

基于相关性建模的图注意力多示例多标签学习

吴嘉欣, 张 健*

中国矿业大学计算机科学与技术学院/人工智能学院, 江苏 徐州

收稿日期: 2025年11月4日; 录用日期: 2025年12月5日; 发布日期: 2025年12月12日

摘 要

多示例多标签学习 (Multi-Instance Multi-Label Learning, MIML) 在图像分类、文本标注等任务中展现了独特优势, 但现有方法在建模样本间复杂依赖关系和标签间相关性时仍存在不足。本文提出一种结合图注意力机制与对比学习的相关性建模多示例多标签学习方法 (GAMIL-C)。首先, 将每个包中的实例构建为图结构, 利用图注意力网络 (Graph Attention Network, GAT) 建模实例间的结构关系, 并引入边特征以刻画实例间的相似性。其次, 在标签空间中通过相关性感知模块捕捉标签间的高阶依赖, 以提升标签预测的准确性。进一步地, 我们引入对比学习策略, 在实例表征和标签表征之间构建判别性约束, 从而增强模型的泛化能力。实验在多个公开数据集上验证了该方法的有效性, 与主流基线方法相比, 在预测精度和宏平均指标上均取得显著提升。结果表明, 该方法能够有效建模实例与标签间的双重相关性, 并通过对比学习进一步提升模型鲁棒性, 为复杂场景下的多示例多标签学习提供了一种新的解决思路。

关键词

多示例多标签学习, 图注意力网络, 相关性建模, 对比学习

Graph Attention-Based Multi-Instance Multi-Label Learning with Correlation Modeling

Jiaxin Wu, Jian Zhang*

School of Computer Science and Technology/School of Artificial Intelligence, China University of Mining and Technology, Xuzhou Jiangsu

Received: November 4, 2025; accepted: December 5, 2025; published: December 12, 2025

Abstract

Multi-Instance Multi-Label Learning (MIML) has shown great potential in tasks such as image

*通讯作者。

文章引用: 吴嘉欣, 张健. 基于相关性建模的图注意力多示例多标签学习[J]. 计算机科学与应用, 2025, 15(12): 209-221. DOI: 10.12677/csa.2025.1512336

classification and text annotation. However, existing approaches still struggle to effectively capture the complex dependencies among instances and the correlations among labels. In this paper, we propose a novel correlation modeling approach for MIML based on graph attention networks and contrastive learning (GAMIL-C). Specifically, we first construct a graph structure for the instances within each bag and employ a Graph Attention Network (GAT) to model the structural relationships among instances, where edge features are incorporated to characterize their similarities. Then, a correlation-aware module is introduced to capture high-order dependencies in the label space, thereby improving label prediction accuracy. Furthermore, a contrastive learning strategy is incorporated to establish discriminative constraints between instance and label representations, which enhances the generalization capability of the model. Extensive experiments conducted on several public datasets demonstrate that our method significantly outperforms strong baselines in terms of prediction accuracy and macro-averaged metrics. The results indicate that the proposed method effectively captures the dual correlations between instances and labels, while contrastive learning further improves model robustness, providing a new perspective for MIML in complex scenarios.

Keywords

Multi-Instance Multi-Label Learning, Graph Attention Network, Correlation Modeling, Contrastive Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在图像处理、文本检索、生物信息学等领域,实际任务往往具有多示例、多标签的特点。例如,一幅图像可以划分为多个区域,每个区域作为一个示例,同时图像可能对应多个语义标签;一篇文档由若干段落或句子组成,同时可能涉及多个主题;蛋白质序列由多个片段构成,且通常具有多种功能标签[1][2]。为了刻画此类复杂的映射关系,多示例多标签学习(Multi-Instance Multi-Label Learning, MIML)被提出并逐渐成为研究热点[3]。其主要难点在于如何同时有效地建模示例间的依赖关系以及标签之间的相关性。

早期的多示例多标签方法主要基于实例聚合、特征转换或标签空间建模。例如, Zhou 等人率先提出了 MIML 的基本框架,为后续研究奠定了基础[4];此后,研究者们通过集成学习、核方法和概率建模等手段,不断提升模型的表达能力和泛化性能[5]。然而,这类方法通常假设示例之间相互独立,忽略了它们潜在的结构关系,因而难以适应更复杂的应用场景。

近年来,图神经网络(Graph Neural Networks, GNNs)在处理非欧几里得结构数据方面展现出强大优势,成为建模示例关系的重要工具[5][6]。其中,图注意力网络(Graph Attention Networks, GATs)通过引入注意力机制,自适应地学习节点间的权重关系,能够更有效地捕捉示例间的关键信息[5]。部分研究已尝试将图神经网络应用于多示例多标签任务,并取得了一定成效[6]。但现有方法仍然存在不足:(1) 示例间的语义相关性差异尚未得到充分建模,导致图结构难以真实反映任务所需关系;(2) 标签预测往往仅依赖独立的分类器,忽视了标签之间的依赖性;(3) 在存在冗余示例或噪声样本的情况下,模型鲁棒性不足,性能下降明显[7]。

针对上述问题,本文提出了一种相关性建模的图注意力多示例多标签学习方法。该方法在特征嵌入阶段通过计算边属性显式建模示例间相关性,使得图表示学习能够更好地捕捉实例间的全局依赖关系;在实例聚合阶段引入多头注意力池化机制,实现对关键示例的自适应选择与加权;在标签建模阶段,利

用标签相关性矩阵增强标签间依赖的刻画能力, 从而提升预测结果的整体一致性。同时, 本文进一步结合对比学习策略, 在表示学习过程中引入全局约束, 提升了模型的鲁棒性与判别性[8] [9]。

2. 相关工作

2.1. 多示例多标签学习

在实际应用中, 多示例多标签学习(MIML)有着广泛的应用场景。例如, 在图像识别任务中, 一幅完整的图像可视为一个“包”(bag), 其中的像素点或局部区域被视作多个“实例”(instances), 而该图像整体往往对应一个或多个语义标签。如何从实例中提取有效特征, 并进一步预测图像的整体标签, 是典型的 MIML 问题。MIML 的核心目标在于刻画由多个实例组成的对象与其多标签集合之间的对应关系。

形式化地, 可以将实例空间记为 X , 标签空间记为 Y 。一个包含 N 个样本的数据集记为: $D = \{(X_1, Y_1), (X_2, Y_2), \dots, (X_N, Y_N)\}$, 其中, 每个样本 $X_i \subseteq X$, 包含 n_i 个实例, 即: $X_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}$, $x_{ij} \in X$ 。与之对应的标签集合为: $Y_i = \{y_{i1}, y_{i2}, \dots, y_{iL}\}$, $Y_i \subseteq Y, y_{ij} \in Y$, 其中 L 表示整个数据集的标签总数。若样本 X_i 与标签 j 相关, 则定义 $y_{ij} = 1$, 否则 $y_{ij} = 0$ 。MIML 的目标是学习一个映射函数 $f: 2^X \rightarrow 2^Y$, 能够同时在包级别和实例级别上为测试样本进行标签预测。

现有的 MIML 方法大体上可以分为两类: 退化方法(degradation methods)和直接方法(direct methods)。前者通过将 MIML 转化为多示例学习(Multi-Instance Learning, MIL)或多标签学习(Multi-Label Learning, MLL)问题来间接求解。例如, MIMLBoost [10]将 MIML 转换为 MIL 问题; MIMLSVM [10]通过 k-medoids 聚类将 MIML 转换为 MLL 问题; 在此基础上 MIMLSVM+ [11]则进一步将 MLL 分解为一系列独立的二分类任务, 而 E-MIMLSVM+ [11]又结合多任务学习以解决标签相关性的问题; 而 MIMLNN 则以两层神经网络结构替代了 MIMLSVM 中的 MISVM 框架。

另一类方法则是直接基于 bag 或实例设计算法, 避免中间转化。例如, M3MIML [12]与 D-MIMLSVM [13]使用正则化技术解决问题; MIMLRBF [14]采用 RBF 神经网络进行实例级学习; MIML-Knn [15]基于 k 最近邻进行建模; KISAR [16]识别每个 bag 中与类别最相关的关键实例; MIMLFast [17]结合排序损失与复杂子概念进行优化; DeepMIML [18]则利用深度神经网络生成实例表示; MIMLLLMC [19]考虑了局部标签相关性; 而 MetaMIML [20]结合异构网络嵌入与元学习, 进一步提升了模型在跨任务场景下的泛化性能。

尽管现有方法在多示例多标签学习问题上取得了丰富的研究进展, 但仍存在一些局限。例如, 许多降解方法依赖于标签消歧、独立同分布假设或复杂的先验知识, 这在实际应用中往往难以满足, 且可能导致性能下降。同时, 直接建模的方法虽然能够一定程度上缓解上述问题, 但仍难以充分捕捉实例间的语义关联与标签之间的相关性, 从而限制了模型的表达能力。相比之下, 本文提出的相关性建模图注意力多示例多标签学习方法, 通过引入图神经网络建模实例间关系, 并结合标签相关性建模与对比学习机制, 有效提升了 bag 级与实例级的预测性能, 避免了对昂贵先验知识或标签消歧步骤的依赖。

2.2. 图注意力网络

图注意力网络(Graph Attention Network, GAT)最早由 Veličković 等人提出[2]。该方法将注意力机制引入图神经网络(Graph Neural Networks, GNNs), 突破了传统图卷积网络依赖拉普拉斯矩阵在频谱域进行复杂计算的限制, 而是直接在空间域通过邻居节点的特征聚合来更新节点表示。

具体来说, GAT 的核心思想是通过可学习的注意力权重来建模不同节点之间的重要性, 从而在信息聚合过程中对关键邻居赋予更高的权重。其过程主要包括两个步骤: 首先, 利用注意力机制计算节点之

间的相关性系数, 即

$$e_{ij} = a\left(\left[Wh_i \parallel Wh_j\right]\right), j \in N_i \quad (1)$$

其中, W 表示可学习的权重矩阵, h_i 和 h_j 分别是节点 v_i 和 v_j 的特征表示, $\parallel$ 表示拼接操作, $a(\cdot)$ 为计算相似度的函数, 通常配合 LeakyReLU 激活函数使用。然后, 对得到的相关性系数进行 softmax 归一化:

$$\alpha_{ij} = \frac{\exp(\sigma(e_{ij}))}{\sum_{k \in N_i} \exp(\sigma(e_{ik}))} \quad (2)$$

其次, 利用注意力系数对邻居节点信息进行加权求和, 得到更新后的节点特征表示:

$$h'_i = \sigma\left(\sum_{j \in N_i} \alpha_{ij} Wh_j\right) \quad (3)$$

$$h'_i(K) = \prod_{k=1}^K \sigma\left(\sum_{j \in N_i} \alpha_{ij}^k W^k h_j\right) \quad (4)$$

此外, GAT 结合了多头注意力机制(multi-head attention), 即通过多个独立的注意力头进行特征聚合, 并将结果拼接或求平均, 从而增强模型的表达能力与稳定性。由于其能够自适应地学习邻居节点之间的依赖关系, GAT 在节点分类、图表示学习等任务中取得了显著效果[2]。

2.3. 对比学习与相关性建模

近年来, 对比学习(Contrastive Learning)作为一种自监督学习范式, 在图表示学习、视觉识别和自然语言处理等领域得到了广泛应用[21]。其核心思想是通过构造正负样本对, 使模型在表示空间中拉近相似样本的距离, 拉远不相似样本的距离, 从而学习到判别性更强的特征表示。在图学习场景中, 图对比学习方法(Graph Contrastive Learning)通过子图采样、数据增强或特征扰动等方式构造对比样本, 有效提升了图神经网络的泛化能力与鲁棒性[22]。

另一方面, 标签相关性建模(Label Correlation Modeling)在多标签学习中具有重要意义。传统方法往往假设标签之间相互独立, 然而在实际任务中, 标签之间往往存在复杂的共现关系。例如, 在图像标注任务中, “海滩”和“海浪”常常同时出现; 在文本分类任务中, “人工智能”和“深度学习”也高度相关。如果忽视这种相关性, 容易导致预测偏差。已有研究通过标签嵌入、图建模或矩阵分解等方式来显式建模标签关系, 从而提升多标签预测的准确性[23]。

将对比学习与标签相关性建模结合, 可以在表示学习和预测层面同时获益。一方面, 对比学习增强了模型的特征表达能力, 帮助实例级与 bag 级的表示更加判别化; 另一方面, 标签相关性建模通过引入相关性矩阵或图结构, 捕捉不同标签之间的依赖关系, 从而优化预测结果。近年来已有部分工作尝试融合这两类方法, 例如在图神经网络的多标签学习中结合对比学习与标签相关性建模, 取得了优于传统方法的效果[24]。这为本文方法的设计提供了重要启发。

3. 方法

在本节中, 我们提出了一种相关性建模的图注意力多示例多标签学习方法(GATMIML)。给定一组多示例 bag, 模型的目标是在仅有 bag-level 标签监督的条件下, 通过联合建模实例间的结构依赖与标签间的相关性, 实现 bag-level 和 instance-level 的多标签预测。与传统方法不同, GATMIML 在框架中同时引入了图注意力网络和标签相关性矩阵, 并结合对比学习机制, 在提升表示学习能力的同时, 有效缓解标签独立性假设带来的限制。

3.1. 算法框架

为便于理解, 本文首先对所提出的 GATMIML 算法进行整体介绍。该算法主要由四个核心模块组成: 特征嵌入模块、图表示学习模块、标签相关性建模模块以及分类与对比学习模块。(1) 在特征嵌入阶段, 输入数据以 bag 形式给出, 每个 bag 含有多个实例。我们利用 K 近邻方法构建实例之间的图结构, 并通过边权函数获得边属性, 从而将实例特征转化为可用于图学习的结构化表示。(2) 在图表示学习阶段, 采用图注意力网络(GAT)对节点进行编码, 通过多头注意力机制聚合邻居信息, 有效建模实例间的依赖关系, 最终得到实例级表示并映射为 logits。(3) 在标签相关性建模阶段, 引入可学习的标签相关性矩阵, 对原始 logits 进行修正, 显式建模标签间的语义依赖关系, 从而提升预测的一致性与准确性。(4) 在分类与对比学习阶段, 利用 sigmoid 函数获得多标签预测结果, 同时通过联合优化分类损失和对比损失来训练模型, 其中分类损失用于监督 bag-level 标签预测, 对比损失则增强实例表示的判别性。通过上述模块的协同作用, GATMIML 能够同时捕捉实例间的结构关系和标签间的语义依赖, 实现更为准确的多示例多标签学习。

3.2. 特征嵌入

在多示例多标签学习任务中, 实例之间往往存在潜在的相关性, 若仅依赖独立的特征表示, 容易忽略 bag 内部的结构信息。为此, 本文在特征嵌入阶段首先将 bag 内实例构建为图结构输入, 以显式建模实例间的关系。设某个 bag 的实例集合为:

$$X_i = \{x_{i1}, x_{i2}, \dots, x_{in_i}\}, x_{ij} \in R^d, \quad (5)$$

其中, d 为特征维度, n_i 表示第 i 个 bag 的实例数量。为了捕捉实例之间的局部关系, 我们采用 K 近邻(KNN)方法在特征空间中为每个实例寻找其最近的 K 个邻居, 从而构建初始邻接矩阵 $A \in R^{n_i \times n_i}$ 。

为了进一步增强图的表达能力, 本文在边上引入边属性(edge attributes), 并通过实例间的余弦相似度进行计算。具体而言, 若节点 p 与节点 q 相连, 则其边权定义为:

$$e_{pq} = \frac{x_p \cdot x_q}{\|x_p\| \|x_q\|}, \quad (6)$$

其中, x_p 和 x_q 分别表示 p 节点与节点 q 的特征向量。最终, 每个 bag 被转化为一个带有边属性的加权图:

$$G = (V, E, X, E_{attr}), \quad (7)$$

其中, V 表示节点集合, E 表示边集合, X 为实例特征矩阵, E_{attr} 为边属性矩阵。通过这一特征嵌入过程, 原始实例集合被转化为结构化的图表示, 既保留了实例的语义信息, 又通过边属性建模捕捉了实例间的潜在相关性, 为后续的图表示学习奠定了基础。

3.3. 图表示学习

在完成特征嵌入后, 本文利用图神经网络进一步建模实例间的高阶依赖关系。我们采用改进的图注意力网络(Graph Attention Network, GATv2)作为基本单元, 通过自适应地分配邻居节点的重要性权重, 实现对节点特征的动态聚合。

设节点 i 的输入特征为 $h_i \in R^d$, 首先通过线性变换映射到新的特征空间:

$$z_i = W h_i, W \in R^{d' \times d}, \quad (8)$$

其中, 其中 d' 为新的特征维度。对于节点 i 与其邻居节点 $j \in N(i)$, 注意力机制计算其相关性系数:

$$e_{ij} = \text{LeakyReLU}\left(a^T \left[z_i \parallel z_j \right] \right), \quad (9)$$

其中, $\parallel$ 表示表示向量拼接, $a \in R^{2d'}$ 为可学习参数向量。接着, 通过 softmax 归一化获得注意力权重:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in N(i)} \exp(e_{ik})}. \quad (10)$$

在得到注意力权重后, 节点 i 的更新特征表示为:

$$h'_i = \sigma \left(\sum_{j \in N(i)} \alpha_{ij} z_j \right), \quad (11)$$

其中 $\sigma(\cdot)$ 为非线性激活函数。为了增强模型的稳定性和表达能力, 本文采用多头注意力机制, 将 K 个注意力头的结果拼接:

$$h' = \parallel_{k=1}^K \sigma \left(\sum_{j \in N(i)} \alpha_{ij}^{(k)} W^{(k)} h_j \right). \quad (12)$$

经过两层 GATv2 迭代更新后, 所有实例节点的特征被进一步编码为高阶语义表示。随后, 我们引入多头注意力池化, 从节点层级聚合到 bag 层级表示:

$$z = \sum_{i=1}^n \beta_i h'_i, \quad (13)$$

其中 β_i 为注意力池化分配的权重, 反映了实例对整体 bag 表示的贡献。通过该过程, 模型不仅能够捕捉实例间的复杂依赖关系, 还能动态选择关键实例, 从而得到判别性更强的 bag-level 表示。

3.4. 标签相关性建模

在多标签学习任务中, 不同标签之间往往存在潜在的语义依赖关系, 例如某些标签之间经常共现, 而另一些则具有明显的互斥性。如果忽略这些相关性, 可能导致预测结果不一致或置信度失衡。为此, 本文在 bag 表示经过分类得到的 logits 基础上, 进一步引入标签相关性建模模块, 以显式刻画标签间的依赖关系并对预测结果进行校正。

具体而言, 首先设一个可学习的参数矩阵 $M \in R^{L \times L}$ 其中 L 为标签总数。将其对称化并通过 $\tanh(\cdot)$ 函数映射到 $[-1, 1]$ 区间, 以得到初始的标签相关性矩阵:

$$S = \frac{1}{2} (M + M^T), \hat{S} = \tanh(S). \quad (14)$$

为避免自相关放大, 去除对角元素, 并对每一行进行归一化裁剪, 从而约束相关性矩阵的行稀疏性, 最终得到标签相关性矩阵 C :

$$C = I + C_{off}, \quad (15)$$

其中 I 为单位矩阵, C_{off} 表示行稀疏化后的非对角部分。这样可以保证标签在自表示时保持权重 1, 同时邻接行和受限于超参数, 避免数值不稳定。

在得到标签相关性矩阵 C 后, 将其用于校正原始 logits。具体做法是将 logits $z \in R^{1 \times L}$ 与其经过相关性传播后的结果 zC 进行加权融合:

$$\tilde{z} = (1 - \alpha)z + \alpha zC, \alpha \in [0, 1], \quad (16)$$

其中超参数 α 控制原始预测与相关性校正预测之间的平衡。通过这种方式, 模型能够在保持单标签判别

能力的同时, 显式利用标签之间的依赖关系。

此外, 为提升所学相关性矩阵的可解释性与稳定性, 对其非对角部分引入稀疏化正则化约束:

$$L_{corr} = \lambda_1 \|C_{off}\|_1 + \lambda_2 \|C_{off}\|_F^2 + \lambda_3 \cdot \frac{1}{L} \sum_{i=1}^L \frac{\sum_{j \neq i} |C_{off}(i, j)|}{L-1}, \quad (17)$$

其中三项分别控制整体稀疏性、数值稳定性和平均行稀疏度。最终, 该模块输出的相关性校正 logits $\tilde{z}$ 被送入 Sigmoid 函数以生成最终的多标签预测概率, 并与分类损失和对比学习损失共同优化整体模型。

3.5. 分类与对比学习模块

在得到 bag 级别的图表示后, 模型通过多层感知机 MLP 进行非线性映射, 获得初步的预测向量 $z \in R^L$, 其中 L 表示标签总数:

$$z = \text{MLP}(h), \quad (18)$$

随后, 引入标签相关性矩阵 $C \in R^{L \times L}$, 以建模不同标签之间的依赖关系。预测结果在与相关性矩阵融合后得到修正表示:

$$\tilde{z} = (1 - \lambda)z + \lambda zC, \quad (19)$$

其中, $\lambda \in [0, 1]$ 为平衡参数, 用于控制原始 logits 与相关性增强表示之间的权重。最终的多标签预测结果通过 Sigmoid 函数计算:

$$\hat{y} = \sigma(\tilde{z}). \quad (20)$$

为了进一步提升实例表示的判别性, 模型在训练阶段引入对比学习目标。设经投影头映射后的表示为 z' , 则对比学习损失基于 InfoNCE 定义为:

$$L_{con} = -\log \frac{\exp(\text{sim}(z'_i, z'_j)/\tau)}{\sum_{k=1}^N \exp(\text{sim}(z'_i, z'_k)/\tau)}, \quad (21)$$

其中, $\text{sim}(\cdot, \cdot)$ 表示余弦相似度, τ 为温度系数。该损失函数鼓励同一 bag 下的增强视图在表示空间中接近, 同时与其他样本保持区分。

最终的优化目标由分类损失 L_{cls} 与对比学习损失共同组成:

$$L = L_{cls} + \beta L_{con}, \quad (22)$$

其中, β 为对比学习损失的权重系数。通过联合优化两类目标, 模型能够在提升分类准确性的同时, 获得更加稳健和判别性的特征表示。

4. 实验

在本节中, 我们通过与多种先进的基线算法进行比较, 系统评估所提出的相关性建模的图注意力多示例多标签学习方法的有效性, 重点考察其在 bag 级别和实例级别多标签分类任务中的表现。

4.1. 数据集与实验设置

本文在六个公开的多示例多标签(MIML)基准数据集上进行了实验, 包括 Letter Carroll、Letter Frost、MSRC v2、Birds、Scene 和 Reuters 数据集[12] [19] [25]。其中, 前 3 个数据集同时提供 bag 级与实例级标签, 可用于 bag 级与实例级多标签分类任务; 后 3 个数据集仅包含 bag 级标签, 因此仅用于 bag 级多

标签分类任务。各数据集的详细信息见表 1, 其中 “#Bags” 表示 bag 的数量, “#Instances” 表示实例数量, “#Labels” 表示标签总数, “#Features” 表示原始特征维度, “#Average Bag Labels” 表示每个 bag 的平均标签数, “#Average Instance Labels” 表示每个实例的平均标签数。

实验在一台配备双 RTX 4070 GPU (CUDA 12.4) 的 Linux 服务器上完成。软件环境为 64 位 Windows 10 操作系统, Python 3.8 编程环境, 以及 PyTorch 1.13 深度学习框架。所有基线方法的超参数均按照各自论文公开的设置进行配置。本文实验结果均取 5 次三折交叉验证的平均值。

在超参数选择上, batch size 取自 {64, 128, 256}, 学习率取自 { $1e-2$, $5e-3$, $1e-3$, $5e-4$ }, 权重衰减取自 { $1e-3$, $5e-4$, $1e-4$ }. 图注意力层数从 {2, 3, 4} 中选择, 注意力头数取自 {2, 4, 8}, 隐藏层维度取自 {128, 256, 512}. 对比学习中温度系数 τ 从 {0.1, 0.2, 0.5, 1.0} 中选择, 平衡超参数 α 、 β 则从 {0.1, 0.5, 1.0, 2.0} 中选择。

Table 1. Statistics of multi-instance multi-label datasets

表 1. 多示例多标签数据集的统计信息

Datasets	#Bags	#Instances	#Labels	#Features	#Average Bag Labels	#Average Instance Labels
Birds	548	10,232	13	38	2.0	18.7
Letter Carroll	166	717	26	16	3.9	4.3
Letter Frost	144	565	26	16	3.6	3.9
MSRC v2	591	1758	23	48	2.5	3.0
Scene	2000	18,000	5	15	1.2	9.0
Reuters	2000	7119	7	243	1.2	3.6

4.2. 评价指标与对比算法

本文采用六个广泛使用的多标签评价指标来评估所提方法在 bag 级与实例级多标签分类任务中的性能[26] [27]。这些指标包括: 汉明损失(Hamming Loss, HL), 用于衡量预测标签与真实标签之间的不一致性; 覆盖度(Coverage, COV), 统计找到所有真实标签所需的排序步数; 一阶错误(One Error, OE), 计算预测排序中得分最高的标签不在真实标签集合中的比例; 排名损失(Ranking Loss, RL), 衡量模型预测标签排序与真实排序之间的不一致性; 平均精度(Average Precision, AP), 衡量预测排序中相关标签的整体位置情况; 以及宏平均 F1 值(Macro-averaging F1, Macro-F1), 计算所有标签的平均 F1 得分。前四个指标越小越好, 后两个指标则越大越好。

为了验证本文方法的有效性, 我们选取了六个具有代表性的 MIML 基线算法进行对比实验: (1) MIM-LSVM [10]: 通过 k-medoids 聚类将 MIML 问题转化为单实例多标签学习问题; (2) MIMLRBF [14]: 基于径向神经网络学习实例表示; (3) KISAR [16]: 识别 bag 中与各类标签相关的关键实例; (4) MIMLFast [17]: 利用排序损失与复杂标签子概念的快速 MIML 算法; (5) MIML-LLMC [19]: 在学习过程中考虑局部标签相关性; (6) MetaMIML [20]: 结合异构网络嵌入与元学习框架的最新方法。需要指出的是, KISAR、MIMLFast 和 MetaMIML 同时适用于 bag 级与实例级多标签分类任务, 从而能够更全面地与本文方法进行对比。

4.3. 实验结果与分析

在本节中, 我们通过大量实验全面评估所提出方法的有效性, 实验重点考 bag 级别的多标签分类性能。通过与多种主流基线方法的对比, 我们验证了本文方法在不同任务场景和评价指标下的优势。

Bag 级别标签预测性能对比

在本小节中, 我们系统评估了本文方法在 bag 级别多标签分类任务中的性能表现, 结果如表 2 所示。对比方法包括六种主流基线: MIMLSVM、MIMLRBF、KISAR、MIMLFast、MIML-LLMC 以及 MetaMIML。其中, MetaMIML 在 Scene 和 Reuters 数据集上的结果直接引用自原始论文, 其余方法的结果均在相同实验设置下, 通过五次三折交叉验证计算平均值和标准差得到。

从表中可以观察到, 本文方法在大多数数据集和指标上均取得了显著优势。具体而言, 在 Birds 与 Reuters 数据集上, 本文方法在六项指标均排名第一, 表现出稳定且全面的性能提升; 在 MSRCv2 数据集上, 本文方法在 5 项指标上取得最优, 在 Ranking Loss 上位列第二; 在 Letter Carroll 与 Letter Frost 数据集上, 本文方法在 Average Precision、Macro-F1 和 Coverage 等核心指标上优于对比方法, 尽管在 Hamming Loss 指标上表现略逊; 在 Scene 数据集上, 本文方法在 Ranking Loss、Macro-F1 和 Coverage 上取得最佳结果, 在 Average Precision 和 One Error 上则略低于 MetaMIML。

整体来看, 在六个数据集、六项评价指标(共 36 组实验)下, 本文方法在 72.2%的情况下取得第一名, 并在 91.7%的情况下保持在前二, 仅在 Letter 系列和 Scene 数据集的 Hamming Loss 指标上未能进入前二。上述结果表明, 本文方法在不同任务场景和评价指标下均表现出显著的稳定性和优越性。其优势主要得益于模型在图结构建模、标签相关性约束与对比学习机制等方面的综合设计, 有效缓解了多实例 bag 内部的标签模糊性问题, 从而提升了整体分类性能与泛化能力。

Table 2. Bag-level multi-label classification results
表 2. 包级多标签分类结果

Dataset	MIMLSVM	MIMLRBF	KISAR	MIMLFast	MIML-LLMC	MetaMIML	Ours
Average Precision ↑							
Birds	0.5681 ± 0.0037	0.8545 ± 0.0002	0.8364 ± 0.0004	0.9016 ± 0.0015	0.4191 ± 0.0217	<u>0.9248 ± 0.0271</u>	0.9315 ± 0.0059
Letter Carroll	0.5032 ± 0.0007	0.6737 ± 0.0012	0.5318 ± 0.0023	0.7082 ± 0.0020	0.6512 ± 0.0127	0.6774 ± 0.0298	<u>0.6842 ± 0.0287</u>
Letter Frost	0.5317 ± 0.0008	0.6663 ± 0.0007	0.5358 ± 0.0007	0.6997 ± 0.0027	0.6380 ± 0.0224	0.6870 ± 0.0819	<u>0.6937 ± 0.0436</u>
MSRC v2	0.7128 ± 0.0005	0.7327 ± 0.0002	0.6449 ± 0.0004	0.6902 ± 0.0039	0.7579 ± 0.0041	<u>0.7946 ± 0.0132</u>	0.7969 ± 0.0175
Scene	0.6534 ± 0.0002	0.7891 ± 0.0009	0.7944 ± N/A	0.7982 ± 0.0013	0.7972 ± 0.0143	0.9160 ± N/A	<u>0.8086 ± 0.0089</u>
Reuters	0.9519 ± N/A	0.9177 ± 0.0002	0.9515 ± N/A	0.9417 ± 0.0003	<u>0.9670 ± 0.0016</u>	0.9120 ± N/A	0.9717 ± 0.0061
One Error ↓							
Birds	0.5518 ± 0.0120	0.1387 ± 0.0003	<u>0.1230 ± 0.0009</u>	0.2486 ± 0.0218	0.6395 ± 0.0312	0.1436 ± 0.0541	0.0594 ± 0.0163
Letter Carroll	0.4457 ± 0.0051	0.2432 ± 0.0047	0.2386 ± 0.0038	0.2035 ± 0.0056	0.2228 ± 0.0268	0.5050 ± 0.0478	<u>0.2181 ± 0.0347</u>
Letter Frost	0.3375 ± 0.0063	0.2097 ± 0.0032	0.2167 ± 0.0035	0.2159 ± 0.0079	0.2153 ± 0.0282	0.4744 ± 0.1330	<u>0.2142 ± 0.0599</u>
MSRC v2	0.2826 ± 0.0012	0.2596 ± 0.0003	0.3939 ± 0.0012	<u>0.2037 ± 0.0050</u>	0.2047 ± 0.0057	0.3478 ± 0.0159	0.1940 ± 0.0237
Scene	0.5324 ± 0.0007	0.3259 ± 0.0003	0.3144 ± 0.0002	0.2926 ± 0.0071	0.3037 ± 0.0187	0.1720 ± N/A	<u>0.2787 ± 0.0161</u>
Reuters	0.0737 ± N/A	0.1331 ± 0.0005	0.0796 ± N/A	-	<u>0.0529 ± 0.0025</u>	0.1770 ± N/A	0.0505 ± 0.0111

续表

Ranking Loss ↓							
Birds	0.2465 ±	<u>0.0721 ±</u>	0.0862 ±	0.1318 ±	0.4973 ±	0.0909 ±	0.0341 ±
	0.0018	<u>0.0007</u>	0.0002	0.0140	0.0292	0.0276	0.0054
Letter Carroll	0.2736 ±	0.1349 ±	0.3489 ±	0.2078 ±	0.1683 ±	0.2377 ±	<u>0.1621 ±</u>
	0.0005	0.0001	0.0049	0.0079	0.0042	0.0417	<u>0.0209</u>
Letter Frost	0.2631 ±	<u>0.1486 ±</u>	0.3531 ±	0.1763 ±	0.1736 ±	0.2224 ±	0.1459 ±
	0.0003	<u>0.0001</u>	0.0005	0.0042	0.0081	0.0346	0.0288
MSRC v2	0.1009 ±	0.0923 ±	0.1182 ±	0.0644 ±	0.0968 ±	0.1331 ±	<u>0.0863 ±</u>
	0.0001	0.0006	N/A	0.0007	0.0017	0.0047	<u>0.0132</u>
Scene	0.3068 ±	0.1750 ±	0.1716 ±	<u>0.1668 ±</u>	0.1723 ±	0.1720 ±	0.1561 ±
	0.0002	0.0006	N/A	<u>0.0112</u>	0.0182	N/A	0.0085
Reuters	0.0264 ±	0.0454 ±	0.0243 ±	0.0177 ±	<u>0.0172 ±</u>	0.1770 ±	0.0168 ±
	N/A	0.0001	N/A	N/A	<u>0.0008</u>	N/A	0.0034
Macro-averaging F1 ↑							
Birds	0.4024 ±	0.5886 ±	0.5430 ±	<u>0.6214 ±</u>	0.1122 ±	0.5092 ±	0.7405 ±
	0.0009	0.0005	0.0003	<u>0.0515</u>	0.0316	0.0511	0.0204
Letter Carroll	0.2119 ±	0.1973 ±	0.2568 ±	<u>0.3333 ±</u>	0.1463 ±	0.1848 ±	0.3357 ±
	0.0008	0.0015	0.0037	<u>0.0013</u>	0.0344	0.0346	0.0475
Letter Frost	0.2008 ±	0.1744 ±	0.2514 ±	<u>0.2829 ±</u>	0.1263 ±	0.1924 ±	0.2918 ±
	0.0070	0.0007	0.0008	<u>0.0081</u>	0.0373	0.0595	0.0426
MSRC v2	0.2772 ±	0.3859 ±	0.3133 ±	0.1909 ±	<u>0.4319 ±</u>	0.3680 ±	0.5551 ±
	0.0003	0.0004	0.0004	0.0002	<u>0.0074</u>	0.0104	0.0285
Scene	0.4641 ±	<u>0.5851 ±</u>	0.5830 ±	-	0.5632 ±	-	0.6123 ±
	0.0003	<u>0.0002</u>	0.0005	-	0.0140	-	0.0137
Reuters	0.8487 ±	0.6400 ±	0.8551 ±	0.8629 ±	<u>0.8740 ±</u>	-	0.8769 ±
	N/A	0.0002	0.0002	0.0033	<u>0.0025</u>	-	0.0133
Hamming Loss ↓							
Birds	0.2090 ±	0.0867 ±	<u>0.0758 ±</u>	0.0835 ±	0.1619 ±	-	0.0568 ±
	0.0008	0.0003	<u>0.0003</u>	0.0239	0.0033	-	0.0025
Letter Carroll	0.1610 ±	0.1228 ±	<u>0.1246 ±</u>	0.1633 ±	0.1254 ±	-	0.1288 ±
	0.0001	0.0004	<u>0.0007</u>	0.0262	0.0034	-	0.0171
Letter Frost	0.1481 ±	0.1122 ±	<u>0.1137 ±</u>	0.1269 ±	0.1208 ±	-	0.1173 ±
	0.0001	0.0005	N/A	0.0180	0.0043	-	0.0129
MSRC v2	0.0814 ±	0.0708 ±	<u>0.0689 ±</u>	0.0697 ±	0.0690 ±	-	0.0605 ±
	N/A	0.0001	<u>N/A</u>	0.0003	0.0010	-	0.0051
Scene	0.3176 ±	0.1757 ±	0.1703 ±	<u>0.1691 ±</u>	0.1700 ±	-	0.1940 ±
	0.0001	0.0003	N/A	<u>0.0038</u>	0.0034	-	0.0065
Reuters	0.0354 ±	0.0661 ±	0.0416 ±	0.0813 ±	<u>0.0286 ±</u>	-	0.0282 ±
	N/A	0.0005	N/A	0.0012	<u>0.0004</u>	-	0.0026
Coverage ↓							
Birds	0.3792 ±	<u>0.1940 ±</u>	0.2140 ±	0.4075 ±	0.5371 ±	-	0.1345 ±
	0.0020	<u>0.0001</u>	0.0004	0.0354	0.0230	-	0.0152
Letter Carroll	0.5549 ±	0.3652 ±	0.6097 ±	0.5143 ±	0.4178 ±	-	<u>0.4054 ±</u>
	0.0015	0.0005	0.0025	0.0692	0.0106	-	<u>0.0272</u>
Letter Frost	0.5394 ±	<u>0.3798 ±</u>	0.6186 ±	0.4610 ±	0.4232 ±	-	0.3786 ±
	0.0008	<u>0.0005</u>	0.0009	0.0275	0.0114	-	0.0529
MSRC v2	0.2270 ±	<u>0.2043 ±</u>	0.2365 ±	0.2153 ±	0.2177 ±	-	0.1957 ±
	0.0007	<u>0.0001</u>	0.0003	0.0387	0.0019	-	0.0160
Scene	0.2967 ±	0.1943 ±	0.1926 ±	<u>0.1815 ±</u>	0.1929 ±	-	0.1806 ±
	0.0001	0.0005	N/A	<u>0.0125</u>	0.0136	-	0.0084
Reuters	0.0470 ±	0.0631 ±	0.0443 ±	0.1152 ±	<u>0.0377 ±</u>	-	0.0313 ±
	N/A	0.0007	N/A	0.0002	<u>0.0008</u>	-	0.0029

4.4. 消融实验

为了验证所提出模型各个关键组件的有效性, 我们分别在 Birds 数据集(包级别)和 Letter Carroll 数据集(实例级别)上进行了消融实验。实验从五个方面逐步去除或替换模块, 包括: (1) 去除对比学习项(w/o Contrastive); (2) 去除标签相关性矩阵(w/o CorrMatrix); (3) 去除相关性矩阵的正则化约束(w/o CorrReg); (4) 将 Focal Loss 替换为传统的 BCE 损失(BCE); (5) 将注意力池化替换为均值池化(w/o AttPooling)。在实验中, 我们分别在包级别(bag-level)和实例级别(instance-level)两个层面进行评估, 并选取 Average Precision (AP)、Coverage 和 Ranking Loss 三项指标作为评价标准。实验结果如表 3 和表 4 所示。

如表 3 所示, 在 Birds 数据集上的包级别结果中, 完整模型在三个指标上均取得最优性能, 其中 AP 达到 0.9315, Coverage 为 0.1345, Ranking Loss 为 0.0341。当移除对比学习或相关性正则化时, 性能均出现一定程度下降, 表明这些设计能够有效提升表征的判别性。进一步观察可以发现, 去除 CorrMatrix 或注意力池化同样会带来 AP 和 Coverage 的下降, 说明标签相关性建模与注意力机制在多标签任务中能够更好地刻画 bag 内部和标签间的复杂关系。相比之下, 将 Focal Loss 替换为 BCE 虽然也导致性能下降, 但影响相对较小。

Table 3. Bag-level ablation study results
表 3. 包级消融研究结果

Algorithm	AP $\uparrow$	Coverage $\downarrow$	Ranking Loss $\downarrow$
Full (Ours)	0.9315 $\pm$ 0.0059	0.1345 $\pm$ 0.0152	0.0341 $\pm$ 0.0054
w/o Contrastive	0.9219 $\pm$ 0.0079	0.1405 $\pm$ 0.0120	0.0355 $\pm$ 0.0051
w/o CorrMatrix	0.9266 $\pm$ 0.0072	0.1383 $\pm$ 0.0119	0.0344 $\pm$ 0.0057
w/o CorrReg	0.9246 $\pm$ 0.0087	0.1385 $\pm$ 0.0105	0.0345 $\pm$ 0.0050
BCE (w/o Focal)	0.9233 $\pm$ 0.0067	0.1357 $\pm$ 0.0109	0.0352 $\pm$ 0.0044
Mean Pooling (w/o AttPooling)	0.9243 $\pm$ 0.0089	0.1401 $\pm$ 0.0108	0.0350 $\pm$ 0.0048

5. 总结

本文针对多示例多标签学习中实例关系建模不足、标签依赖性刻画不充分以及模型鲁棒性较弱等问题, 提出了一种基于相关性建模的图注意力 MIML 方法。该方法在图构建阶段显式建模示例间的语义相关性, 使得图表示学习能够更好地捕捉全局依赖; 在实例聚合阶段引入多头注意力池化机制, 实现对关键信息的自适应选择; 在标签预测阶段结合标签相关性矩阵, 提升了多标签预测结果的整体一致性; 同时, 融入对比学习策略, 在全局层面对表示空间进行约束, 有效增强了模型的判别性与鲁棒性。实验结果表明, 该方法在多个公开数据集和评价指标上均取得了优于现有方法的性能, 验证了其在 bag 级与 instance 级任务上的有效性。未来, 我们将进一步探索该框架在多模态输入、未知标签以及不完全监督场景下的扩展与应用, 以提升其在更复杂实际任务中的适应性与泛化能力。

Table 4. Friedman statistical p-values for each evaluation metric
表 4. 针对每个评估指标的 Friedman 统计 p 值

Evaluation metrics	Bag level p-values	Instance level p-values
Average Precision	0.0058	0.0719
One Error	0.0336	0.0293
Ranking Loss	0.0051	0.0421

续表

Macro F1	0.0001	0.0421
Hamming Loss	0.0034	0.0293
Coverage	0.0016	0.0293

基金项目

国家自然科学基金项目：基于模块化推理的多标签深度概率图神经网络研究(62206297)，掘进巷道围岩加卸载破坏震源特征及灾变自动预警方法研究(52174221)资助。

参考文献

- [1] Zhang, Z., Cui, P. and Zhu, W. (2020) Deep Learning on Graphs: A Survey. *IEEE Transactions on Knowledge and Data Engineering*, **34**, 249-270. <https://doi.org/10.1109/tkde.2020.2981333>
- [2] Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P. and Bengio, Y. (2018) Graph Attention Networks. *International Conference on Learning Representations (ICLR)*, Vancouver, 30 April 2018, 1-12.
- [3] Zhou, J., Cui, G., Hu, S., Zhang, Z., Yang, C., Liu, Z., et al. (2020) Graph Neural Networks: A Review of Methods and Applications. *AI Open*, **1**, 57-81. <https://doi.org/10.1016/j.aiopen.2021.01.001>
- [4] Zhou, Y., Zhang, X. and Huang, T. (2019) Multi-Instance Multi-Label Learning with Sparse Feature Selection. *IEEE Transactions on Knowledge and Data Engineering*, **31**, 1093-1106.
- [5] Chen, X., Fan, H., Girshick, R. and He, K. (2020) Improved Baselines with Momentum Contrastive Learning. *arXiv:2003.04297*.
- [6] Wang, Y., Mao, C., Wu, C. and Wong, K.C. (2022) Contrastive Graph Learning for Graph Neural Network Pre-Training. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **45**, 5511-5525.
- [7] Yan, J., Yang, M. and Zhang, C. (2018) Multi-Instance Multi-Label Learning with Applications to Scene Classification. *International Journal of Computer Vision*, **126**, 398-425.
- [8] Xu, C., Tao, D. and Xu, C. (2016) Multi-Label Multi-Instance Learning with Application to Scene Classification. *Advances in Neural Information Processing Systems (NeurIPS)*, Barcelona, 5 December 2016, 1-9.
- [9] 王立威, 张学工. 多示例多标签学习研究进展[J]. 计算机学报, 2017, 40(8): 1745-1762.
- [10] Zhou, Z. and Zhang, M. (2006) Multi-Instance Multi-Label Learning with Application to Scene Classification. In: Schölkopf, B., Platt, J. and Hoffman, T., Eds., *Advances in Neural Information Processing Systems 19*, The MIT Press, 1609-1616. <https://doi.org/10.7551/mitpress/7503.003.0206>
- [11] Ying-Xin Li, Shuiwang Ji, Kumar, S., Jieping Ye, and Zhi-Hua Zhou, (2012) Drosophila Gene Expression Pattern Annotation through Multi-Instance Multi-Label Learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, **9**, 98-112. <https://doi.org/10.1109/tcbb.2011.73>
- [12] Zhang, M. and Zhou, Z. (2008) M3MIML: A Maximum Margin Method for Multi-Instance Multi-Label Learning. 2008 *Eighth IEEE International Conference on Data Mining*, Pisa, 15-19 December 2008, 688-697. <https://doi.org/10.1109/icdm.2008.27>
- [13] Zhou, Z., Zhang, M., Huang, S. and Li, Y. (2012) Multi-Instance Multi-Label Learning. *Artificial Intelligence*, **176**, 2291-2320. <https://doi.org/10.1016/j.artint.2011.10.002>
- [14] Zhang, M. and Wang, Z. (2009) MIMLRBF: RBF Neural Networks for Multi-Instance Multi-Label Learning. *Neurocomputing*, **72**, 3951-3956. <https://doi.org/10.1016/j.neucom.2009.07.008>
- [15] Zhang, M. (2010) A K-Nearest Neighbor Based Multi-Instance Multi-Label Learning Algorithm. 2010 *22nd IEEE International Conference on Tools with Artificial Intelligence*, Arras, 27-29 October 2010, 207-212. <https://doi.org/10.1109/ictai.2010.102>
- [16] Li, Y., Hu, J., Jiang, Y. and Zhou, Z. (2021) Towards Discovering What Patterns Trigger What Labels. *Proceedings of the AAAI Conference on Artificial Intelligence*, **26**, 1012-1018. <https://doi.org/10.1609/aaai.v26i1.8285>
- [17] Huang, S., Gao, W. and Zhou, Z. (2019) Fast Multi-Instance Multi-Label Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**, 2614-2627. <https://doi.org/10.1109/tpami.2018.2861732>
- [18] Feng, J. and Zhou, Z. (2017) Deep MIML Network. *Proceedings of the AAAI Conference on Artificial Intelligence*, **31**, 1884-1890. <https://doi.org/10.1609/aaai.v31i1.10890>

-
- [19] Yang, M., Tang, W. and Min, F. (2022) Multi-Instance Multi-Label Learning Based on Parallel Attention and Local Label Manifold Correlation. 2022 *IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA)*, Shenzhen, 13-16 October 2022, 1-10. <https://doi.org/10.1109/dsaa54385.2022.10032401>
 - [20] Qiu, S., Wang, M., Yang, Y., Yu, G., Wang, J., Yan, Z., *et al.* (2023) Meta Multi-Instance Multi-Label Learning by Heterogeneous Network Fusion. *Information Fusion*, **94**, 272-283. <https://doi.org/10.1016/j.inffus.2023.02.010>
 - [21] Chen, T., Kornblith, S., Norouzi, M. and Hinton, G. (2020) A Simple Framework for Contrastive Learning of Visual Representations. *International Conference on Machine Learning (ICML)*, 1597-1607.
 - [22] You, Y., Chen, T., Sui, Y., Chen, T., Wang, Z. and Shen, Y. (2020) Graph Contrastive Learning with Augmentations. *Advances in Neural Information Processing Systems (NeurIPS)*, **33**, 5812-5823.
 - [23] Zhang, M. and Zhou, Z. (2014) A Review on Multi-Label Learning Algorithms. *IEEE Transactions on Knowledge and Data Engineering*, **26**, 1819-1837. <https://doi.org/10.1109/tkde.2013.39>
 - [24] Wu, J., Wang, W., Song, L. and Chen, Z. (2021) Multi-Label Graph Convolutional Networks with Label Correlations. *Proceedings of the AAAI Conference on Artificial Intelligence*, **35**, 10078-10086.
 - [25] Sebastiani, F. (2002) Machine Learning in Automated Text Categorization. *ACM Computing Surveys*, **34**, 1-47. <https://doi.org/10.1145/505282.505283>
 - [26] Dai, J., Huang, W., Zhang, C. and Liu, J. (2024) Multi-Label Feature Selection by Strongly Relevant Label Gain and Label Mutual Aid. *Pattern Recognition*, **145**, Article 109945. <https://doi.org/10.1016/j.patcog.2023.109945>
 - [27] Zhang, Y., Huo, W. and Tang, J. (2024) Multi-Label Feature Selection via Latent Representation Learning and Dynamic Graph Constraints. *Pattern Recognition*, **151**, Article 110411. <https://doi.org/10.1016/j.patcog.2024.110411>