

一种协同过滤式零次学习方法^{*}

杨博^{1,2}, 张钰雪晴^{1,3}, 彭羿达^{1,2}, 张春旭^{1,2}, 黄晶^{1,2}



¹(符号计算与知识工程教育部重点实验室(吉林大学), 吉林 长春 130012)

²(吉林大学 计算机科学与技术学院, 吉林 长春 130012)

³(吉林大学 软件学院, 吉林 长春 130012)

通讯作者: 黄晶, E-mail: huangjing@jlu.edu.cn

摘要: 深度学习算法在很多有监督学习任务上达到了令人满意的结果,但其依赖于大量标注样本,并且使用特定类别训练的分类器,只能对这些类别进行分类.零次学习希望计算机像人类一样,能够结合历史经验与知识进行推理,无需使用大量新类别样本训练,便可达到识别新类别的效果.发现了零次学习任务存在“冷启动”以及矩阵稀疏两个特点,这些特点在推荐任务中同样存在.受推荐任务启发,将零次图像分类任务建模为矩阵填充问题,借鉴推荐领域中协同过滤算法,将稀疏的样本标签矩阵视为非稀疏的视觉特征矩阵和类别特征矩阵的内积结果,进而实现对新类别样本的分类预测.此外,构建了基于类间语义关联的语义图结构,使用图神经网络进行已知类别和新类别之间的知识迁移,以较小代价为类别学得准确的语义特征.在 3 个经典零次学习数据集上分别进行传统零次学习和广义零次学习实验,实验结果表明:提出的协同过滤式零次学习方法能够有效提升分类精度,且训练代价较小.

关键词: 零次学习;协同过滤;矩阵填充;图神经网络

中图法分类号: TP183

中文引用格式: 杨博,张钰雪晴,彭羿达,张春旭,黄晶.一种协同过滤式零次学习方法.软件学报,2021,32(9):2801–2815.
http://www.jos.org.cn/1000-9825/6418.htm

英文引用格式: Yang B, Zhang YXQ, Peng YD, Zhang CX, Huang J. Collaborative filtering based zero-shot learning. Ruan Jian Xue Bao/Journal of Software, 2021,32(9):2801–2815 (in Chinese). http://www.jos.org.cn/1000-9825/6418.htm

Collaborative Filtering Based Zero-Shot Learning

YANG Bo^{1,2}, ZHANG Yu-Xue-Qing^{1,3}, PENG Yi-Da^{1,2}, ZHANG Chun-Xu^{1,2}, HUANG Jing^{1,2}

¹(Key Laboratory of Symbolic Computation and Knowledge Engineering for Ministry of Education (Jilin University), Changchun 130012, China)

²(College of Computer Science and Technology, Jilin University, Changchun 130012, China)

³(College of Software, Jilin University, Changchun 130012, China)

Abstract: Many deep learning algorithms have achieved satisfactory results on many supervised learning tasks, but they rely on a large number of labeled samples, and the classifiers trained with specific categories can only classify these categories. Zero-shot learning wishes that the computer can reason like a human, it uses historical knowledge to infer the characteristics of new objects and has the ability to recognize novel categories without lots of samples. It is found that there are sparse matrix and “cold-start” phenomena in zero-shot learning task, these phenomena are also in the recommendation tasks. Inspired by the recommendation tasks, the zero-shot classification task is modeled as a matrix completion problem, hoping to learn from the collaborative filtering algorithms in the

* 基金项目: 国家自然科学基金(61876069, 62172185); 吉林省自然科学基金(20200201036JC); 吉林省科技厅重点研发项目(20180201044 GX, 20180201067GX)

Foundation item: National Natural Science Foundation of China (61876069, 62172185); Jilin Province Natural Science Foundation (20200201036JC); Jilin Province Key Scientific and Technological Research and Development Project (20180201044GX, 20180201067GX)

收稿时间: 2020-11-12; 修改时间: 2021-02-11; 采用时间: 2021-07-12

recommendation field, which regards the sparse labeled matrix as the product of the visual feature matrix and semantic feature matrix, and then classifies the novel samples. In order to make the semantic representation of each category more accurate, a semantic graph structure is constructed based on the semantic relations between categories and a graph neural network is applied on it for information transferring between known and novel categories. Traditional zero-shot learning and generalized zero-shot learning experiments are performed on three classic zero-shot learning data sets. The experimental results show that the collaborative filtering based zero-shot learning method proposed in this study can effectively improve the classification accuracy, and the training cost is relatively small.

Key words: zero-shot learning; collaborative filtering; matrix completion; graph neural network

深度学习算法在图像识别^[1,2]、语音识别^[3,4]、自动驾驶^[5,6]等大数据处理和人工智能任务中取得了巨大成功,然而这类有监督学习方法存在的限制是:模型过度依赖大量训练样本,每个类别需要几百甚至上千个样本,才可使模型达到令人满意的识别精度以及泛化能力;并且训练好的模型只能对已训练过类别的样本进行识别,无法应用于新类别样本.与之相比,人类的推理过程更加灵活,能够结合历史经验与新知识进行推理.例如,小孩子认识马这种动物,在学习到“斑马是具有黑白条纹的马”这条新知识后,便可推理出从未见过的新动物——斑马.零次学习任务希望计算机模仿人的推理过程,通过将历史经验迁移至新任务,无需使用大量新类别样本训练,便可达到识别新类别的效果.

在零次学习任务中,已知类别集合和新类别集合不相交,通常需要引入外部语义知识使得零次学习模型有效.目前,一些工作通过引入类别属性、词向量或类别描述等外部知识为各类别学得更准确的特征表示,进而构建类别语义与视觉特征之间的映射关系^[7].然而在大多数零次学习方法中,语义知识表示过程是独立完成的,类间互不影响,不能很好表现类间关联,且分类准确率仍有提升空间.知识图谱是一种显式组织知识以及知识间关联的数据结构,有工作将知识图谱引入零次学习,利用外部知识库中类间关系构建图谱,并借助图卷积神经网络实现图谱上的类间知识传递^[8].相较于直接使用语义知识,引入知识图谱中的类间关系能够有效提升对新分类的预测效果.然而,将知识图谱引入零次学习存在以下难点:高质量知识图谱构建较为困难;需要针对不同数据集分别构建对应的图谱;构造的图谱中可能存在噪声.

本文通过研究发现:零次学习与协同过滤推荐有内在联系,两者都可以建模为低秩矩阵填充问题.零次学习任务具有两个基本特点:新类别样本在模型训练过程中不出现,对新类别样本的分类可视作“冷启动”情况;样本标签矩阵是稀疏矩阵,仅有样本真实所属类别位置非空.这两个特点导致零次学习模型无法得到充分训练,很难完成对新类别样本的准确分类.冷启动和矩阵稀疏问题同样存在于推荐任务中,可采用协同过滤的思想进行处理.协同过滤算法通过挖掘用户关联和物品关联,为某些没有历史行为数据的新用户及新物品完成推荐,从而缓解冷启动和稀疏性带来的推荐不准确问题^[9].理论上,协同过滤算法将基于稀疏评分矩阵的预测问题建模为低秩矩阵填充问题,使用用户-物品评分矩阵中的已知值估计其中的缺失值.若将评分矩阵视为用户特征矩阵和物品特征矩阵的内积,则可采用矩阵分解方法解决低秩矩阵填充问题^[10].

基于上述发现,本文拟将零次学习和协同过滤这两个领域有机结合起来,提出一种新颖的协同过滤式零次学习方法,将零次学习任务建模为标签矩阵填充问题,借鉴协同过滤思想,将稀疏的标签矩阵分解为非稀疏的视觉特征矩阵和语义特征矩阵,进而实现对新类别样本的分类预测.该方法借鉴协同过滤思想,挖掘已知类别样本标签矩阵,学习视觉特征与类别语义知识的对应模式,并进一步迁移至对新类别样本的分类任务.为了使用较小代价为各类别学得准确的语义表示,本文以语义图的形式引入知识,建立已知类别与新类别之间的语义关联.本文使用图神经网络实现图上节点间的信息传递,促进已知类别和新类别之间的语义知识交互,打破现有方法中语义表示学习过程的独立性限制.此外,本文在目标函数中额外加入了图结构约束正则项,以维持类别在原始语义特征空间中的语义关联.针对传统零次学习任务和广义零次学习任务,在 3 组经典零次学习数据集上对本文所提出的协同过滤式零次学习方法进行验证,实验结果表明:本文方法能够普遍提升各任务的评估指标,并在各任务下维持稳定且优异的效果.据我们所知,目前尚未有从协同过滤的角度研究零次学习的相关工作.

本文第 1 节描述零次学习任务所对应的机器学习问题定义.第 2 节回顾现有零次学习的相关工作.第 3 节介绍本文提出的协同过滤式零次学习方法,包括视觉特征抽取、语义知识表示以及协同过滤式零次学习算法等

方面,第 4 节采用实验验证本文方法在各项任务中的表现并进行分析.最后一节对本文工作进行总结,并进一步对将协同过滤思想引入零次学习的未来研究进行分析和展望.

1 零次学习定义

在零次学习中,已知类别集合记作 S ,新类别集合记作 U , S 和 U 不相交,即 $S \cap U = \emptyset$.对于 $S \cup U$ 集合中的每个类 k ,可表示为一个语义特征向量 $c_k, c_k \in \mathbb{R}^m, m$ 为语义空间维度,可以由类别属性、词向量或类别描述等语义知识构成.零次学习可分为传统零次学习任务和广义零次学习任务:

- (1) 在传统零次学习任务中,训练集记作 $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$,其中, x_i^s 和 y_i^s 分别表示已知类别样本与其类别标签, N_s 表示训练集样本总量;测试集记作 $D_u = \{x_i^u\}_{i=1}^{N_u}$,其中, x_i^u 表示测试样本, N_u 表示测试集样本总量.传统零次学习任务旨在学得一个映射函数 f ,对 D_u 中的每个 x_i^u ,给出其在集合 U 上的预测结果 y_i^u ;
- (2) 在广义零次学习任务中,训练集为 $D_s - \{(x_i^s, y_i^s)\}_{i=1}^{N_s}$,测试集为 $D_u \cup \{x_i^s\}_{i=1}^{N_s}$,表示从原始训练集中划分出 N_e 个已知类别样本 x_i^s 至测试集中.广义零次学习任务同样旨在学得一个映射函数 f ,对测试集中每个样本,给出其在集合 $S \cup U$ 上的预测结果,相较于传统零次学习任务难度更高.

2 相关工作

从模型训练过程是否包含新类别样本的角度,现有零次学习方法可分为归纳式零次学习^[11-13]和直推式零次学习^[14-16]两大类:前者的训练过程不包含新类别样本,后者包含新类别样本但不包含其标签.尽管直推式零次学习可以有效缓解由于训练样本和测试样本分布不同导致的领域漂移问题,但在模型学习过程中获取到所有新类别样本这一假设不太符合实际场景,因此,本文工作主要聚焦于归纳式零次学习.在零次学习任务中,由于训练集样本类别和测试集样本类别不相交,所以零次学习任务的实现通常依赖于引入外部语义知识,进而构建视觉空间和语义空间的映射关系,如图 1 所示.

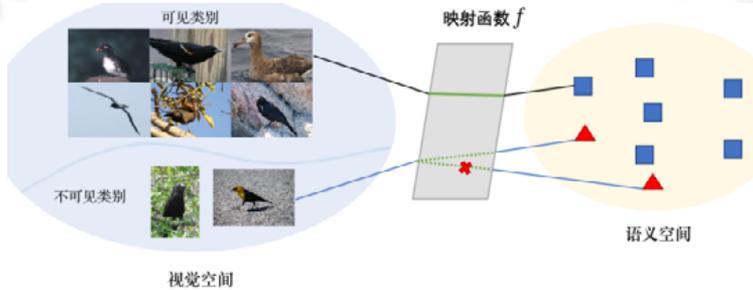


Fig.1 Compatibility learning framework between visual space and semantic space

图 1 视觉空间与语义空间的匹配学习框架

根据嵌入空间的不同选择,现有归纳式零次学习方法可分为 3 类.

- 第 1 类方法以 Lampert 等人^[17-19]为代表,选择语义空间作为嵌入空间,将样本的视觉特征嵌入至语义空间,与类别原型计算距离并进行匹配;
- 然而, Radovanovic 等人^[20]发现:第 1 类方法在使用语义空间作为嵌入空间时,需要将高维视觉特征做低维映射,使得空间发生萎缩,点与点之间更加稠密,加重由于最近邻计算产生的枢纽点问题.因此,第 2 类方法选择更高维的视觉空间作为嵌入空间,建立从类别语义原型到视觉特征空间的映射,在一定程度上缓解枢纽点问题^[12,21,22];
- 此外,以 Changpinyo 等人^[23-26]为代表的最后一类方法选择除视觉空间和语义空间之外的第三方空间作为嵌入空间,视觉特征及语义原型都向其做映射.

零次学习的特点与推荐任务相似,面临冷启动和矩阵稀疏问题.由于模型在训练过程中未见过新类别样本,

对新类别样本的分类属于冷启动;且样本标签矩阵中只有样本真实所属类别位置不为空,矩阵非常稀疏.在推荐任务中,冷启动指为没有历史行为数据的新用户做个性化推荐,矩阵稀疏指大量用户只与一小部分物品完成交互,表现在用户-物品评分矩阵中存在大量空值.Goldberg^[27]于 1992 年首次提出了协同过滤的概念,它的优势在于能够通过评估用户关联及商品关联,挖掘交互矩阵蕴含的信息.协同过滤算法^[28,29]的本质是考虑大量用户的偏好信息来对某一用户的偏好作出预测,通过计算用户相似性及物品相似性,可以为某些没有历史行为数据的新用户及新物品完成推荐.此外,将评分矩阵分解为有效的用户特征矩阵和物品特征矩阵,利用低维、连续特征矩阵的稠密性,缓解矩阵稀疏导致的评分预测不准确问题.协同过滤推荐可分为 3 种:基于用户的协同过滤、基于物品的协同过滤以及基于模型的协同过滤.基于用户的协同过滤通过观测用户与不同物品的交互,评估用户之间相似程度,使用高相似度用户的偏好信息进行推荐.基于物品的协同过滤的原理与前者类似,主要思想是评估物品之间相似程度,将高度相关的物品推荐给用户.基于模型的协同过滤采用不同的机器学习思想建模推荐任务,主要工作有:以文献[30]为代表的关联规则方法;以 Gibbs 抽样算法^[31]为代表的聚类算法;以及使用奇异值分解^[32]、降低矩阵稀疏性的矩阵分解算法等.

零次学习需要通过引入外部语义知识建立类间关联.知识图谱是一种显式地组织知识及知识关联的数据结构,通过为零次学习引入知识图谱中的结构化信息和复杂关系,可以更好地进行知识迁移.Wang 等人^[8]于 2018 年提出将领域知识图谱与语义向量相结合,知识图谱中以类别作为节点,节点初始表示为类别的词向量特征,图谱中的边采用知识库中不同类别之间的关系.通过利用图卷积神经网络(graph convolutional neural network,简称 GCN)在不同类别之间迁移信息,最后每个节点的输出特征代表对此类别的分类参数(即全连接层参数).实验结果证明,基于语义向量和知识图谱的方法比仅仅基于语义向量的方法效果更加优异.同年, Kampffmeyer 等人^[33]指出:多层 GCN 结构需要传递并吸收图中较远节点的知识,由于在每一层都要执行 Laplacian 平滑,这种操作会稀释知识从而导致性能降低.文中提出的基于密集图的传播模型(dense graph propagation)通过建立节点与较远邻居的直连边,来更好地利用知识图谱的层次图结构.

3 模型描述

本文借鉴协同过滤算法的思想,将零次分类建模为矩阵填充问题,利用已知类别标签矩阵中非空位置信息指导图像视觉特征以及类别语义特征的学习.模型整体框架如图 2 所示,分为视觉特征抽取、语义知识表示以及协同过滤式零次学习这 3 个主要模块.本文采用深度卷积模型作为视觉特征提取部分的网络架构,为图像样本提取像素级特征;语义知识表示部分涉及用于知识迁移的语义图构建以及知识表示学习过程,建立已知类别和新类别间关联,借助图神经网络进行图上的信息传递,为每个节点(类别)学得更好的特征表示;协同过滤式零次学习部分描述如何使用协同过滤思想建模零次学习过程并解决零次分类问题,是本文的核心模块.

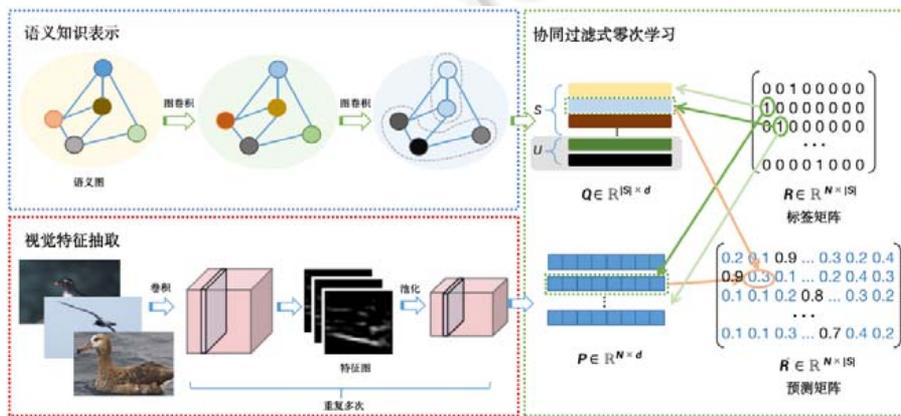


Fig.2 Overall framework of collaborative filtering based zero-shot learning

图 2 协同过滤式零次学习的总体框架

3.1 视觉特征抽取

零次学习主要面向零次图像分类任务,模型输入通常为未处理的原始图片.本文采用广泛应用于计算机视觉领域中的深度卷积神经网络,将图像转换为模型可识别数据,提取图片中各目标的有效视觉特征.通过将各卷积层、激活层、池化层组合在一起,图像特征不断被提取和压缩,最终每张图片表示为一个 d 维视觉特征向量,所有图像的视觉特征向量构成视觉特征矩阵 $P, P \in \mathbb{R}^{N \times d}, N$ 为训练集样本总量(这里不区分传统零次学习和广义零次学习).

3.2 语义知识表示

能否为类别学得准确的特征表示,使潜在的语义特征矩阵能够表达更多信息,是影响协同过滤效果的重要模块.为了挖掘已知类别与新类别之间的语义关联,实现知识迁移,本文通过构建语义图的方式建立类间关联.语义图定义为 $G=(H,A)$,图中节点由已知类别和新类别共同构成, H 表示节点特征矩阵,每个节点 h_i 的初始表示为类别语义知识(本文选用类别属性特征),即 $h_i^0 = c_i \in S \cup U$. A 表示图的邻接矩阵, a_{ij} 表示第 i 个类别与第 j 个类别的语义关系,构造过程如下:

首先计算两两节点间的语义距离 d_{ij} ,在这里,选择欧氏距离进行计算:

$$d_{i,j} = d(h_i^0, h_j^0) = d(c_i, c_j) = \sqrt{(c_{i_1} - c_{j_1})^2 + (c_{i_2} - c_{j_2})^2 + \dots + (c_{i_m} - c_{j_m})^2} = \sqrt{\sum_{k=1}^m (c_{i_k} - c_{j_k})^2} \quad (1)$$

其中, c_{i_k}, c_{j_k} 分别表示两个节点各维度的语义特征, m 为语义知识维度.计算得到 d_{ij} 之后,应用转换函数 $\kappa(\cdot)$ 将此距离关系转化为语义关联,见公式(2):

$$a_{ij} = \kappa(d_{ij}) = \exp(-d_{ij}/\varepsilon) \quad (2)$$

其中, ε 表示转换系数,是一个大于 1 的正数.两个节点的初始语义特征越相近,则其计算出的语义距离越小,那么语义关联越大;反之,不相似的节点之间的语义关联较小.不难发现:邻接矩阵 A 的对角线位置值均为 1,表示每个类别与自己本身的关联最大;其余位置值均在 0-1 之间,表示两个类别的语义关联系数.

语义图构建完毕后,本文选用图神经网络进行图上节点的表示学习.在图卷积过程中,每个节点在特征表示的更新过程中会受到相关性较大的邻域节点的影响,相似节点的特征表示趋于相似,不相似节点的特征表示的差异更大.图 2 语义知识表示模块中的虚线表示相似节点的特征表示趋于相似这一过程.由于图上的消息传递机制在所有节点对中共享,语义图中同时包含已知类别与新类别,所以通过已知类别样本训练好的传递机制可直接泛化至新类别,从而为新类别学得合适的特征表示.图的具体更新过程如下:首先,在训练之前随机初始化所有可学习参数 W 和 b, A 为上述过程计算出来的语义关系矩阵(图邻接矩阵).在每次迭代中,每个节点 i 压缩并整合来自其邻域节点传递的信息,在此过程中,邻接矩阵数值越大,越会促进此位置节点向节点 i 的消息传递.然后,使用整合后的信息与节点 i 上一次迭代后的特征表示共同对节点 i 的表示进行更新.具体更新公式如下:

$$h_i^{(t+1)} = \sigma \left(\sum_{j \in N_i} \hat{A} h_j^{(t)} W^{(t)} + b^{(t)} \right) \quad (3)$$

$$\hat{A} = D^{-\frac{1}{2}} (A + I) D^{-\frac{1}{2}} \quad (4)$$

其中: h_i^t 表示节点特征, $h_i^0 = c_i$; W 和 b 表示图卷积神经网络中的可学习参数矩阵; N_i 表示节点 i 的邻居节点集合;

\hat{A} 表示经标准化处理后的对称邻接矩阵; I 为单位矩阵; D 为图的度矩阵, $D^{-\frac{1}{2}}$ 起归一化作用.模型趋于使用相似度更高的邻居信息来对当前节点的表示进行更新.使用图卷积神经网络 $g(\cdot)$ 多层迭代消息传递过程,最终为每个节点 i 学得合适的语义特征表示 $g(h_i^0), g(h_i^0) \in \mathbb{R}^d$, 最终所有节点的语义特征构成语义特征矩阵 Q .

3.3 协同过滤式零次学习

在推荐任务中,使用“用户-物品评分矩阵”描述用户对物品的偏好信息,由于用户只会与其中一小部分物品产生交互,故评分矩阵中会存在很多空白.这些空白的存在导致很难对矩阵进行分析,所以在分析之前,一般需

要先对其进行补全处理.矩阵中空白位置的未知信息蕴含在非空位置之中,此时可借助这些非空位置数据,估计并填充某位置的缺失值.协同过滤算法通过将稀疏的评分矩阵视为非稀疏的用户特征矩阵和物品特征矩阵的内积结果,可缓解矩阵稀疏问题的影响.两矩阵根据相关性进行低秩重构,使用重构矩阵的对应元素值作为对原评分矩阵缺失值(即空白位置)的评分预测,对原稀疏矩阵进行补全.此外,协同过滤算法可通过挖掘用户间关联及物品间关联,推断新用户对其未交互的物品的喜好程度,进而为新用户进行推荐.

在零次分类任务中,每张图像 x_i 的真实标签可编码为一条独热向量,如图 3 所示表中的一列.由所有独热向量组成的样本标签矩阵 R 中,存在大量空白位置,是一个稀疏矩阵.注意:在该矩阵中,未知类别对应的行元素以及测试图片对应的列元素,都是未被观测的值,需要通过算法进行填充或补全.对于所有新类别而言,由于在训练过程中不包含任何新类别的样本,因此对新类别样本的分类任务相当于“冷启动”问题.受协同过滤推荐过程启发,本文将样本标签矩阵 R 视为两个矩阵的内积结果,即:由卷积神经网络提取出的图像特征向量构成的视觉特征矩阵 P 以及由类别的语义知识构成的语义特征矩阵 Q .第 3.1 节和第 3.2 节分别辅助视觉特征矩阵 P 与语义特征矩阵 Q 的学习.矩阵 R 中,“1”的位置表示样本 x_i 属于此类别,使用此信息可对图像隐特征表示和类别隐特征表示的准确程度进行监督.通过将学得 P, Q 矩阵进行点积,可预测每个图像样本对各类别的预测概率,对矩阵填充补全.对新类别样本的预测过程可看作是根据此样本提取出的视觉特征,将与视觉特征关联性最大的类别作为预测结果,该过程与协同过滤推荐过程类似.



Fig.3 Collaborative filtering based zero-shot learning

图 3 协同过滤式零次学习

一些零次学习方法使用已知类别样本预训练卷积神经网络,将预训练好的分类器参数(即全连接层参数)对语义特征向量进行监督,将语义特征向量与分类器参数之间的距离度量作为损失函数.这种做法对于语义特征的约束过于严格,且上游分类的预训练效果是其零次分类模型的效果上限.本文目标函数由零次分类损失和图结构约束正则项两部分加权求和而成,模型目标函数如公式(5)所示:

$$L=L_s+\lambda L_c \tag{5}$$

其中, λ 是权衡零次分类损失及语义图结构约束正则项部分的超参数.

目标函数的第 1 部分 L_s 是零次分类损失.从有监督学习视角来看,通过使用图像特征矩阵 P 和学得的语义特征矩阵 Q 两个矩阵做点积,待拟合矩阵为图片真实的标签矩阵 R ,直接对应了图像分类任务.在实现过程中,仅要求预测标签矩阵中最高值的位置正确,放宽了约束,且不依赖于上游任务效果.为了对协同过滤效果进行评估,本文选用交叉熵损失函数作为此过程的目标函数,如公式(6)所示:

$$L_s = -\frac{1}{N} \sum_{i=0}^N \sum_{k \in S} y_{i,k} \log(p_i \cdot q_k) = -\frac{1}{N} \sum_{i=0}^N \sum_{k \in S} y_{i,k} \log(f(x_i), g(c_k)) \tag{6}$$

其中:样本 x_i 由图像特征抽取网络 $f(\cdot)$ 提取出的视觉特征 $f(x_i)$, 记作 p_i, N 为训练集样本总量;由图卷积神经网络

$g(\cdot)$ 更新后的标签语义特征向量 $g(c_k)$ 记作 q_k , S 为已知类别集合; $y_{i,k}$ 为指示变量,表示样本 x_i 是否属于类别 k . p_i 与 q_k 相乘表示样本 x_i 属于已知类别 k 的预测概率.

目标函数的第 2 部分 L_c 为图结构约束正则项.通过加入图的结构信息,防止语义知识表示后类间语义关联发生改变.原始语义空间中类别特征及类间语义关联对语义知识表示过程起重要作用,控制着图中节点间信息交互的强弱,是度量已知类别和新类别之间关系的重要参考项.所以,为了保持原始的类别间拓扑结构,防止两个类别在表示学习过程中的语义关联产生变化,在目标函数中加入额外的图结构约束正则项,如公式(7)所示:

$$L_c = \|sim_1 - sim_2\| = \sum_{i=1}^{s \cup u} \sum_{j=1}^{s \cup u} \|sim(c_i, c_j) - sim(g(c_i), g(c_j))\|_2 \quad (7)$$

其中, sim_1 表示原始语义空间中类别间的语义相似度; sim_2 表示经第 3.2 节语义知识表示过程后,类别在高维特征空间中的语义相似度; c_i, c_j 分别表示两个类别节点的初始语义特征; $g(c_i), g(c_j)$ 分别表示经第 3.2 节语义知识表示过程中,图卷积神经网络 $g(\cdot)$ 更新后的节点特征表示. $sim(\cdot, \cdot)$ 的计算同第 3.2 节中类别语义相关系数 $a_{i,j}$ 的计算过程. 差值越小,越能表示在映射后空间中对对应类别仍能维持原始空间中的类别语义关系.

零次学习的推断过程分为两部分完成:首先,使用训练好的图卷积神经网络 $g(\cdot)$ 对语义图进行更新,将所有新类别的语义特征 c_k 转换为对应的高维特征向量 $q_k, k \in N, N$ 为新类别集合;然后,使用卷积神经网络 $f(\cdot)$ 为每个测试样本 x_i' 提取视觉特征 $p_i, i \in N_u$ (在广义零次学习任务中, $i \in N_u \cup N_s$). 将 p_i 与各个新类别语义特征 q_k 做点积,预测其所属概率,取最大值位置作为最终分类,见公式(8):

$$y_i^* = \arg \max_{k \in u} \frac{p_i^* q_k}{\|p_i\|_2 * \|q_k\|_2} \quad (8)$$

4 实验比较

为了评估本文提出方法的有效性,分别在传统零次学习任务和广义零次学习任务上进行实验,并就任务中的各项性能与其他基准方法进行比较.实验结果表明:本文提出的协同过滤式零次学习(CF-ZSL)方法可以在较短的训练时间内普遍提高各任务的分类准确率,能够在不同难度的任务下维持稳定且优异的效果.

4.1 实验数据集和实验设置

以下是本文选用的实验数据集.

- SUN^[34]:细粒度场景图像数据集;
- CUB-200-2011 Birds(CUB)^[35]:细粒度鸟类图像数据集;
- Animals with Attributes 2(AWA2)^[18]:粗粒度动物图像数据集.

表 1 展示了各数据集的统计特征.其中,各数据集除提供图像样本之外,额外提供类别级的属性特征,属性个数见表 1 中属性维度.这些属性由人工标注,标注标准是对于给定的类别,此属性出现的次数百分比.以 AWA2 数据集为例,属性特征囊括了对颜色、条纹以及毛发等的特征描述,是区分不同类别的重要的考量因素.所有数据集均根据零次学习任务的标准划分准则 PS 进行训练集和测试集的划分(PS 划分:负责图像特征抽取的深度卷积神经网络作为零次学习模型的一部分,通常使用 ImageNet 图像数据集^[36]进行预训练,为了保证实验的公平性,经 PS 划分后的测试集类别均不出现在 ImageNet 数据集之中).

Table 1 Statistics of zero-shot learning data sets

表 1 零次学习数据集统计特征

数据集	属性维度	训练类别数量	测试类别数量	图片数量
SUN	102	645	72	14 340
CUB	312	150	50	11 788
AWA2	85	40	10	37 322

传统零次学习和广义零次学习的任务设置不同,广义零次学习任务在测试阶段包含一定数量的训练类别样本,其对应的训练集和测试集的划分方式不同,具体划分情况见表 2.相比于传统零次学习,原有的训练样本转换为测试样本,导致训练样本数量减少.因而,广义零次学习更具有挑战性,对模型的要求更高.

Table 2 Data set partition

表 2 数据集划分

数据集	传统零次学习		广义零次学习		
	训练样本量	测试样本量	训练样本量	测试样本量(已知类别)	测试样本量(新类别)
SUN	12 892	1 439	10 314	2 578	1 439
CUB	8 814	2 966	7 052	1 762	2 966
AWA2	29 408	7 913	23 526	5 882	7 913

实验基于 Ubuntu 16.04 操作系统及一块 GeForce GTX 1080 Ti GPU,使用 Pytorch 深度学习框架实现.对于模型的视觉特征抽取部分,采用在 ImageNet 数据集上预训练好的 ResNet50^[37]网络作为视觉特征抽取网络.为了增强在不同数据集上的适应能力,在抽取特征之前,使用各数据集的训练集样本对其进行微调,以适应当前任务.语义知识表示部分,本文选用两层图卷积神经网络完成图上的信息传递,其输入特征维度为类别的初始属性维度,邻接矩阵为构造好的类别语义关系矩阵,中间隐层维度为超参数,输出的视觉特征维度为 $d, d=2048$.本文将在第 4.8 节对图卷积神经网络的层数设置进行对比实验,借此说明将图卷积神经网络层数选择为 2 的原因.每层网络使用概率系数为 0.5 的 Dropout 层以及批量归一化算法进行网络解耦,并加入 ReLU 激活函数进行非线性转换,表 3 中为不同数据集下具体的图神经网络结构.模型训练过程中,网络权重使用 Xavier 方法进行初始化,学习率初始化为 0.005.在 1 000 次迭代之后,学习率调整为 0.001,并使用 Adam 梯度下降算法进行权重更新.

Table 3 Architecture of graph neural network

表 3 图神经网络结构

网络层	SUN	CUB	AWA2
L1	输入维度:717×102 邻接矩阵:717×717 GCN,BN,ReLU 输出维度:717×256	输入维度:200×312 邻接矩阵:200×200 GCN,BN,ReLU 输出维度:200×512	输入维度:50×85 邻接矩阵:50×50 GCN,BN,ReLU 输出维度:50×128
L2	输入维度:717×256 邻接矩阵:717×717 GCN,BN,ReLU 输出维度:717×2048	输入维度:200×512 邻接矩阵:200×200 GCN,BN,ReLU 输出维度:200×2048	输入维度:50×128 邻接矩阵:50×50 GCN,BN,ReLU 输出维度:50×2048
输出	特征矩阵:717×2048	特征矩阵:200×2048	特征矩阵:50×2048

4.2 对比算法

为了评估本文提出方法的有效性,选取多个具代表性的经典零次学习算法以及近年提出的新算法进行比较,各基准算法的提出时间以及主要思想见表 4.

Table 4 Baseline algorithms

表 4 基准算法

基准方法	发表时间	主要思想
DAP/IAP ^[38]	2009	零次学习领域开创方法,学习一组视觉空间中的属性分类器
DEWISE ^[39]	2013	使用预训练的词向量特征与图像视觉特征进行相似度度量
CONSE ^[40]	2013	组合类标签向量,将图像特征嵌入语义空间
CMT ^[41]	2014	提出对双线性兼容性学习框架的非线性扩展
ALE ^[42]	2015	提出针对给定图像,正确类排序高于不正确类的兼容性函数
SJE ^[43]	2015	利用多种模态的语义信息相互补充
ESZSL ^[44]	2015	将正则化项加入线性兼容方法中,对特征重构过程进行惩罚
SSE ^[45]	2015	将源数据和目标数据混合建模,属于直推式零次学习设置
LATEM ^[46]	2016	使用线性函数模型集合,针对不同类别自动选择最好的线性函数模型
SYNC ^[12]	2016	使用虚构类学习模型空间和语义类嵌入空间之间的映射
SAE ^[13]	2017	引入自编码器结构,加入视觉特征或语义原型的重构约束
PSR ^[11]	2018	在自编码器结构基础上,显式加入语义类别之间的关系约束
ZSKL ^[47]	2018	应用核方法学习视觉特征和属性空间之间的非线性映射
MLSE-ZSL ^[48]	2019	提出潜在语义编码器,该编码器利用视觉特征和潜在语义表示进行学习
SABR ^[49]	2019	使用生成对抗网络,学习一个既有区分度又语义相似的嵌入空间

4.3 评价标准

基于类别的分类准确率(per-class classification accuracy)是进行零次学习算法效果评价的常用指标,其优点是能有效降低单个类别对整体结果产生的影响,计算过程见公式(9):

$$ACC_c = \frac{1}{\|C\|} \sum_{c \in C} \frac{\# \text{ correct predictions in } c}{\# \text{ samples in } c} \quad (9)$$

其中, C 为待测试集中的全部类别.

对于广义零次学习任务,采用已知类别准确率(u)、未知类别准确率(s)、调和准确率(h)这三者作为评价指标.三者的计算过程如下:其中, ACC_{tr}, ACC_{te} 分别表示通过公式(9)计算出的已知类别准确率和未知类别准确率,调和准确率计算过程见公式(10):

$$H_{acc} = \frac{2 \times ACC_{tr} \times ACC_{te}}{ACC_{tr} + ACC_{te}} \quad (10)$$

4.4 传统零次学习任务实验结果

传统零次学习任务中,未知类别分类准确率指标侧重考查模型对新类别样本的分类准确性,各项对比算法在 SUN,CUB 和 AWA2 数据集上的最终结果如表 5 所示.本文采用相同的实验设置以及评估原则,来保证比较的公平性.从表中可以观察到,本文提出的 CF-ZSL 算法的分类准确率要优于所有对比算法.这表明通过协同过滤过程能够充分学习图像视觉特征与类别语义信息之间的潜在影响关系,优化零次学习建模的效果,提升模型对新类别的分类能力.

Table 5 Compare the classification accuracy under the traditional ZSL setting (%)

表 5 比较传统零次学习任务分类准确率 (%)

方法	SUN	CUB	AWA2
DAP	39.9	40.0	46.1
IAP	19.4	24.0	35.9
CONSE	38.8	34.3	44.5
CMT	39.9	34.6	37.9
SSE	51.5	43.9	61.0
LATEM	55.3	49.3	55.8
ALE	58.1	54.9	62.5
DEVISE	56.5	52.0	59.7
SJE	53.7	53.9	61.9
ESZSL	54.5	53.9	58.6
SYNC	56.3	55.6	46.6
SAE	40.3	33.3	54.1
PSR	61.4	63.8	56.0
ZSKL	70.5	61.7	51.7
MLSE	62.8	64.2	67.8
SABR	62.8	63.9	65.2
CF-ZSL	71.1	66.2	69.2

能够在不同粒度数据集上有均衡表现,也是对零次学习算法能力的一种考量.粒度刻画的是数据集中各类别的差异程度,例如:AWA2 是粗粒度动物数据集,各类别是不同种类的动物(如狮子、海豹和蝙蝠等),彼此之间差异较大;而 CUB 是细粒度鸟类数据集,各个类别是细化分的不同鸟类.从表中可以观察到:无论是在粗粒度数据集 AWA2 还是在细粒度数据集 SUN 和 CUB 上,CF-ZSL 算法的准确率均有明显提升.实验结果表明:本文提出的语义图加图神经网络的语义表示过程能够有效挖掘类别间语义关联,受类别粒度的影响较小,在各粒度数据集上显示出均衡且优秀的表现.

4.5 广义零次学习任务实验结果

广义零次学习任务描述一种更通用的场景,测试集不仅包含新类别样本,还包含一些从原始训练集划分出的已知类别样本,且每个样本的预测范围为包含已知类别 S 和新类别 U 的所有类别.相较于传统零次分类任务,广义零次分类任务难度较高,要求模型具有更强的泛化能力.表 6 展示了广义零次学习任务的实验结果,其中, u ,

s, h 分别表示未知类别分类准确率、已知类别分类准确率以及调和准确率.有些基准算法仅针对传统零次学习任务,表中所列基准算法均为进行过广义零次学习任务的算法.从表中可以看出:所有基准模型很难在已知类别和新类别的准确率上达到一个均衡的表现,并且新类别的准确率也均低于其在传统零次学习任务中的结果.例如:CONSE 方法在已知类别的分类准确率(s)大幅超过其在新类别的分类准确率(u),且 u 值非常低,几乎无法对新类别样本进行正确预测.这种现象是由实验性质导致的,由于每张样本都要向所有类别去做预测,而已知类别和新类别中可能存在非常相似的类别,以 AWA2 数据集为例,新类别中的蓝鲸与已知类别中的虎鲸在视觉特征上较为相似,这种情况提升了分类的难度,更加考验算法能否有效准确地区分已知类别和新类别中差异较小的那些类别.CF-ZSL 方法在这种实验设置下,相较于各个基准模型仍可获得较好的结果;同时,在未知类别准确率指标上总能达到更优的结果.

Table 6 Compare the classification accuracy under the generalized ZSL setting (%)

表 6 比较广义零次学习任务分类准确率 (%)

方法	SUN			CUB			AWA2		
	u	s	h	u	s	h	u	s	h
CONSE	6.8	39.9	11.6	1.6	72.2	3.1	0.5	90.6	1.0
SSE	2.1	36.4	4.0	8.5	46.9	14.4	8.1	82.5	14.8
DEVISE	16.9	27.4	20.9	23.8	53.0	32.8	17.1	74.7	27.8
SJE	14.7	30.5	19.8	23.5	59.2	33.6	8.0	73.9	14.4
ESZSL	11.0	27.9	15.8	12.6	63.8	21.0	5.9	77.8	11.0
SYNC	7.9	43.3	13.4	11.5	70.9	19.8	10.0	90.5	18.0
ALE	21.8	33.1	26.3	23.7	62.8	34.4	14.0	81.8	23.9
PSR	20.8	37.2	26.7	24.6	54.3	33.9	20.7	73.8	32.3
ZSKL	20.1	31.4	24.5	24.2	63.9	35.1	18.9	82.7	30.8
MLSE	20.7	36.4	26.4	22.3	71.6	34.0	23.8	83.2	37.0
CF-ZSL	29.4	45.8	35.8	36.3	72.9	48.5	33.3	82.0	47.4

4.6 基于局部特征的零次学习实验

现有零次学习方法多数基于整张图片提取的全局特征作为视觉特征,但是全局特征着重捕获图像中整体信息,忽略了局部区域的判别性差异,使用其表示细粒度图像可能会导致次优结果.与之相比,局部区域特征更关注图像中的一些重点区域,可能会捕获与类别语义描述相关的更多局部信息,特征更具判别性.在细粒度 CUB 数据集中,局部区域由多位领域专家进行标注.图 4 为 CUB 数据集中标注的局部区域,一般可选取背部、喙、腹部、胸部、腿部、翅膀和尾部这 7 个主要局部区域用于概括整张图片图像的视觉特征.

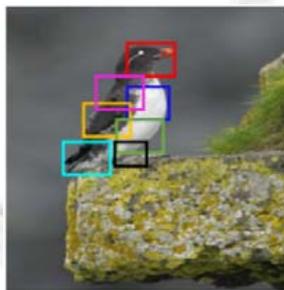


Fig.4 Local regions in CUB image^[35]

图 4 CUB 数据集图像局部区域^[35]

本文希望尝试使用更加细粒度的局部特征进行零次学习实验,结果见表 7.本文在这里使用的对比算法均为使用局部特征的零次学习方法,是近两年发表在计算机视觉领域顶级会议上的代表性工作.从表中可以看出:相较于使用全局特征的 CF-ZSL 方法,实验结果由 66.2% 提升至 80.2%,实现了 14% 的准确率提升.而相较于其他使用局部特征的基准算法,CF-ZSL 方法能够达到最高的准确率指标.由此可证明:针对 CUB 这类细粒度数据集,零次学习模型应该使用更加细粒度的局部特征作为图像视觉特征,识别效果更加精准.

Table 7 Compare the classification accuracy under the local region feature based ZSL setting (%)

表 7 比较基于局部区域特征的零次学习分类准确率 (%)

方法	准确率
SGA ^[50] (local)	75.3
SGMA ^[51] (local)	71.0
RGEM ^[52] (local)	76.1
CF-ZSL (global)	66.2
CF-ZSL (local)	80.2

4.7 消融实验

4.7.1 映射函数实验分析

为了验证语义图以及图神经网络的作用,本节与使用多层感知机这种类别独立更新的方法进行对比。两组实验在相同的参数设定下各自进行 10 组实验,并对实验结果取最大值作为最终结果,实验结果见表 8。可以看出:使用图卷积神经网络作为语义知识表示模型能够获得更好的分类准确率,本质上能为各个类别学得更恰当的特征表示。本文分析此实验结果背后的原因是:由于多层感知器模型在进行特征映射的过程中,各类别的更新过程是相互独立、互不影响的,已知类别和新类别间的信息不能很好地进行传递;而图卷积神经网络从图结构数据出发,可以在非欧式空间中进行消息传递;此外,图卷积神经网络表现出拉普拉斯平滑的作用,可为相似类别学得相似的特征表示,使得最终分类变得更加准确。

Table 8 Comparative result between GCN and MLP (%)

表 8 GCN 与 MLP 的准确率对比结果 (%)

实验设置	SUN	CUB	AWA2
MLP	55.3	54.1	51.2
GCN	71.1	66.2	69.2

4.7.2 图结构约束正则项实验分析

为了验证公式(6)中图结构约束正则项对实验结果的影响,本节对比不加入图结构约束的传统零次学习实验(即损失函数仅为公式(5)中的 L_s),对算法进行评价。两组实验在相同的参数设定下各自进行 10 组实验,取实验结果最大值作为最终结果。表 9 中,加入图结构约束之后,算法的分类性能有一定程度上的提升,证明了原始语义空间中的类别语义以及类间语义关联对整个语义知识表示过程起到重要作用。

Table 9 Ablation study of graph structure constraint regularization (%)

表 9 图结构约束正则项消融实验 (%)

实验设置	SUN	CUB	AWA2
无约束正则项	68.6	64.9	67.8
有约束正则项	71.1	66.2	69.2

4.7.3 图神经网络结构分析

为了验证语义知识表示部分图神经网络结构对模型准确率的影响,本节使用不同层数的图神经网络结构进行实验,并通过尝试不同的隐层维度,将相同层数下的最优值作为最终结果,见表 10。

Table 10 Effect of graph neural network structure (%)

表 10 图神经网络结构影响 (%)

实验设置	SUN	CUB	AWA2
1 层 GCLayer	69.8	54.1	70.8
2 层 GCLayer	71.1	66.2	69.2
3 层 GCLayer	66.3	63.4	66.1
4 层 GCLayer	57.9	60.8	64.0
5 层 GCLayer	56.7	57.1	61.7

实验结果表明:对于 CUB 和 SUN 数据集而言,GCN 层数过少导致知识不能很好地通过图来传播(1 层 GCN 就只能利用它的邻接节点);而随着 GCN 层数增加,图中节点之间传递消息的次数增加,可能会导致较远节点产生的知识稀释问题;同时,模型参数量随之增加,容易造成过拟合,提升了训练难度.而在 AWA2 数据集上,一层 GC Layer 效果比较好.原因可能是 AWA2 的语义特征矩阵的类别维度和属性维度都最小(50×85),随着 GCN 层数增加,节点包含的语义信息会被迅速稀释;此外,AWA2 是唯一的粗粒度数据集,类间差异相对较大,GCN 层数增加也可能导致原本差异明显的类别变得相似.

4.8 收敛速度实验分析

图 5 是损失函数收敛曲线.横坐标代表训练迭代次数,纵坐标代表损失函数值.从图中可以看出:随着训练的进行,损失函数值迅速降低并收敛.在迭代 500 次时,AWA2,CUB,SUN 这 3 个数据集上的测试分类准确率分别为 64.8%、63.8% 和 67.8%,均达到近似最优的测试分类准确率.经分析,本模型中需要学习的参数只涉及语义知识表示模块中的图神经网络,参数量较少,降低了模型收敛难度,这也是本文方法在各数据集上表现出强泛化能力的主要原因.

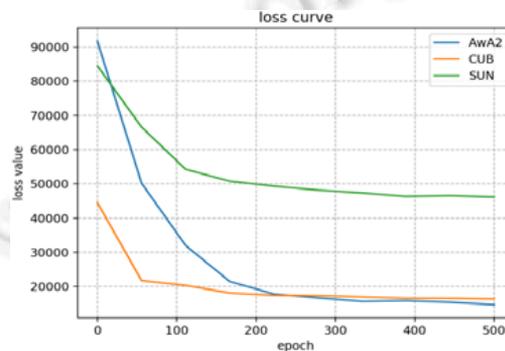


Fig.5 Convergence curve of loss value

图 5 损失函数收敛曲线

4.9 模型误分类图像分析

为了进一步检验算法的有效性,对 AWA2 数据集中被模型误分类的部分样本进行检查.如图 6 所示,左侧一列是能够被正确分类的图片,右侧三列是一些误分类的图片.



Fig.6 Misclassified images

图 6 误分类图像

从图中可以看出:右侧这些图像不能很明显地体现所属类别的判别性特征,与该类别的常规图像特征以及类别属性描述都差异较大,即使是从人类的认知角度,这些样本也确实很难被正确分类。

5 总结

本文面向零次图像分类任务提出了一种协同过滤式零次学习方法,通过借鉴推荐领域中的协同过滤思想,将零次学习过程建模为一个矩阵填充问题,建立了零次学习领域与推荐领域的桥梁。根据已知类别标签矩阵提供的丰富信息推断图像及类别的隐特征表示,从而使用视觉特征矩阵和语义特征矩阵重构标签矩阵,实现对新类别样本的分类。此外,通过构建类别语义图来建立类别间的语义关联,将已知类别知识迁移至新类别,并应用图卷积神经网络更新节点,为每个节点学得更好的特征表示。本文方法是端到端的轻量级模型,迭代 300 次~500 次即可达到近似最优的测试准确率。实验结果表明:在传统零次学习任务以及广义零次学习任务上,本文提出的 CF-ZSL 方法在 AWA2, CUB 和 SUN 这 3 个零次学习数据集上均能达到稳定且优秀的实验结果。通过设计不同的损失函数或采用不同的协同过滤模型,可能会更好地发挥协同过滤算法的优势。我们相信:在融合推荐领域和零次学习领域的方向上,仍存在很多潜力有待挖掘。

致谢 感谢赖永老师在论文修改过程中提出的建设性意见,感谢夏日婷、于东然、刘丁菠以及李俊达同学对本文工作提出的宝贵建议。

References:

- [1] Krizhevsky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2012,25:1097–1105.
- [2] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2016. 770–778.
- [3] Hinton G, Deng L, Yu D, Dahl GE, Mohamed A, Jaitly N, Senior A, Vanhoucke V, Nguyen P, Sainath TN, Kingsbury B. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal processing Magazine*, 2012,29(6):82–97.
- [4] Graves A, Mohamed A, Hinton G. Speech recognition with deep recurrent neural networks. In: *Proc. of the IEEE Int'l Conf. on Acoustics, Speech and Signal Processing*. 2013. 6645–6649.
- [5] Fagnant D, Kockelman K. Preparing a nation for autonomous vehicles: Opportunities, barriers and policy recommendations. *Transportation Research Part A: Policy and Practice*, 2015,77:167–181.
- [6] Yao Q, Tian Y. A model predictive controller with longitudinal speed compensation for autonomous vehicle path tracking. *Applied Sciences*, 2019,9(22):4739.
- [7] Xian Y, Schiele B, Akata Z. Zero-shot learning—the good, the bad and the ugly. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 4582–4591.
- [8] Wang X, Ye Y, Gupta A. Zero-shot recognition via semantic embeddings and knowledge graphs. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 6857–6866.
- [9] Su X, Khoshgoftaar T. A survey of collaborative filtering techniques. In: *Proc. of the Advances in Artificial Intelligence*. 2009.
- [10] Koren Y, Bell R. *Advances in Collaborative Filtering. Recommender Systems Handbook*. 2015. 77–118.
- [11] Annadani Y, Biswas S. Preserving semantic relations for zero-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2018. 7603–7612.
- [12] Changpinyo S, Chao WL, Gong B, Sha F. Synthesized classifiers for zero-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2016. 5327–5336.
- [13] Kodirov E, Xiang T, Gong S. Semantic autoencoder for zero-shot learning. In: *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*. 2017. 3174–3183.
- [14] Fu Y, Hospedales T, Xiang T, Gong S. Transductive multi-view zero-shot learning. *IEEE Trans. on Pattern Analysis and Machine Intelligence*. 2015,37(11):2332–2345.
- [15] Fu Z, Xiang T, Kodirov E, Gong S. Zero-shot learning on semantic class prototype graph. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2017,40(8):2009–2022.

- [16] Kodirov E, Xiang T, Fu Z, Gong S. Unsupervised domain adaptation for zero-shot learning. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 2452–2460.
- [17] Chen L, Zhang H, Xiao J, Liu W, Chang S. Zero-shot visual recognition using semantics-preserving adversarial embedding networks. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 1043–1052.
- [18] Lampert CH, Nickisch H, Harmeling S. Attribute-based classification for zero-shot visual object categorization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2013,36(3):453–465.
- [19] Li Y, Zhang J, Zhang J, Huang K. Discriminative learning of latent features for zero-shot recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 7463–7471.
- [20] Radovanovic M, Nanopoulos A, Ivanovic M. Hubs in space: Popular nearestneighbors in high-dimensional data. *Journal of Machine Learning Research*, 2010,11(Sept.):2487–2531.
- [21] Wang X, Ye Y, Gupta A. Zero-shot recognition via semantic embeddings and knowledge graphs. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 6857–6866.
- [22] Zhang L, Xiang T, Gong S. Learning a deep embedding model for zero-shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2017. 2021–2030.
- [23] Changpinyo S, Chao WL, Gong B, Sha F. Classifier and exemplar synthesis for zero-shot learning. *Int'l Journal of Computer Vision*, 2020,128(1):166–201.
- [24] Changpinyo S, Chao WL, Sha F. Predicting visual exemplars of unseen classes for zero-shot learning. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 3476–3485.
- [25] Tsai H, Huang L, Salakhutdinov R. Learning robust visual-semantic embeddings. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2017. 3571–3580.
- [26] Liu S, Long M, Wang J, Jordan M. Generalized zero-shot learning with deep calibration network. In: Proc. of the Advances in Neural Information Processing Systems. 2018. 2005–2015.
- [27] Goldberg D, Nichols D, Oki B, Terry D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 1992,35(12):61–70.
- [28] Linden G, Smith B, York J. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 2003, 7(1):76–80.
- [29] Pazzani MJ, Billsus D. Content-based recommendation systems. In: Proc. of the Adaptive Web. Berlin, Heidelberg: Springer, 2007. 325–341.
- [30] Deng AL, Zhu YY, Shi BL. A collaborative filtering recommendation algorithm based on item rating prediction. *Ruan Jian Xue Bao/Journal of Software*, 2003,14(9):1621–1628 (in Chinese with English abstract). <http://www.jos.org.cn/1000-9825/14/1621.htm>
- [31] Ungar L, Foster D. Clustering methods for collaborative filtering. In: Proc. of the AAAI Workshop on Recommendation Systems, Vol.1. 1998. 114–129.
- [32] Vozalis M, Margaritis K. Using SVD and demographic data for the enhancement of generalized collaborative filtering. *Information Sciences*, 2007,177(15):3017–3037.
- [33] Kampffmeyer M, Chen Y, Liang X, Wang H, Zhang Y, Xing E. Rethinking knowledge graph propagation for zero-shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2019. 11487–11496.
- [34] Xiao J, Hays J, Ehinger K, Oliva A, Torralba A. Sun database: Large-scale scene recognition from abbey to zoo. In: Proc. of the 2010 IEEE Computer Society Conf. on Computer Vision and Pattern Recognition. IEEE, 2010. 3485–3492.
- [35] Wah C, Branson S, Welinder P, Perona P, Belongie S. The caltech-ucsbirds-200-2011 dataset. Technical Report, CNS-TR-2011-001, 2011.
- [36] Deng J, Dong W, Socher R, Li L, Li K, Li F. Imagenet: A large-scale hierarchical image database. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2009. 248–255.
- [37] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 770–778.
- [38] Lampert CH, Nickisch H, Harmeling S. Learning to detect unseen object classes by between-class attribute transfer. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. IEEE, 2009. 951–958.
- [39] Frome A, Corrado G, Shlens J, Bengio S, Dean J, Mikolovet T. Devise: A deep visual-semantic embedding model. In: Proc. of the Advances in Neural Information Processing Systems. 2013.
- [40] Norouzi M, Mikolov T, Bengio S, Singer Y, Shlens J, Frome A, Corrado G, Dean J. Zero-shot learning by convex combination of semantic embeddings. arXiv preprint arXiv:1312.5650, 2013.

- [41] Socher R, Ganjoo M, Sridhar H, Bastani O, Manning C, Ng A. Zero shot learning through cross-modal transfer. In: Proc. of the Advances in Neural Information Processing Systems. 2014.
- [42] Akata Z, Perronnin F, Harchaoui Z, Schmid C. Label-embedding for image classification. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2015,38(7):1425–1438.
- [43] Akata Z, Reed S, Walter D, Lee H, Schiele B. Evaluation of output embeddings for fine-grained image classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2015. 2927–2936.
- [44] Romera-Paredes B, Torr P. An embarrassingly simple approach to zero-shot learning. In: Proc. of the Int'l Conf. on Machine Learning. PMLR, 2015. 2152–2161.
- [45] Zhang Z, Saligrama V. Zero-shot learning via semantic similarity embedding. In: Proc. of the IEEE Int'l Conf. on Computer Vision. 2015. 4166–4174.
- [46] Xian Y, Akata Z, Sharma G, Nguyen Q, Hein M, Schiele B. Latent embeddings for zero-shot classification. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2016. 69–77.
- [47] Zhang H, Koniusz P. Zero-shot kernel learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2018. 7670–7679.
- [48] Ding Z, Liu H. Marginalized latent semantic encoder for zero-shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2019. 6191–6199.
- [49] Paul A, Krishnan NC, Munjal P. Semantically aligned bias reducing zero shot learning. In: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition. 2019. 7056–7065.
- [50] Yu Y, Ji Z, Fu Y, Guo J, Pang Y, Zhang Z. Stacked semantics-guided attention model for fine-grained zero-shot learning. In: Advances in Neural Information Processing Systems. 2018. 5995–6004.
- [51] Zhu Y, Xie J, Tang Z, Peng X, Elgammal A. Semantic-guided multi-attention localization for zero-shot learning. In: Advances in Neural Information Processing Systems. 2019.
- [52] Xie G, Liu L, Zhu F, Zhao F, Zhang Z, Yao Y, Qin J, Shao L. Region graph embedding network for zero-shot learning. In: Proc. of the European Conf. on Computer Vision. Cham: Springer-Verlag, 2020. 562–580.

附中文参考文献:

- [30] 邓爱林,朱扬勇,施伯乐.基于项目评分预测的协同过滤推荐算法.软件学报,2003(9):1621–1628. <http://www.jos.org.cn/1000-9825/14/1621.htm>



杨博(1974—),男,博士,教授,博士生导师,CCF 杰出会员,主要研究领域为知识发现与知识工程,网络分析理论与应用,推荐系统,多智能体系统.



张春旭(1996—),女,学士,主要研究领域为数据挖掘,网络表示学习.



张钰雪晴(1997—),女,硕士,主要研究领域为零次学习,深度学习.



黄晶(1975—),女,博士,副教授,博士生导师,CCF 专业会员,主要研究领域为大规模网络数据挖掘与学习,智能大数据处理,复杂网络分析,深度学习,多 Agent 系统,数据驱动的智能传染病防控.



彭羿达(1996—),男,学士,主要研究领域为计算机视觉.