中国科学: 信息科学 2024 年 第 54 卷 第 6 期: 1386–1399
SCIENTIA SINICA Informationis
论文

属性知识自反绎下的半监督表示学习

沈阳[1], 孙旭豪[1], 徐赫洋[1], 魏秀参[2,3*]

1. 南京理工大学计算机科学与工程学院, 南京 210094
2. 东南大学计算机科学与工程学院, 南京 210096
3. 新一代人工智能技术与交叉应用教育部重点实验室 (东南大学), 南京 210096
* 通信作者. E-mail: weixs.gm@gmail.com

收稿日期: 2023–08–29; 修回日期: 2023–11–08; 接受日期: 2023–11–30; 网络出版日期: 2024–06–07

国家重点研发计划青年科学家项目 (批准号: 2021YFA1001100)、国家自然科学基金面上项目 (批准号: 62272231)、江苏省自然科学基金青年基金项目 (批准号: BK20210340)、中央高校基本科研业务费专项资金项目 (批准号: 4009002401) 和中国人工智能学会 – 华为 MindSpore 学术奖励基金项目 (批准号: CAAIXSJLJJ-2022-001B) 资助

摘要 机器学习结合逻辑推理的方法可以大幅提升模型的鲁棒性与可解释性. 近年来, 已有工作从给定的具体知识库出发, 通过反绎学习的范式或是其衍生范式来促进机器学习中模型的更新过程. 然而, 在表示学习任务中, 即便存在这样的知识库, 其往往也是不完备或含有噪声的. 且在真实环境下, 即便领域专家也无法精准定量地描述不同对象的属性表示信息. 因此, 本文针对半监督表示学习任务, 提出了一种可根据少量有标记样本构建弱领域属性知识库并结合无标记数据与基于启发式规则扩张领域知识库推理的反绎学习方法. 该方法可有效解决表示学习任务下缺少强领域知识与真实环境下高质量标注数据较少这两个问题. 在人工合成的数据集与真实环境下的数据集中的实验对比结果均验证了我们提出的方法的有效性.

关键词 人工智能, 机器学习, 反绎学习, 半监督学习, 特征表示, 细粒度属性

1 引言

机器学习中, 表示学习模型训练的过程本质上是模型从海量数据中寻找有意义的统计特征的过程, 而模型性能的优劣则直接由统计特征的好坏决定. 然而, 仅通过数据驱动的方式完成模型的构建并进行训练, 需要大量的高质量有标记数据, 这在实际应用中往往需要昂贵的代价 [1], 例如油气管网缺陷检测 [2]. 因此, 近年来研究者通过尝试利用无标记数据辅助模型训练的半监督方法 [3] 来降低表示学习模型对标记数据的依赖.

另一方面, 以往的研究者认为仅通过数据驱动训练获得的表示学习模型缺乏可解释性 [4]. 例如在细粒度表示学习任务中 [5], 人们可以通过 “钩状的喙”、“白色的前额”、“米黄色的冠” 等属性信息的有

引用格式: 沈阳, 孙旭豪, 徐赫洋, 等. 属性知识自反绎下的半监督表示学习. 中国科学: 信息科学, 2024, 54: 1386–1399, doi: 10.1360/SSI-2023-0252
Shen Y, Sun X H, Xu H Y, et al. Attribute-aware knowledge based self-abductive for semi-supervised representation learning (in Chinese). Sci Sin Inform, 2024, 54: 1386–1399, doi: 10.1360/SSI-2023-0252

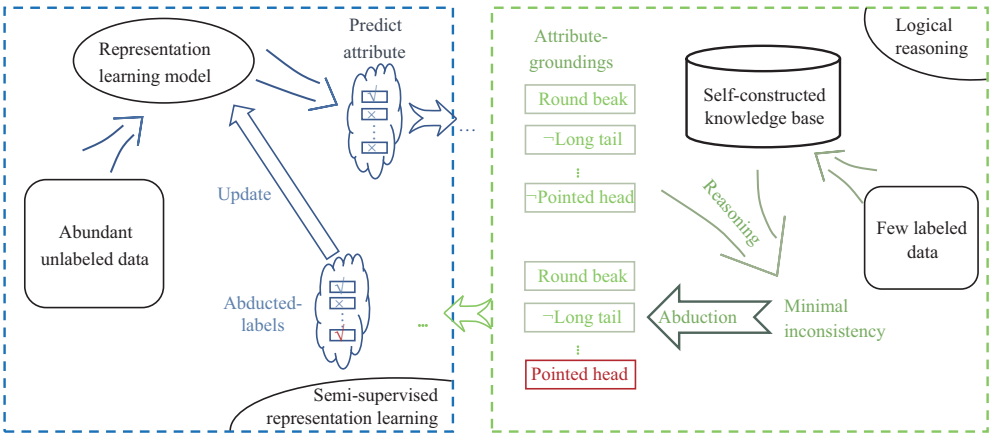

图 1 (网络版彩图) 属性知识自反绎下的半监督表示学习框架

Figure 1 (Color online) Attribute-aware knowledge based self-abductive learning framework for semi-supervised representation learning

机组合来识别一只 “黑脚信天翁”, 但却无法得知模型是否确实关注到了这些属性信息, 或是无法得知模型的哪个模块提取到了这样的属性信息. 因此, 越来越多的学者考虑将由符号表示的领域知识引入到机器学习的模型中 [6, 7]. 这样的将数据驱动的机器学习方法和由知识驱动的逻辑溯因集成在一个统一的框架下的学习范式也被认为是下一代人工智能的关键研究领域之一 [8].

反绎学习 (abductive learning, ABL) [7, 9] 即为该范式下的一种机器学习框架. 它主要包含感知与逻辑溯因两个模块, 其中感知模块可以是任何常见的表示学习模型, 用于从原始输入中提取任务所需的高维特征表达, 也可以称为从原始数据中获取具体事实信息; 逻辑溯因模块则通过额外的强领域知识进行逻辑推理, 一般强领域知识由一阶逻辑表达构成 (知识通常指先验知识、领域知识或背景信息, 也可以是关于特定领域的规则、模式、先前观察到的数据或经验. 知识可以指导推理过程, 帮助推理得到最可能的解释或假设). 然而, 在实际应用中, 极大部分情况下无法获取这样的由一阶逻辑表示的充分的强领域知识, 只能获取由具体事实表示的知识或不充分的弱领域知识 [8]. 同样地, 以实际应用中的表示学习任务为例, “钩状的喙”、“白色的前额” 等属性信息均可以认为是 “黑脚信天翁” 的弱属性知识, 即该鸟类可能包含这些属性信息. 但由于数据本身的噪声或鸟类在自然环境下生长状态的变化, 完备的属性知识标注需要大量领域专家通过海量数据样本才可以获得, 无法满足实际应用的需求. 因此, 本文提出了一种可以根据少量有标记样本构建弱领域属性知识库, 并通过无标记数据与反绎学习优化表示学习模型的方法.

图 1 概述了本文方法的框架. 具体而言, 首先模型以自监督学习的方式提取所有有标记样本的属性信息, 并将不同类别相同部位的属性进行对齐, 以此构建弱领域属性知识库. 接着, 对于输入的无标记数据, 需要通过待优化的表示学习模型完成属性预测, 并将得到的属性信息约束为弱领域属性知识库可以识别的表示. 而后通过基于启发式规则扩张领域知识库的属性表示反绎学习方法 (ground abductive learning for attribute representation, AR-GABL), 依据最小不一致原则修改错误的属性预测, 并对相应的无标记数据添加反绎标记. 最后, 通过带有反绎标记的数据更新表示学习模型.

本文的其余部分组织如下: 第 2 节介绍反绎学习与半监督表示学习的相关方法与研究现状; 第 3 节介绍本文提出的基于属性知识自反绎的半监督表示学习方法的问题定义与具体步骤; 第 4 节通过对比实验验证了提出方法的有效性; 第 5 节总结全文.

2 相关工作

2.1 反绎学习

近年来, 一些研究者提出建立一种混合模型, 希望将机器学习的感知模型和逻辑推理的推理模型相结合. 这样的代表性框架包括反绎学习 [7, 10, 11] 与 DeepProbLog [12] 等. 以反绎学习为例, 其中的反绎推理是连接机器学习模型和领域知识的纽带, 反绎推理根据由一阶逻辑语言表示的强领域知识对机器学习模型预测的结果进行修改. 反绎学习利用反绎推理修改后的结果更新机器学习模型, 提高机器学习模型性能, 如此往复.

在传统的监督学习中, 训练机器学习模型需要使用大量的有标记数据. 反绎学习的应用场景与传统的监督学习不同, 反绎学习假设难以取得足够数量的监督数据对模型进行充分训练, 但是可以获取充足的无标记数据和任务的领域知识. 然而, 对于具体任务, 我们往往无法获得充足的领域知识, 且这些知识无法构成传统反绎学习要求的一阶逻辑表达. 因此, 在此基础上提出了基于启发式规则扩张领域知识库的反绎学习方法 (grounded abductive learning, GABL) [8], 该方法在领域知识库中增广了一条与具体领域知识无关的反绎推理规则, 该规则从弱领域知识库中寻找与预测结果最为接近的具体事实作为推理结果. 然而, 这些改进的反绎学习方法仍然需要较为完备的领域知识作为支撑. 在本文涉及的表示学习任务中, 人们又难以通过先验的领域知识对机器学习的表征作出改进. 因此, 本文提出基于自监督学习的方式, 通过少量的有标记样本进行弱领域知识库的自构建.

2.2 半监督表示学习

半监督学习 (semi-supervised learning, SSL) 旨在使用给定的少量有标记样本, 并结合大量未标记的数据完成模型的训练任务 [3, 13~17]. 通常, 大规模高质量的标记数据可以帮助模型获得更高的性能, 然而, 这些有标记的样本在一些任务中是非常昂贵的. 例如, 在医疗图像识别任务中, 通常需要使用昂贵的设备进行测量后由多个领域专家共同分析得到标记. 而半监督学习的出现则大幅降低了对高质量有标记数据量的需求. 表示学习则旨在通过机器学习模型找到比原始数据更好的高维特征表达. 借助良好的特征表示, 机器学习模型可以完成后续的分类、聚类、回归等任务. 本文主要研究通过表示学习完成相关分类任务.

现有的半监督表示学习方法一般可以划分为两大类: 基于传统机器学习的方法与基于深度学习的方法. 而基于深度学习的半监督表示学习方法则可以进一步划分为 5 类.

• 基于数据生成的方法 [18]: 从有标记的训练数据集中建模数据的分布, 而后依据这样的数据分布, 通过对抗生成网络 [19] 等方法生成新的训练数据.

• 一致性正则化方法 (consistency regularization) [20]: 此类方法认为一个输入在微小扰动后, 其预测应该是一致的. 因此主要通过无标记数据找到数据集所在的平滑流形 (smooth manifold).

• 基于图的方法 [21]: 从原始数据集中提取一个图, 图中的每个节点代表一个训练样本, 每条边表示节点对的相似性度量.

• 伪标记方法 [22]: 通过对无标记数据添加伪标记的形式将数据添加至训练集, 并依此完成模型的更新.

• 混合方法: 将上述 4 种方法中的部分进行混合优化.

本文用于半监督表示学习的方法属于伪标记类的方法. 与以往添加伪标记的方法不同, 本文对无标记数据添加的伪标记通过生成的弱属性知识库的反绎推理获得, 即伪标记需要通过属性知识库的推

演得到, 而对于无法推理的无标记数据, 则不使用这些数据完成模型的更新过程. 因此本文方法具有更强的可解释性.

3 基于属性知识自反绎的半监督表示学习

3.1 问题定义

本文的主要目标是通过自监督的方式生成训练数据的实例级别的属性信息, 并结合少量有标记数据构建一个弱领域知识库 $\mathcal{GKB}$. 弱领域知识库中将包含每个有标记样本的类别信息及其对应的关键属性的高维特征表示. 这些与类别相关的属性知识将为反绎学习的推理过程提供指导. 而后, 根据基于启发式规则扩张领域知识库的反绎学习方法, 优化目标任务中的表示学习模型. 方法的框架如图 2 所示.

具体而言, 任务的输入包含一组带有标记的训练数据 $\mathcal{D}_L = \{x_1, x_2, \ldots, x_m\}$, 一组不含标记的训练数据 $\mathcal{D}_U = \{x_{m+1}, x_{m+2}, \ldots, x_{m+n}\}$, 以及待构建的弱领域知识库 $\mathcal{GKB}$. 其中, 每个数据 $x_i \in \mathcal{D}_L$ 都是输入空间 $\mathcal{X}$ 中的一个点, x_i 所对应的已知的真实标记 $y_i \in \mathcal{Y}$; 每个数据 $x_j \in \mathcal{D}_U$ 同样为输入空间 $\mathcal{X}$ 中的一个点, x_j 所对应的未知的真实标记 $y_j \in \mathcal{Y}$, $\mathcal{Y}$ 是训练数据 $\mathcal{D}_L$ 与 $\mathcal{D}_U$ 的标记空间. 待构建的弱领域知识库 $\mathcal{GKB}$ 由带有标记的样本集合 $\mathcal{D}_L$ 生成. 将待优化的表示学习模型及对应的下游分类任务分类器的组合定义为 $M : \mathcal{X} \mapsto \mathcal{Y}$. 对于任意的无标记输入 $x' \in \mathcal{D}_U$, 根据模型 M 可以得到一个预测结果 $\hat{y} = M(x')$. 而后通过在弱领域知识库 $\mathcal{GKB}$ 的 “监督” 下增广反绎逻辑程序的方式, 对模型预测结果进行反绎推理, 完成对模型 M 的优化任务.

3.2 弱领域属性知识库自构建

在反绎学习中, 通常借助逻辑溯因帮助机器学习模型进行训练, 其中逻辑溯因由一阶逻辑定义的强领域知识使用反绎推理实现. 然而这样定义完备的由一阶逻辑表示构成的强领域知识在很多实际应用中可能无法完全获得. 因此, 本文采用基于启发式规则扩张领域知识库的反绎学习方法 [8], 从弱领域知识库中寻找与预测结果最为接近的具体事实作为推理结果. 然而, 在很多表示学习的任务中, 这样的弱领域知识库仍然很难根据已有的知识构建. 因此, 对于一批无标记的样本, 本文通过构造 k 个对象属性查询 $\mathcal{Q}$, 依据自监督学习的方式完成属性信息的提取过程, 并依此构建弱领域属性知识库 $\mathcal{GKB}$.

如下, 我们以一般的自监督学习方式详述该过程的具体步骤. 首先, 对于给定的输入数据 x_i, 通过两种不同的数据增强方式生成两个增强后的输入 $x_i^{(1)}$ 与 $x_i^{(2)}$, 将 $x_i^{(1)}$ 与 $x_i^{(2)}$ 通过编码器后得到各自的高维空间特征, 其中, 编码器的参数在整个模型的训练过程中被冻结. 接着, 随机初始化 k 个对象属性查询 $\mathcal{Q}$ 与高维空间特征解码器 [23]. 将 $x_i^{(1)}$ 与 $x_i^{(2)}$ 的高维空间特征作为解码器的值 (value), 分别添加位置编码后作为解码器的键 (key), $\mathcal{Q}$ 则直接作为解码器的查询 (query), 而后通过解码器得到最终的对象属性信息. 其中, $\mathcal{Q}$ 可以理解为对于一批数据, 预定义了 k 个属性特征查询, 经过解码器输出的 k 个特征向量则代表对于 $\mathcal{Q}$ 中的每一个查询, 数据 x_i 对其的实际表达强度. 当 $k \geqslant 2$ 时, 还需要对解码器输出的 k 个特征向量进行正交约束以避免模型学习到数据中相似的属性 [24]. 在实现过程中, 该过程可以用任意其他自监督学习框架替换.

通过无标记数据与有标记数据的集合完成对解码器及属性查询 $\mathcal{Q}$ 的预训练任务后, 仅将少量的有标记数据 $x_i \in \mathcal{D}_L$ 输入编码器与解码器, 得到每个有标记数据对应的 k 个属性特征向量, 并将其直接作为每个类别的属性知识. 弱监督知识库 $\mathcal{GKB}$ 即由这些属性特征向量与其对应的标记构成, 可以

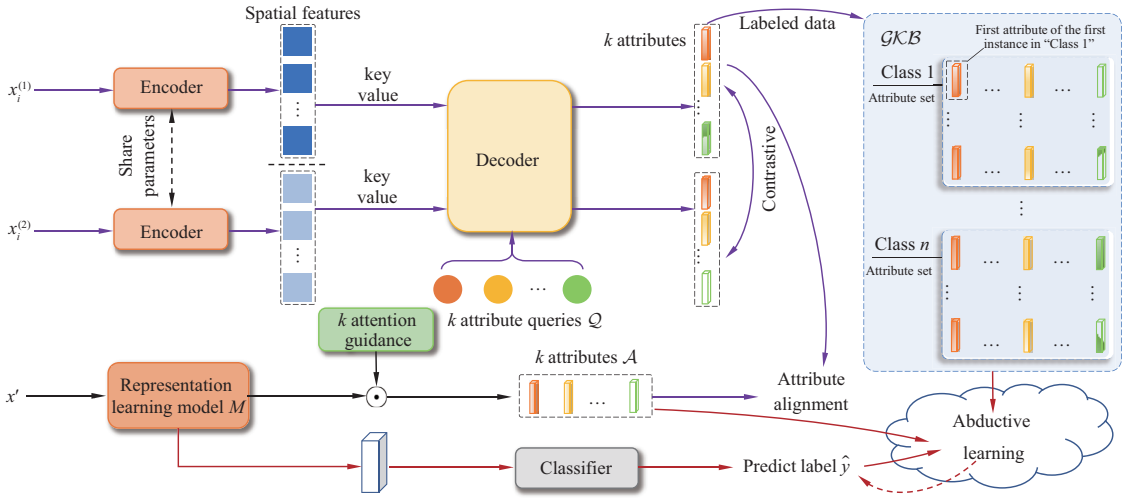

图 2 (网络版彩图) 本文方法详细流程图. 其中, 紫色箭头代表该部分仅在弱领域属性知识库自构建阶段训练, 红色箭头代表该部分仅在基于启发式规则扩张领域知识库的属性表示反绎学习阶段训练, 黑色箭头代表该部分在两个阶段均进行训练

Figure 2 (Color online) Details of the proposed method. The purple arrows indicate that these parts are trained only during the ground domain attribute-aware knowledge-based self-construction phase. The red arrows denote that these parts are trained during the ground abductive learning phase. The black arrows denote that these parts are trained in both phases

表示为

$$\mathcal{GKB} = \{(y_1, a_{11}, a_{12}, \ldots, a_{1k}), \ldots, (y_m, a_{m1}, a_{m2}, \ldots, a_{mk})\}, \tag{1}$$

其中, y_i 表示数据 x_i 的类别标记, a_{ij} 表示数据 x_i 的第 j 个属性特征. 对于表示学习模型 M, 在该阶段则需要训练 k 个注意力引导以获取对齐的属性信息进行反绎学习, 用于分类任务的分类器在该阶段不进行训练.

3.3 基于启发式规则扩张领域知识库的属性表示反绎学习 (AR-GABL)

在完成弱领域知识库 $\mathcal{GKB}$ 的构建后 (即完成弱领域属性知识库的构建), 我们需要在 $\mathcal{GKB}$ 的基础上进行增广反绎逻辑来完成对表示学习模型 M 经过分类器后得到的预测结果 $\hat{y}$ 进行反绎推理. 具体的反绎程序可以由下述的一阶逻辑表达进行描述:

$$\begin{aligned} \text{program}(\hat{y}, \bar{y}, \mathcal{GKB}) \leftarrow\ & \hat{y} \in \mathcal{GKB} \\ & \wedge \text{between}(0, F(k), \text{distance}(f(\hat{y}), f(\bar{y}))) \\ & \wedge \text{confidence}(\bar{y}, M) \geqslant \text{threshold}, \end{aligned} \tag{2}$$

其中, $F(k)$ 代表关于属性查询数量 k 的度量函数, $f(\hat{y})$ 代表对于预测结果为 $\hat{y}$ 的数据, 取其对应模型输出的 k 个属性, $f(\bar{y})$ 则代表在弱领域知识库中, 取类别为 $\bar{y}$ 的 k 个属性的中心向量, distance$(\cdot)$ 为距离度量函数, confidence$(\cdot)$ 为置信度度量函数, threshold 为置信度阈值.

具体而言, 在根据 $\mathcal{GKB}$ 对伪标记 $\hat{y}$ 进行反绎推理的过程中, 首先, 需要获取数据的属性标记并判断其是否属于弱领域知识库 $\mathcal{GKB}$ 可以进行反绎推理的范围: 对于任意输入的无标记数据 $x_i \in \mathcal{D}_u$, 通过表示学习模型 M 经过分类器后得到其预测结果 $\hat{y}$ 及样本 x_i 对应的 k 个属性标记 $\mathcal{A}_i$, 由于 $\mathcal{GKB}$ 中包含了带有标记数据的每个样本的属性信息, 因此通过遍历 $\mathcal{GKB}$ 中存储的每个属性是否均存在与

算法 1 弱领域属性知识库的自构建

输入: 有标记数据集 $\mathcal{D}_L$, 无标记数据集 $\mathcal{D}_U$, 模型 M, 编码模型 Encoder, 解码模型 Decoder, 训练轮数 E_1, k 个随机初始化属性查询 $\mathcal{Q}$, 损失函数 $\mathcal{L}_1$ 与 $\mathcal{L}_2$.

输出: 表示学习模型 M 与弱领域属性知识库 $\mathcal{GKB}$.

for $e = 1$ to E_1 **do**
  for $x \in \{\mathcal{D}_L, \mathcal{D}_U\}$ **do**
    $x_1 = \text{Augmentation}_1(x)$;
    $x_2 = \text{Augmentation}_2(x)$;
    $\text{Attribute}_1 = \text{Decoder}(\text{Encoder}(x_1), \mathcal{Q})$;
    $\text{Attribute}_2 = \text{Decoder}(\text{Encoder}(x_2), \mathcal{Q})$;
    $\mathcal{A} = M(x)$;
    $\text{Loss}_1 = \mathcal{L}_1(\text{Attribute}_1, \text{Attribute}_2)$;
    $\text{Loss}_2 = \mathcal{L}_2(\text{Attribute}_1, \mathcal{A})$;
    $\text{Update}((\text{Decoder}, \mathcal{Q}, M), \text{Grad}(\text{Loss}_1 + \text{Loss}_2))$;
  end for
end for
利用 $\mathcal{D}_L$, $\mathcal{Q}$, Encoder 与 Decoder 构建 $\mathcal{GKB}$;

返回: M, $\mathcal{GKB}$.

$\mathcal{A}_i$ 中对应属性的相似性大于阈值 τ 的序列, 以判断是否可以基于弱领域知识库 $\mathcal{GKB}$ 进行反绎推理, 其中相似性度量函数记为 $\varphi_1(\cdot)$; 而后, 对于 $\mathcal{GKB}$ 中存储的每个样本的属性序列, 计算其与 $\mathcal{A}_i$ 的距离, 取距离最小的向量所指代的类别作为反绎标记 $\bar{y}$, 并判断该最小距离是否在 0 与 $F(k)$ 间, 若距离大于 $F(k)$ 则直接认为反绎标记 $\bar{y}$ 无效; 最后, 计算模型 M 对于反绎标记 $\bar{y}$ 的预测置信度, 若置信度大于预定义的 threshold, 则认为反绎标记有效, 将二元组 $(x_i, \bar{y})$ 添加到待训练样本中 (即对 x_i 设置伪标记 $\bar{y}$), 否则认为反绎标记无效. 表示学习模型 M 的参数仅通过具有反绎标记的数据进行训练与更新.

3.4 模型训练

完整的模型训练一共分为两个阶段. 第 1 阶段通过自监督学习的方式得到稳定的数据解码器与表示学习模型 M, 并通过有限的有标记数据构建弱领域属性知识库 $\mathcal{GKB}$, 其完整构建方式如算法 1 所示. 其中, $\text{Augmentation}_1(\cdot)$ 与 $\text{Augmentation}_2(\cdot)$ 指代弱领域属性知识库自构建过程中对输入数据 x_i 作用的两种不同的数据增强方式, $\mathcal{L}_1$ 与 $\mathcal{L}_2$ 均为均方差损失.

根据算法 1 得到的表示学习模型 M 与弱领域属性知识库 $\mathcal{GKB}$, 第 2 阶段则基于启发式规则扩张领域知识库的反绎学习方法进一步训练模型 M, 其具体形式可以根据算法 2 表示. 其中 $M_{e-1}(\cdot)$ 代表第 $e-1$ 轮更新得到的模型, M_0 即为算法 1 输出的模型. abduce$(\cdot)$ 即为一次反绎推理过程, 其中反绎程序 P 遵循式 (2).

4 实验分析

本文选取半监督图像识别作为典型任务, 通过定量与定性实验来验证本文方法在半监督表示学习问题上的有效性.

4.1 实验数据与设定

我们使用两类数据集来验证所提出方法的有效性.

算法 2 基于启发式规则扩张领域知识库的属性表示反绎学习

输入: 无标记数据集 $\mathcal{D}_U$, 表示学习模型 M, 弱领域属性知识库 $\mathcal{GKB}$, 反绎程序 P, 训练轮数 E_2.
输出: 训练后的模型 M.

```
for e = 1 to E_2 do
    D̄ = [];
    for x ∈ D_U do
        ŷ = M_{e−1}(x);
        ȳ = abduce(ŷ, GKB, P);
        if ȳ ≠ None then
            D̄.append((x, ȳ));
        end if
    end for
    通过 D̄ 中的数据更新模型 M 的参数;
end for
```

返回: M.

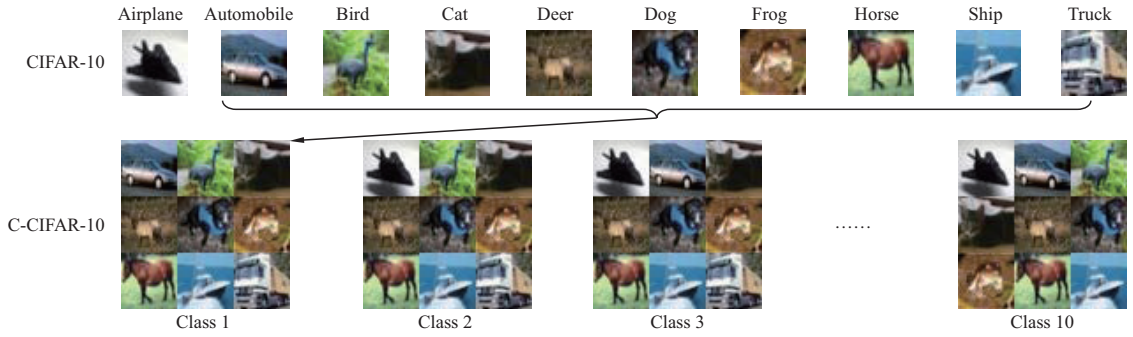

图 3 (网络版彩图) **C-CIFAR-10 合成数据集示例. C-CIFAR-10 的类别 1 由 CIFAR-10 数据集的后 9 种类别 (视作 9 种 "属性") 组合得到; 在实际合成 C-CIFAR-10 的过程中, 每个类别中包含的 9 种 "属性" 的位置也会随机生成**

Figure 3 (Color online) Examples of the C-CIFAR-10 dataset. Class 1 of C-CIFAR-10 is synthesized by combining the last 9 classes (regarded as 9 'attributes') of the CIFAR-10 dataset. During the synthesis of the C-CIFAR-10 dataset, the positions of the 9 'attributes' within each class are also randomly generated

第一类数据集通过人工合成得到. 具体而言, 我们将 CIFAR-10 [25] 中的每个类别 (分别记 airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck 编号为 1∼10) 视作一个 "属性", 而后根据这些 "属性" 重新构建数据集. 重构后的类别 1 中每张图像均包含编号 "2∼10" 中的图像, 类别 2 则包含编号 "1" 及 "3∼10" 的图像, 依此类推, 共 10 个类别 (如图 3 所示). 每个类别包含 5000 张图像用于训练, 1000 张图像用于测试, 并记该数据集为 C-CIFAR-10. 在本文研究的半监督任务下, 我们设置有标记图像数量占总图像数量的 1%, 2%, 5% 与 10%.

第二类数据集为真实世界中带有属性标记的数据集. 我们使用 CUB200-2011 [26] 与 Stanford Dogs [27] 中的图像数据. CUB200-2011 数据集一共涵盖了 200 个细粒度鸟类的子类别, 共包含 11788 张图像, 其中 5994 张图像用于训练, 5794 张图像用于测试. 在本文研究的半监督任务下, 设置有标记图像数量占总图像数量的 3% (每类 1 张有标记数据), 10% (每类 3 张有标记数据), 20% (每类 6 张有标记数据) 以进行实验. Stanford Dogs 数据集则涵盖了 120 个细粒度狗类的子类别, 共包含 20580 张图像, 其中每个类 100 张图片用于训练, 剩余的共 8580 张图像用于测试. 在本文研究的半监督任务下,

表 1 合成数据集下各方法的半监督分类性能对比

Table 1 Comparisons of semi-supervised classification performance on the synthetic dataset

Method	Labeled data ratio (%)	Top-1 Acc. (%)
Baseline	1 / 2 / 5 / 10	15.94 / 18.42 / 25.47 / 41.50
Pseudo-label	1 / 2 / 5 / 10	11.29 / 13.22 / 26.34 / 47.73
Self-Tuning	1 / 2 / 5 / 10	12.49 / 13.51 / 27.49 / 48.49
FlexMatch	1 / 2 / 5 / 10	12.72 / 13.76 / 28.02 / 49.72
DST	1 / 2 / 5 / 10	13.16 / 13.93 / 28.93 / 50.21
Pi-model	1 / 2 / 5 / 10	12.54 / 13.36 / 28.87 / 49.65
Mean teachers	1 / 2 / 5 / 10	12.96 / 14.24 / 29.46 / 50.93
Ours w/o $\mathcal{GKB}$	1 / 2 / 5 / 10	16.32 / 18.83 / 25.91 / 42.81
Ours	1 / 2 / 5 / 10	**23.18** / **32.81** / **48.30** / **60.94**

设置有标记图像数量的比例与 CUB200-2011 数据集相同.

4.2 评价指标及实验方法

本文仅考虑使用表示学习优化得到的模型完成分类任务. 我们采用 Top-1 准确率来进行评估. 基线方法使用 ImageNet-1K [28] 预训练参数初始化模型, 通过动量为 0.9 的随机梯度下降法对训练集中的有标记样本训练 80 轮. 权重衰减设置为 0.0005, 批次大小为 16. 训练期间将输入图像随机裁剪, 用双线性插值法缩放为 224 × 224, 再经过随机水平翻转并对亮度、对比度、饱和度和色调进行数据增强. 骨干网络为 ResNet-18 [29]. 为公平起见, 我们还将传统的基于伪标记的半监督学习算法 Pseudo-label [30], 目前基于伪标记的最优半监督学习方法 Self-Tuning [31], FlexMatch [32] 与 DST [33] 以及半监督领域经典的算法 Pi-model [34] 与 Mean teachers [35] 做相应设定修改后作为对比方法进行结果的定量比较. 其中, Pseudo-label [30] 通过训练中的模型对无标签数据进行预测, 并将概率最高的类别视作无标签数据的伪标签; Self-Tuning [31] 提出了一种通用的伪标签组对比机制, 减轻了模型对伪标签的依赖程度; FlexMatch [32] 采用课程学习的方式通过伪标签的状态来自适应调整半监督训练过程中不同置信度的阈值; DST [33] 设计了减少训练偏差与数据偏差的方法, 解决了半监督学习下训练不稳定及可能产生马太效应的问题; Pi-model [34] 基于 Π 模型, 对未标记的数据进行两次预测, 将两次预测结果间的均方误差作为无监督的损失, 并对训练样本的预测结果进行指数移动平均; Mean teachers [35] 则在 Pi-model 的基础上作出改进, 降低了大数据集标签更新周期长、无法适用于在线学习的问题.

本文方法在自监督学习阶段使用 CLIP [36] 模型作为编码器并冻结参数, 解码器则随机初始化参数, 属性表征维度设置为 256 维, 训练轮数为 300, 批次大小为 128, 数据增强方式与 BYOL [37] 相同. 表示学习模型 M 的框架为 ResNet-18 [29], k 个注意力引导由 1×1 卷积生成 [38]. 训练完成后由训练集中的有标记样本的属性构成弱领域属性知识库. 在反绎学习阶段, 模型通过算法 2 标注无标记训练样本并使用基线方法在有标记样本和伪标记样本上进行训练. 距离度量损失 $\mathcal{L}_1$ 与 $\mathcal{L}_2$ 均为余弦相似度损失函数, 超参数 threshold 为 0.5.

4.3 合成数据集下的实验结果与分析

首先在合成数据集上验证方法的有效性. 对于通过 CIFAR-10 重构得到的新数据集 C-CIFAR-10, 将原始数据集中的每个类别视作一个属性, 即 C-CIFAR-10 中每个类别共包含 9 个属性, 因此超参数 k 恒等于 9. 如表 1 所示, 首先通过有标记的数据训练了基线模型, 而后在提出方法的框架中进行训

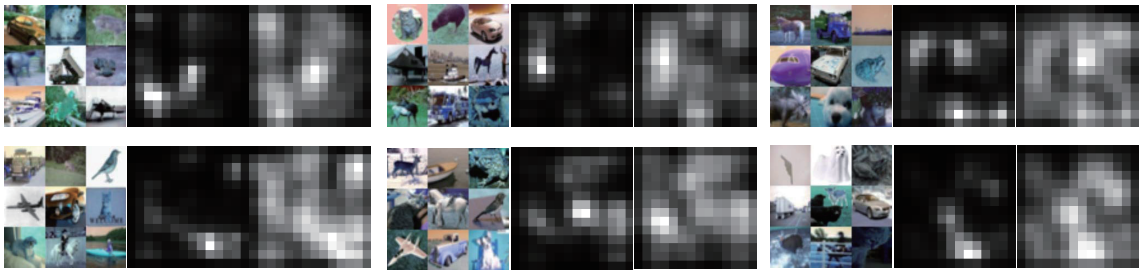

图 4 (网络版彩图) C-CIFAR-10 合成数据集上的激活图可视化. 每个三元组由 C-CIFAR-10 原始类别图像、没有使用 AR-GABL 时模型对该图像的激活和我们提出的方法优化后的表示学习模型对该图像的激活三部分组成

Figure 4 (Color online) Activation map on the C-CIFAR-10 dataset. Each triplet consists of the following: the original C-CIFAR-10 class image, the activation map without adopting $\mathcal{GKB}$, and the activation map generated by our method

练, "w/o $\mathcal{GKB}$" 代表没有使用 AR-GABL (参考 3.3 小节) 部分, 直接对模型进行微调. 不难看出, 通过本文方法构建的弱领域属性知识库对模型进行反绎推理后, 半监督分类性能得到了大幅提升. 另一方面, 基于伪标记的传统方法与最优方法为无标记数据集中的所有对象均添加了伪标记. 这些方法无法很好地关注对象的属性, 导致在带标记样本量极少且缺乏合成数据集分布下良好的预训练模型的情况下, 这些方法得到的伪标记质量较低且错误标记数量大, 甚至劣于基线方法的半监督分类精度. 而本文的方法针对对象属性构建了弱领域属性知识库, 可使其更好地刻画属性信息从而良好地融合知识与感知, 仅对该知识库可识别的无标记数据添加确信的伪标签, 因此可以获得更好的半监督分类性能.

而后在 C-CIFAR-10 数据集中进行了可视化以验证本文提出的方法对表示学习模型的改进动机. 如图 4 所示, 在没有使用 AR-GABL (参考 3.3 小节) 进行反绎推理时, 模型对 "属性" 的学习能力非常有限 (即只能激活部分属性). 而在添加该反绎推理的步骤进行模型更新后, 表示学习模型可以学习到绝大部分的属性信息.

4.4 真实数据集下的实验结果与分析

本小节在真实的自然鸟类数据集 CUB200-2011 以及狗类数据集 Stanford Dogs 中进行实验. 我们固定属性数量为 4, 即超参数 k 等于 4. 如表 2 所示. 对于 CUB200-2011 数据集, 当有标记数据仅有 3% 时 (即每个类别仅包含 1 张有标记数据), 对比基线方法, 本文方法获得了近 4% 的性能提升, 而现有的基于伪标记的最优方法则与基线方法的半监督分类性能几乎相同或仅有微小的提升; 对于 Stanford Dogs 数据集, 当有标记数据占 3% 时 (即每个类包含 3 张有标记数据), 对比基线方法, 本文方法仍然获得了约 3% 的性能提升, 同样地, 现有的基于伪标记的方法与一些常见的半监督学习方法相比基线方法的半监督分类性能仅有微小的变化. 说明以往的方法在样本量极少的情况下, 无法有效地为无标记数据添加正确的有利于模型更新的伪标记, 也证明了本文提出的方法在真实环境下也可以根据极其有限的有标记数据构建弱领域属性知识库并基于该知识库为无标记数据中的部分对象添加确信的伪标记.

当有标记数据提升至 10% 与 20% 时, 对于 CUB200-2011 数据集, 本文方法获得了平均约 7% 的性能提升, 而对于 Stanford Dogs 数据集, 本文方法获得了平均约 5% 的性能提升. 一方面证明了本文提出的方法在有标记数据数量提升的情形下, 可以稳定地根据自构建得到的弱领域属性知识库, 反绎推理得到更多的无标记数据的有效反绎标签. 另一方面, 弱领域属性知识库的完备程度也直接影响了方法的性能, 当弱领域属性知识库中包含的知识量极少时, 对于极大部分的无标记数据, 反绎推理过程会判断其不属于可识别范围; 而当弱领域属性知识库中包含的知识量增加时, 对于无标记数据的有

表 2 真实环境数据集下半监督分类性能对比

Table 2 Comparisons of semi-supervised classification performance on real-world datasets[a)]

Method	Labeled data ratio (%)	CUB200-2011 Top-1 Acc. (%)	Stanford Dogs Top-1 Acc. (%)
Baseline	3	12.15	36.55
Pseudo-label	3	12.53	36.84
Self-Tuning	3	12.67	37.02
FlexMatch	3	12.59	36.95
DST	3	12.44	36.77
Pi-model	3	12.72	36.48
Mean teachers	3	12.94	36.79
Ours w/o $\mathcal{GKB}$	3	12.24	36.70
Ours	3	**15.81**	**39.61**
Baseline	10	28.73	44.54
Pseudo-label	10	29.65	45.23
Self-Tuning	10	30.46	45.64
FlexMatch	10	30.97	45.89
DST	10	29.37	44.91
Pi-model	10	30.64	44.72
Mean teachers	10	31.26	45.35
Ours w/o $\mathcal{GKB}$	10	28.96	44.75
Ours	10	**37.14**	**50.40**
Baseline	20	43.49	54.53
Pseudo-label	20	44.94	55.68
Self-Tuning	20	45.27	56.16
FlexMatch	20	45.58	56.42
DST	20	46.08	56.87
Pi-model	20	45.67	55.74
Mean teachers	20	45.98	56.40
Ours w/o $\mathcal{GKB}$	20	43.75	55.09
Ours	20	**52.38**	**61.82**

a) The highest accuracy with the same labeled data ratio is marked in bold.

效应用率也会大幅提升.

而后在 CUB200-2011 与 Stanford Dogs 数据集中进行了可视化以进一步验证本文提出的方法对表示学习模型的改进动机. 如图 5 所示, 在没有使用 AR-GABL (参考 3.3 小节) 进行反绎推理时, 模型只能观测到部分属性信息. 而在添加该反绎推理的步骤进行模型更新后, 表示学习模型可以学习到更多的属性信息.

4.5 消融实验

本小节对超参数 k (属性查询 $\mathcal{Q}$ 的数量) 进行消融实验. 首先在合成数据集 C-CIFAR-10 上进行探索. 由于 C-CIFAR-10 中每个类均包含了 9 个人为定义的属性, 因此在实验过程中我们固定 $k = 9$. 本小节选取 k 为 3, 6, 9 与 12, 分别在有标记数据比例为 1%, 2%, 5% 与 10% 时进行实验以探究当

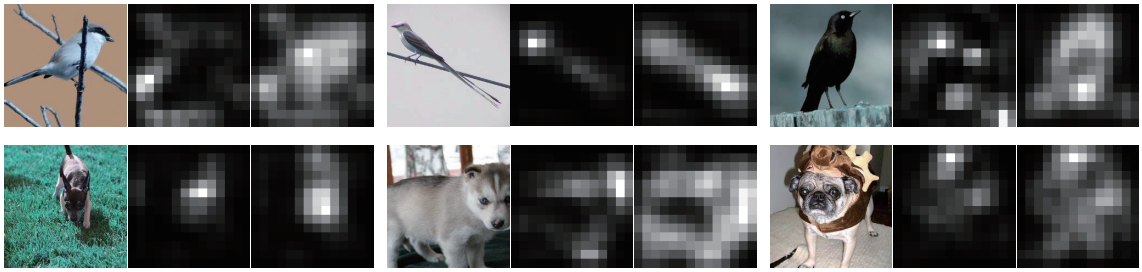

图 5 (网络版彩图) 真实数据集下激活图可视化. 上方三组图像为 CUB200-2011 数据集下的可视化, 下方三组图像为 Stanford Dogs. 每个三元组由原始类别图像、没有使用 AR-GABL 时模型对该图像的激活和我们提出的方法优化后的表示学习模型对该图像的激活三部分组成

Figure 5 (Color online) Activation map on the CUB200-2011 (top three sets of images) and Stanford Dogs (three sets of images below) datasets. Each triplet consists of the following: the original image, the activation map without adopting $\mathcal{GKB}$, and the activation map generated by our method

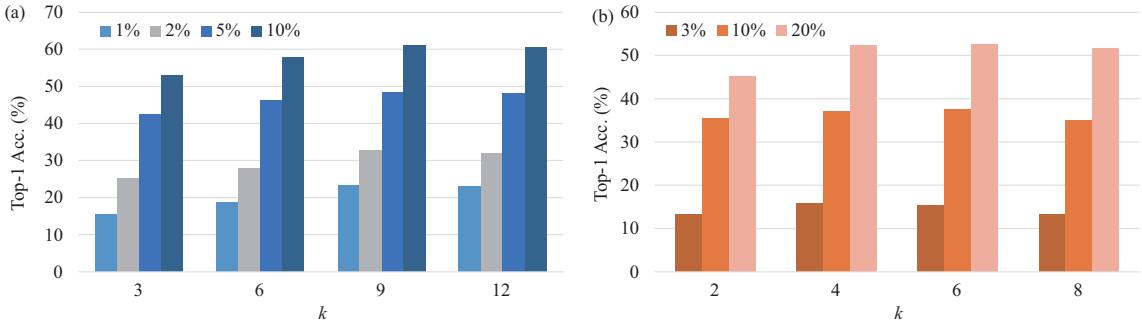

图 6 (网络版彩图) 超参数 k 的消融实验. (a) 在合成数据集 C-CIFAR-10 上的消融实验; (b) 在真实数据集 CUB200-2011 上的消融实验

Figure 6 (Color online) Ablation studies on the hyper-parameter k. (a) Ablation studies under the C-CIFAR-10 dataset; (b) ablation studies under the CUB200-2011 dataset

k 小于及大于人为设定属性量时的效果. 消融实验结果如图 6(a) 所示. 可以看到, 对于合成数据集 C-CIFAR-10, 当 k 小于 9 时, Top-1 准确率指标随着 k 增大稳步提升; 当 k 大于 9 时, 虽然随着 k 增大模型参数量也会增加, 但是模型的 Top-1 准确率性能反而略有下降. 图 4 的可视化结果也表明, 当 $k = 9$ 时, 模型很好地捕获了人工合成数据集约束的 9 个特征.

而后我们在真实数据集 CUB200-2011 中进行消融实验. 选取 k 为 2, 4, 6 与 8, 在有标记数据比例分别为 3%, 10%, 20% 时, 探究当 k 逐步变大时, 模型提取属性特征的能力. 消融实验结果如图 6(b) 所示. 可以看到, 对于真实数据集 CUB200-2011, 当 k 小于 4 时, Top-1 准确率指标随着 k 增大提升较为明显; 而当 k 大于 4 时, 虽然模型参数量会随之增加, 但模型的 Top-1 准确率性能提升极其微小, 且在 $k = 8$ 时, 模型性能开始略微下降. 同时, 结果也表现出在 k 大于 4 时, 模型对超参数 k 的选取并不敏感. 为了平衡模型的参数量进行公平对比, 在本文实验部分, 选取 $k = 4$ 进行真实环境数据集的实验.

5 总结与展望

本文提出了一种可以根据少量有标记样本构建弱领域属性知识库, 并通过无标记数据与反绎学习优化表示学习模型的方法. 方法通过自监督的方式自动提取所有有标记样本的属性信息, 以此构建弱

领域属性知识库, 解决了表示学习任务下标注完备的对象属性知识并构建知识库困难的问题. 方法还通过基于启发式规则扩张领域知识库的反绎学习方法, 依据最小不一致原则修改表示学习模型得到的错误的属性预测, 并对相应的无标记数据添加反绎标记, 给予表示学习模型更强的可解释性. 通过在人工合成数据集上对比实验及可视化结果, 验证本文提出的方法可以有效地提取有标记对象的不同属性信息并构建有助于更新表示学习模型的弱领域属性知识库. 在真实环境下, 数据集中的实验也进一步证明了提出方法的有效性. 未来, 我们希望将方法扩展至目标检测任务中, 提升细粒度半监督目标检测任务的性能.

参考文献

1 Zhou Z H. A brief introduction to weakly supervised learning. Natl Sci Rev, 2018, 5: 44–53
2 Alobaidi M H, Meguid M A, Zayed T. Semi-supervised learning framework for oil and gas pipeline failure detection. Sci Rep, 2022, 12: 13758
3 Yang X, Song Z, King I, et al. A survey on deep semi-supervised learning. IEEE Trans Knowl Data Eng, 2023, 35: 8934–8954
4 Cai L W. Research and application of abductive learning based on weak domain knowledge. Dissertation for Master's Degree. Nanjing: Nanjing University, 2021 [蔡乐文. 弱领域知识下的反绎学习方法研究与应用. 硕士学位论文. 南京: 南京大学, 2021]
5 Wei X S, Song Y Z, Aodha O M, et al. Fine-grained image analysis with deep learning: a survey. IEEE Trans Pattern Anal Mach Intell, 2022, 44: 8927–8948
6 Bengio Y. The consciousness prior. 2017. ArXiv:1709.08568
7 Zhou Z-H. Abductive learning: towards bridging machine learning and logical reasoning. Sci China Inf Sci, 2019, 62: 076101
8 Cai L W, Dai W Z, Huang Y X, et al. Abductive learning with ground knowledge base. In: Proceedings of the 30th International Joint Conference on Artificial Intelligence, Montreal, 2021. 1815–1821
9 Dai W Z, Xu Q, Yu Y, et al. Bridging machine learning and logical reasoning by abductive learning. In: Proceedings of the 33rd Conference on Neural Information Processing Systems, Vancouver, 2019. 2811–2822
10 Huang Y X, Dai W Z, Cai L W, et al. Fast abductive learning by similarity-based consistency optimization. In: Proceedings of the 35th Conference on Neural Information Processing Systems, 2021. 26574–26584
11 Huang Y X, Dai W Z, Jiang Y, et al. Enabling knowledge refinement upon new concepts in abductive learning. In: Proceedings of the 37th AAAI Conference on Artificial Intelligence, Washington DC, 2023. 7928–7935
12 Manhaeve R, Dumancic S, Kimmig A, et al. DeepProbLog: neural probabilistic logic programming. In: Proceedings of the 32nd Conference on Neural Information Processing Systems, Montreal, 2018. 3749–3759
13 Hu E L, Chen S C, Yin X S. Manifold contraction for semi-supervised classification. Sci China Inf Sci, 2010, 53: 1170–1187
14 Li Z C, Tang J H. Semi-supervised local feature selection for data classification. Sci China Inf Sci, 2021, 64: 192108
15 Wei X-S, Cui Q, Yang L, et al. RPC: a large-scale and fine-grained retail product checkout dataset. Sci China Inf Sci, 2022, 65: 197101
16 Wei X-S, Xu H-Y, Yang Z W, et al. Negatives make a positive: an embarrassingly simple approach to semi-supervised few-shot learning. IEEE Trans Pattern Anal Mach Intell, 2024, 46: 2091–2103
17 Wei X-S, Xu S-L, Chen H, et al. Prototype-based classifier learning for long-tailed visual recognition. Sci China Inf Sci, 2022, 65: 160105
18 Springenberg J T. Unsupervised and semi-supervised learning with categorical generative adversarial networks. 2015. ArXiv:1511.06390
19 Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks. Commun Acm, 2020, 63: 139–144
20 Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: Proceedings of the 14th International Conference on Neural Information Processing Systems, Vancouver, 2001. 585–591
21 Iscen A, Tolias G, Avrithis Y, et al. Label propagation for deep semi-supervised learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019. 5070–5079

22 Zhou Z H, Li M. Semi-supervised learning by disagreement. Knowl Inf Syst, 2010, 24: 415–439
23 Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, 2017. 6000–6010
24 Wei X S, Shen Y, Sun X, et al. Attribute-aware deep hashing with self-consistency for large-scale fine-grained image retrieval. IEEE Trans Pattern Anal Mach Intell, 2023, 45: 13904–13920
25 Alex K. Learning Multiple Layers of Features from Tiny Images. Technical Report. 2009
26 Wah C, Branson S, Welinder P, et al. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report. 2011
27 Khosla A, Jayadevaprakash N, Yao B, et al. Novel dataset for fine-grained image categorization: Stanford Dogs. In: Proceedings of CVPR Workshop on Fine-Grained Visual Categorization (FGVC), 2011. 1–2
28 Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Miami, 2009. 248–255
29 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 770–778
30 Lee D H. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. In: Proceedings of the 30th International Conference on Machine Learning, Atlanta, 2013. 896–901
31 Wang X, Gao J, Long M, et al. Self-Tuning for data-efficient deep learning. In: Proceedings of the 38th International Conference on Machine Learning, Vienna, 2021. 10738–10748
32 Zhang B, Wang Y, Hou W, et al. FlexMatch: boosting semi-supervised learning with curriculum pseudo labeling. In: Proceedings of the 35th Conference on Neural Information Processing Systems, 2021. 18408–18419
33 Chen B, Jiang J, Wang X, et al. Debiased self-training for semi-supervised learning. In: Proceedings of the 36th Conference on Neural Information Processing Systems, New Orleans, 2022. 32424–32437
34 Laine S, Aila T. Temporal ensembling for semi-supervised learning. 2016. ArXiv:1610.02242
35 Tarvainen A, Valpola H. Mean teachers are better role models: weight-averaged consistency targets improve semi-supervised deep learning results. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, 2017. 1195–1204
36 Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning, Vienna, 2021. 8748–8763
37 Grill J B, Strub F, Altché F, et al. Bootstrap your own latent-a new approach to self-supervised learning. In: Proceedings of the 34th International Conference on Neural Information Processing Systems, Vancouver, 2020. 21271–21284
38 Shen Y, Sun X, Wei X S, et al. SEMICON: a learning-to-hash solution for large-scale fine-grained image retrieval. In: Proceedings of European Conference on Computer Vision, Tel Aviv, 2022. 531–548

Attribute-aware knowledge based self-abductive for semi-supervised representation learning

Yang SHEN[1], Xuhao SUN[1], Heyang XU[1] & Xiushen WEI[2,3*]

1. *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing* 210094*, China*;
2. *School of Computer Science and Engineering, Southeast University, Nanjing* 210096*, China*;
3. *Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications* (*Southeast University*)*, Ministry of Education, Nanjing* 210096*, China*
* Corresponding author. E-mail: weixs.gm@gmail.com

Abstract Integrating logical reasoning with machine learning holds the potential to substantially enhance model robustness and interpretability. In recent years, prevailing approaches have often been initiated with specific knowledge bases, leveraging abductive learning paradigms or their derivatives to optimize machine learning models. However, even in the presence of such knowledge bases, they frequently prove to be incomplete or noisy when applied to representation learning tasks. Additionally, domain experts may encounter challenges in accurately characterizing the attributed properties of various objects in real-world contexts. Focusing on semi-supervised representation learning tasks, our proposed method constructs a weak domain attribute knowledge base using a limited number of labeled samples and conducts self-abductive learning through grounded abductive learning with unlabeled data. This approach effectively addresses the limitations posed by insufficient strong domain knowledge in representation learning tasks and the scarcity of high-quality labeled data in real-world environments. Experimental comparisons conducted on both synthetic and real-world datasets validate the effectiveness of our proposed method.

Keywords artificial intelligence, machine learning, abductive learning, semi-supervised learning, feature representation, fine-grained attributes