忽视了对这些概念之间的显式关系建模。因此，我们提出了：

一个基于知识图谱的端到端零样本行为识别框架，其可以

联合建模 行为-属性、属性-属性、行为-行为之间的关系。

具体的，我们设计了：

一个双支图卷积神经网络，其包括

一个分类器支

分类器支输入所有概念的词向量并产生对应概念的分类器。

一个实例支

实例支将属性的词向量和和每个视频实例的属性得分映射到一个特征空间中。

最后，学习到的分类器在产生的属性特征上进行评估，

并通过一个分类损失进行端到端地整体优化。

实验结果表明提出方法具有很好的效果。

**请关注专知公众号（点击上方蓝色专知关注）**

* **后台回复“GNN2019” 就可以获取讲义PPT的下载链接~**

视频地址：

https://www.bilibili.com/video/av53421272/

**1. Introduction**

近期有监督行为识别方面的研究有了长足的进展，这主要得益于鲁棒的深度学习方法框架和大规模的标注数据。然而，随着不断增长的行为类别，传统的有监督方法受到了类别可扩展性的限制。这些方法需要大量的、高花费的标注视频，使得这些方法很难泛化到未知类上。为了解决这个问题，零样本行为识别（Zero-Shot Action Recognition, ZSAR）吸引了学界的广泛关注，其可以不使用任何标注样本而能够识别未知类别。

现有的ZSAR工作主要基于两种方式：

（1）如图1 (a)所示，一些方法使用人类定义的属性来进行分类，其仅仅利用了行为-属性之间的关系来区分新的行为类别。另一方面，由于属性很难预先定义，因此在实际场景中，这些基于属性的方法很难以泛化到任意的未知类上。

（2）另一些方法使用行为名称的语义表示（如词向量等）在一个语义空间中建模行为-行为之间的关系，如图1 (b)所示。即使这些方法简单且高效，这种词向量空间只能隐式地表示行为-行为之间的关系。

另外，这些方法很难利用到视频的其他辅助信息。最近，受到物体和行为之间的强相关关系，许多方法把物体作为属性来进行零样本行为分类，并且获得了良好的效果。这些方法使用预训练的物体分类器来寻找视频中的物体，然而，其仅仅以固定的相似度权重考虑了行为-物体（属性）之间的关系，因此缺乏良好的端到端训练。

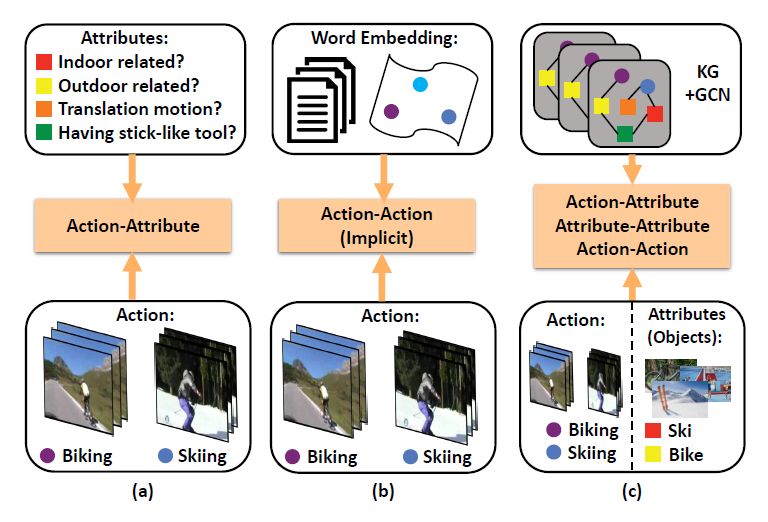


图1 3种零样本视频识别框架

除了上述提到的问题，大多数已有方法仅仅聚焦于视频的视觉特征而忽略了外部知识信息对零样本分类的指导作用。实际上，人类具有显著地能力来根据自身经验识别现实世界中的客观概念实体。因此，用结构化的知识信息建模各种概念（行为类别、属性等）之间关系是非常直观的。并且，这些知识信息有助于指导已知类上学习到的模型向未知类上迁移。

近期，知识图谱成功的应用到了各种计算机视觉任务中，如物体检测、多标签图片分类、零样本物体识别等等。通过在已知方法中引入知识图谱，实验效果获得了显著的提升。这说明了知识图谱确实具有补充现有方法所存在的知识鸿沟的能力。因此，在零样本行为识别中使用知识图谱也是非常有潜力的。另外，目前的方法大多忽略了视频的时序建模，比如直接在所有视频帧上使用均值池化等。但是，许多研究表明使用时序信息对视频理解是十分有帮助的。

**2. Our Methods**

受启发于上述观察，如图1 (c)所示我们提出了一个新颖的零样本视频分类方法，在一个端到端的框架中使用知识图谱来直接地、全面地建模行为-属性、属性-属性、行为-行为之间的关系。

事实上，这几类关系都可以直接或者间接地提升零样本学习的效果。这里，为了避免繁琐的属性标注，我们使用物体作为属性信息。为了高效地使用知识图谱中的知识信息，我们使用图卷积网络来在概念节点建模和传递信息。

具体的，我们提出了一个双支图卷积网络（Two-Stream GCN, TS-GCN），其包括一个分类器支和一个实例支。知识图谱被有机地嵌入到了这两个分支中来建模上述三种关系类型，如图2所示。

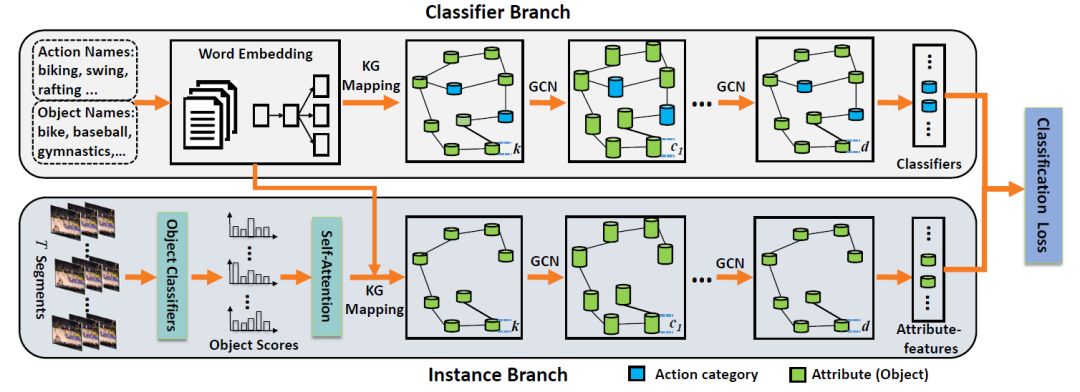
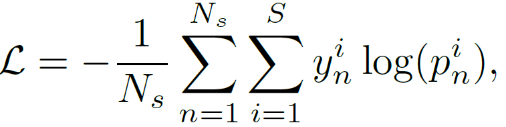
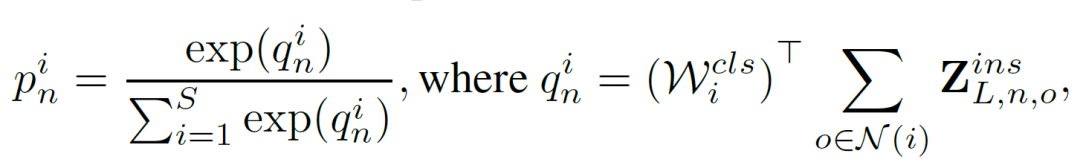


图2 TS-GCN框架

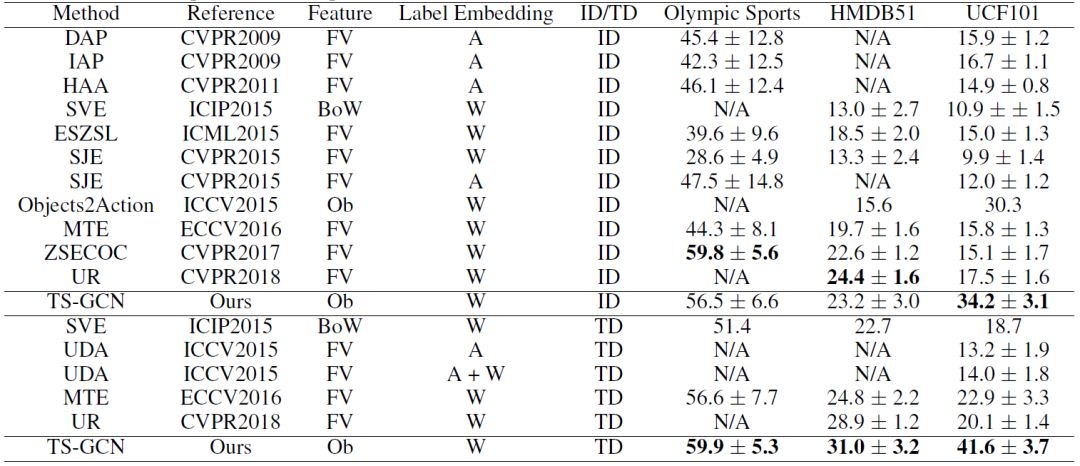
分类器支以所有概念的词向量表示为输入，对不同的行为种类产生分类器参数。实例支根据视频中的物体得分产生相应的属性特征。我们最终使用分类器支和实例支的输出，以一个分类损失来优化整个框架，如下式所示：





另外，为了建模视频的时序信息，我们在实例支中使用了一个自注意力模型来建模视频中动态变化的物体得分分布。在训练过程中，可见类上的分类器参数通过有监督的方式进行学习。

在测试阶段，训练好的模型以未知类上的视频特征为输入，产生在未知类上的预测分数。在三个视频数据集上的结果表示我们的方法取得了较好的效果，如下表所示。



**3. Take Home Message**

1. 考虑更丰富的知识信息，如边的类型等；结合图推理方法等；
2. 研究动态的图学习方法，以适应节点数目动态变化的场景；
3. 大规模图网络学习方法，值得研究；

**4. Reference**

**[1]** Junyu Gao, Tianzhu Zhang, Changsheng Xu. I Know the Relationships: Zero-Shot Action Recognition via Two-Stream Graph Convolutional Networks and Knowledge Graphs. AAAI, 2019.

**[2]**Wang, Xiaolong, Yufei Ye, and Abhinav Gupta. Zero-shot recognition via semantic embeddings and knowledge graphs. CVPR, 2018.

**[3]**Kampffmeyer, Michael, et al. Rethinking knowledge graph propagation for zero-shot learning. CVPR, 2019.

**[4]**Lee, Chung-Wei, et al. Multi-label zero-shot learning with structured knowledge graphs. CVPR, 2018.

**转载于人工智能前沿讲习班**