

MGE-BERT: 融合标签排序与情感图感知的多标签情感分类模型^①



关 慧^{1,2}, 韩志远¹

¹(沈阳化工大学 计算机科学与技术学院, 沈阳 110142)

²(辽宁省化工过程工业智能化技术重点实验室, 沈阳 110142)

通信作者: 韩志远, E-mail: zyhan6857@163.com

摘 要: 目前多标签文本分类研究已经整合了标签信息, 但在情感分析领域, 现有方法常忽略了基于情感本身强度和极性的标签相关性, 这些相关性对于精确分类至关重要. 为了解决这些问题, 本文提出了多标签交互和情感图感知的 MGE-BERT 模型. 该模型首先通过情感强度关联和情感层次关联对情感标签进行优先排序, 然后将排序后的标签与文本数据结合, 作为输入导入 BERT 模型. 在此过程中, 采用了句法分析技术和情感词典, 通过独特的构图方法构建了复杂的依赖图和情感图. 为了进一步增强标签信息与文本特征的深度融合, 本文将 BERT 模型的输出作为图卷积网络 (GCN) 的输入, 使其能够更精确地捕捉并传递节点间的上下文关系. 实验结果表明, 在 SemEval2018 Task-1C 数据集和 GoEmotions 数据集上进行的实验中, 本文提出的 MGE-BERT 模型相比于最先进的模型, *Macro-F1* 得分分别提高了 1.6% 和 2.0%.

关键词: 多标签情感分类; 情感标签排序; 情感图感知; 情感词典; 情感强度加权

引用格式: 关慧, 韩志远. MGE-BERT: 融合标签排序与情感图感知的多标签情感分类模型. 计算机系统应用, 2025, 34(3): 268–276. <http://www.c-s-a.org.cn/1003-3254/9803.html>

MGE-BERT: Multi-label Sentiment Classification Model Integrating Label Ranking and Sentiment Graph Perception

GUAN Hui^{1,2}, HAN Zhi-Yuan¹

¹(College of Computer Science and Technology, Shenyang University of Chemical Technology, Shenyang 110142, China)

²(Key Laboratory of Intelligent Technology for Chemical Process Industry in Liaoning Province, Shenyang 110142, China)

Abstract: Currently, research on multi-label text classification integrates label information. However, in the field of sentiment analysis, existing methods often overlook the correlations of labels based on the intensity and polarity of emotions themselves, which are crucial for accurate classification. To address these issues, this study proposes the MGE-BERT model which features multi-label interaction, graph enhancement, and emotion perception. The model first prioritizes sentiment label sorting through the correlations of sentiment intensity and hierarchy and then combines these sorted labels with text data as inputs into the BERT model. During this process, syntactic analysis techniques and sentiment lexicons are employed, and through a unique graph construction method, intricate dependency and emotion graphs are built. To further enhance the in-depth integration of label information and text features, the study uses BERT outputs as inputs to graph convolutional network (GCN), enabling it to capture and transmit contextual relationships between nodes more precisely. Experimental results demonstrate that the proposed MGE-BERT model outperforms state-of-the-art models, achieving improvements in *Macro-F1* scores by 1.6% and 2.0% on the SemEval2018 Task-1C and

① 收稿时间: 2024-09-09; 修改时间: 2024-10-14; 采用时间: 2024-10-23; csa 在线出版时间: 2025-01-21

CNKI 网络首发时间: 2025-01-22

GoEmotions datasets, respectively.

Key words: multi-label sentiment classification; sentiment label sorting; sentiment graph perception; sentiment dictionary; sentiment intensity weighting

情感是传达思想的核心要素. 情感分析的目的是识别个体对某事物的看法, 这可能表现为情感倾向 (正面或负面) 或具体的情绪状态 (例如快乐、愤怒、悲伤等). Ekman 等^[1]识别出 6 种基本情绪: 愤怒、厌恶、快乐、恐惧、悲伤和惊讶. Plutchik 等^[2]和 Parrot 等^[3]也提出了他们的情绪分类理论, Plutchik 采用情绪轮盘模型, 而 Parrot 则使用树状模型, 这些理论为我们提供了理解情感复杂性和多样性的不同视角.

在多标签情感分析任务中, 由于不同情感标签之间存在内在联系, 因此对情感标签进行排序^[4]变得尤为重要. 这些标签可以根据情绪的强度或其他相关属性进行排序. 例如, 在某些文本中, 惊讶、激动和快乐可能会同时出现, 而厌恶和信任等情绪则可能较少出现或不明显. 这种排序机制在情感分析中相对容易实现, 而在其他类型的分类任务, 如新闻文章分类中, 一篇文章可能同时包含“政治”“经济”和“社会”等多个主题, 这些主题标签之间并不存在层级或强度的关联, 因此不适合进行排序. 引入情感标签排序机制可以帮助我们优先识别和处理更显著的情感标签, 从而提升模型对文本情感的识别能力和分类的准确性.

1 相关工作

近年来, 多标签分类技术因其在文本分类、图像标注和生物信息学等多个领域的广泛应用而受到关注, 这主要得益于数据量的增长和复杂性的提升. 传统处理方法主要分为两类: 一种是问题转换法, 例如 binary relevance (BR)^[5]方法, 它将多标签问题分解为多个二元或多类问题; 另一种是问题适应法, 如 ranking support vector machine (RankSVM)^[6], 它通过改进现有算法以更好地适应多标签任务的需求. 近年来, 基于深度学习的技术显著推动多标签分类技术的进步, 这些技术包括卷积神经网络 (CNN)、循环神经网络 (RNN) 和 Transformer 模型^[7]. 例如, 有研究者利用 CNN 来构建词向量, 进而应用于多标签分类任务, 尽管这种方法可能会忽略文本的全局语义信息. 另一些研究者提出采用层次化注意力机制 LSTM 模型, 以捕捉文本的局部

和全局特征. 还有研究展示了 FastText 模型在新闻文本情感分析中的高效性能.

在情感识别研究领域, 由 NTUA-SLP 团队^[8]开发的基于多层自注意力机制的双向长短时记忆网络 (BiLSTM) 在 SemEval2018 的竞赛任务中取得显著成绩. PlusEmo2Vec^[9]通过扩展训练数据集, 提升了情感分析的精确度. Ameer 等^[10]利用图表示的优势, 提出图注意力网络, 该网络能够整合语义和句法信息, 为多标签情感分类提供一种新颖的方法. 为更有效地利用多标签分类任务中标签之间的相互关系, Pal 等^[11]提出了 MAGNET 模型, 该模型利用图神经网络来模拟标签间的内在联系, 并结合 BiLSTM 来捕捉文本数据的上下文信息. 然而, 现有的多标签情感分类方法在文本特征提取和情感关联性分析方面仍有待提高. 针对这一挑战, 本文介绍了 MGE-BERT 模型, 其创新之处如下.

(1) 根据本文所知, 这是第 1 次将基于情绪强度和层级的情感标签排序同时应用于全局信息和图构建的群体中.

(2) 本文构建了情感标签和文本的依赖性及情感图, 通过考虑句子的情绪极性、利用情感词典以及标签之间的相关性, 动态地分配权重.

(3) 本文采用图卷积网络来优化图和 BERT 模型的输出, 从而提高模型在多标签情感文本分类中的准确性和泛化能力.

2 MGE-BERT 模型

本文提出了一种专注于情感文本的多标签分类模型 MGE-BERT, 如图 1 所示, 该模型的结构由几个关键层组成, 包括: 数据预处理层、BERT 层、图卷积层、特征融合层和分类层^[12]. 在数据预处理层, 对数据集进行去噪和优化, 仅保留具有语义信息的文本; BERT 层在自注意力机制下考虑上下文语义信息; 图卷积层在图结构上进行卷积操作, 以有效地学习和传播图数据中的特征信息; 特征融合层进一步整合前面的特征信息; 最后, 通过加权求和的方式进行句子情感极性的预测.

2.1 数据预处理

由于社交媒体的评论往往受到字数限制, 用户通常会使用缩写和简写的方式来表达自己的意见, 例如: “lol”代表“笑”, “omg”代表“天啊”等. 另外, 用户可以使用哈希标签(#)来标记主题或话题, 例如“#COVID19”. 以及用户经常在评论中使用“@”符号来与其他用户进

行互动. 为避免以上对情感分析的结果产生干扰, 本文使用针对 Twitter 评论的特定特征而设计的数据预处理工具^[13]. 该工具提供了不同的功能, 如标记化、规范化、拼写纠正和分段. 本文使用该工具对文本进行标记, 将单词转换为小写, 规范化用户提及、URL 和重复字符.

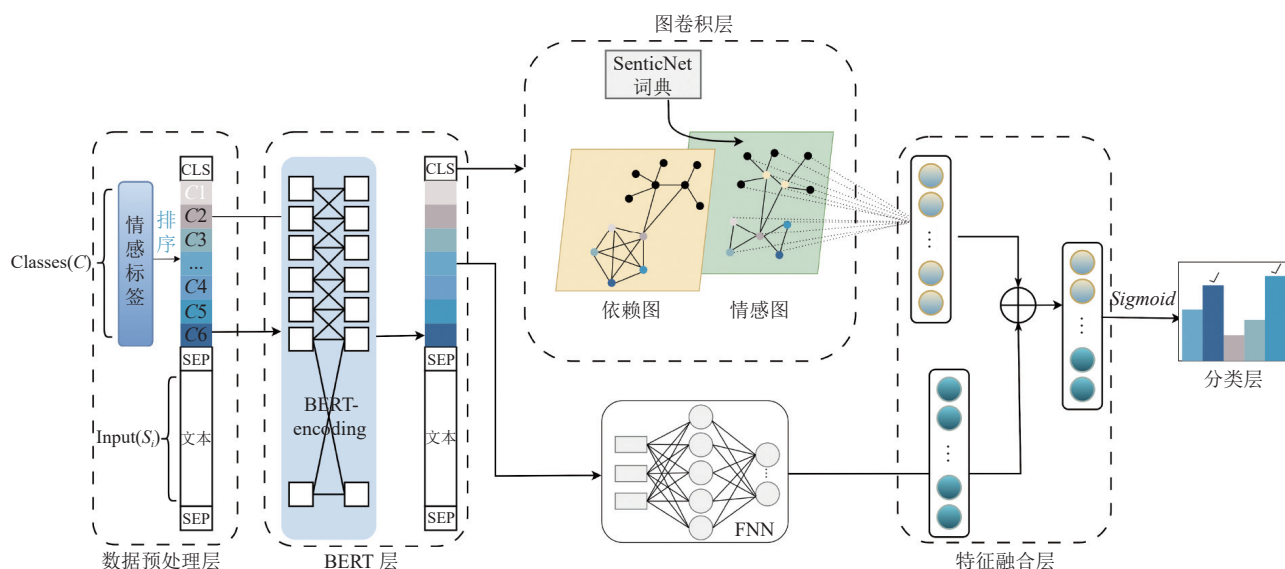


图1 MGE-BERT 模型框架

2.2 BERT 特征提取

本文采用预训练模型 BERT 来进行文本特征的提取. 为了能够让计算机处理文本序列, 需要将文本序列转换成数值向量. 序列的每一部分都会被嵌入层投影到一个 d 维的连续向量空间中. 考虑一个由 n 个元素组成的标记序列 X , 序列中的每个元素 x_i 代表第 i 个标记. 在词嵌入后, 将得到一个 $n \times d$ 维数的矩阵. 在每个序列的开始处添加一个特殊的“[CLS]”标记, 并将句子嵌入添加到词嵌入中. 为捕获词序, 还添加了位置嵌入. 词嵌入、句子嵌入和位置嵌入的嵌入维度是相同的. 输入序列在嵌入层完成后被送入 BERT 模块. BERT 模块是由堆叠的 Transformer 组成的. 该模块在输入嵌入层加上了位置编码, 然后再通过多头注意力机制模块计算多头自注意力, 最后通过前馈神经网络模块、残差连接和归一化层得到最后的输出.

在 BERT 模型中, 文本特征提取的复杂性尤为显著, 尤其是在多标签情感分类任务中. 通过以往的研究认识到, 同一词汇在不同位置出现时, 其语境和含义可能会有所不同. 为了更准确地捕捉这种变化, 本文特别

考虑了情感强度的关联和情感层次的因素, 这两个关键因素对于理解文本的整体情感至关重要.

首先, 情感标签之间的相互联系是不可忽视的. 情感标签之间存在相互关联, 且同一词汇在不同的情感背景下可能表现出不同的情感色彩. 例如, 单词“happy”在描述积极事件时可能与“joy”相关联, 而在描述悲伤事件时可能与“sadness”相关联. 这种情感强度的关联性对于理解文本的整体情感至关重要. 本文通过计算每个情感标签的情感强度 S_i 来量化这种关联性, 这反映了标签在文本中表达情感的显著性. 情感强度的计算公式如下:

$$S_i = \sum_{j \in L} f_{ij} \times w_j \quad (1)$$

其中, f_{ij} 表示标签 i 与标签 j 同时出现的次数, w_j 表示标签 j 的权重. 其次, 情感层次的因素也是我们在模型设计中考虑的. 在情感任务中, 相同极性的情感标签 (如“joy”和“happiness”) 更容易同时出现, 而不同极性的情感标签 (如“joy”和“sadness”) 则不太可能共现. 这种情感层次的关联要求模型能够有效地捕捉并处理复

杂的情感关联. 我们通过构建情感层次关联 H_{ij} 来反映标签 i 和 j 之间的情感关联度, 公式如式 (2):

$$H_{ij} = \frac{1}{1 + e^{-\alpha(S_i - S_j)}} \quad (2)$$

其中, α 是一个用于控制关联敏感度的参数. 为确保排序结果能够反映标签在文本中的整体重要性, 本文结合情感强度和情感层次关联, 计算每个标签的综合排序值 R_i . 综合排序值的计算公式如式 (3):

$$R_i = S_i \times \sum_{j \in L} H_{ij} \quad (3)$$

其中, R_i 表示标签 i 的综合排序值. S_i 是标签 i 的情感强度. H_{ij} 是标签 i 和 j 之间的情感层次关联度. 通过这种方法, BERT 模型能够更有效地处理文本中的情感信息, 尤其是在多标签情感分类任务中. 这不仅能够提升情感信息的提取效果, 还减少了情感标签在分类过程中的混淆, 使得模型在处理复杂情感表达时更为精确和有效.

最终, 本文使用一个基于优先队列的算法来根据综合排序值 R_i 对标签进行排序. 算法步骤如下.

- 初始化一个空的优先队列 Q .
- 对于每个标签 i , 计算其综合排序值 R_i .
- 令标签 i 和对应的 R_i 插入到队列 Q 中.
- 从优先队列 Q 中依次取出标签, 按照 R_i 的值从高到低进行排序.

图2呈现了最终的排序结果. 本文通过将11种情绪按消极到积极的顺序排列, 得到一个高效的标签排序体系: 包括悲观、悲伤、害怕、厌恶、愤怒、期待、惊讶、信任、爱、喜悦和乐观. 在这一体系的最高层, 积极与消极情绪的界限清晰, 这极大地提升了多标签情感分类的精确度. 更深层次的结构则为不同层次的情感聚类提供了合理的框架.

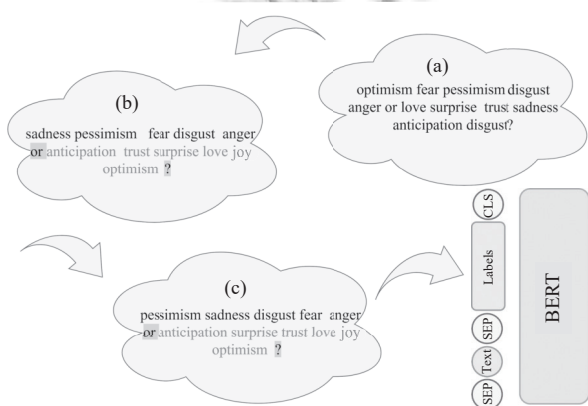


图2 情感标签排序

在图2中, 文本特征提取的流程被展示出来. 其中, 情感标签首先根据情绪的强度和层级关系进行排序, 随后与文本内容结合, 形成模型的输入. 这种排序方法让 MGE-BERT 模型在面对多标签情感分类任务时, 能更精确地识别文本中的情感元素, 进而提升模型的分类效果.

2.3 GCN 特征提取

BERT 模型虽然得到各个标签的特征, 但是不能充分挖掘出标签内部的关系. 本文构建依赖图和情感图, 与其他构图方法不同的地方如下.

(1) 本文将标签和文本串联起来, 共同构建图, 考虑到情绪标签之间的相关性. 这一步骤有助于捕捉标签内的潜在关系, 从而更全面地理解情感信息.

(2) 本文使用情感词典作为边的权重, 并根据句子的情绪极性调整边的权重, 以增强图中情感词的重要性. 这不仅有助于解决矩阵中潜在的稀疏性问题, 还强调了情感词的关键作用, 为模型提供更有针对性的信息.

构图方法如下: 依赖图 D 通过 Spacy 句法分析得到依赖关系树, 并将关系信息映射到一个 $n \times n$ 的矩阵中, 形成句子中单词之间的连接:

$$D_{i,j} = 1, \text{ if } T(w_i, w_j) \quad (4)$$

其中, $D_{i,j}$ 是单词 w_i 和 w_j 在句子中的依赖关系. 由于需要维护情感信息和依赖信息, 构造无向图 $D_{i,j}=D_{j,i}=1$, 并设置自环 $D_{i,i}=1$. 在依赖图 D 基础上引入 SenticNet 情感词库, 通过创建情感图来构建情感矩阵. 设 S 为一个有 n 个单词的句子, 用 $S = \sum w_i$ 表示, 情感矩阵是基于从外部情感常识知识中获得的单词的情感分数建立的:

$$A_{i,j} = |S(w_i) - S(w_j)| \quad (5)$$

其中, $A_{i,j}$ 表示单词之间的权重, $S(w_i) \in [-1, 1]$ 表示从 SenticNet 获得的单词 w_i 的整体情感得分, 如果存储库中没有保存 w_i , 则 $S(w_i) = 0$.

在构建依赖图的过程中, 本文注重于句子的句法结构, 将情感标签与文本单词同等对待, 通过 Spacy 句法分析工具识别并映射单词间的依赖关系, 形成无向图中的节点连接. 这一步骤不区分情感标签之间的极性, 而是将它们作为整体文本结构的一部分.

然而, 在构建情感图时, 本文首先考虑句子的整体情感倾向. 这一步骤中, 我们采用了不同于依赖图的连接策略. 具体来说, 我们取消了积极和消极情感标签之

间的直接连接,以避免不同极性标签间的潜在冲突.这种策略允许模型更清晰地捕捉和表达文本的情感一致性.此外,本文根据句子的极性对情感标签进行附加权重.这意味着对于积极倾向的句子,我们增加积极情感标签的权重;而对于消极倾向的句子,我们增加消极情感标签的权重.假设 P 和 N 分别代表积极和消极的情感标签的权重,则可通过式 (6) 调整情感标签的权重:

$$W_{label} = \begin{cases} P \times S(label), & \text{if sentence is positive} \\ N \times S(label), & \text{if sentence is negative} \end{cases} \quad (6)$$

这两种图的使用增强了图结构中的信息,因为从依赖图获得的词关系和从情感图推断出的语义分数产生了用于理解给定文本数据中的情感差异.

图3 通过例句“I love it when people trust me”构建

依赖关系树,图4分别展示了如何构建依赖图和情感图.在依赖图中,边代表单词之间的连接关系,所有边的权重均为1,这反映了句子内部单词的结构联系.与此不同,情感图中的边具有不同的权重,这些权重表示单词或情感之间的关联强度.由于该句子主要表达积极情绪,因此负面情感标签的权重被大幅降低.这种处理方式使模型能够更深入地理解句子中单词间的情感联系及其上下文,从而提高情绪分析的准确性.

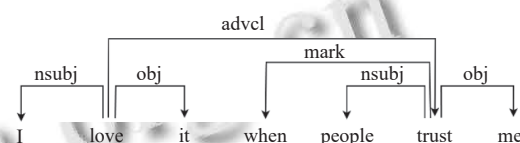


图3 依赖关系树

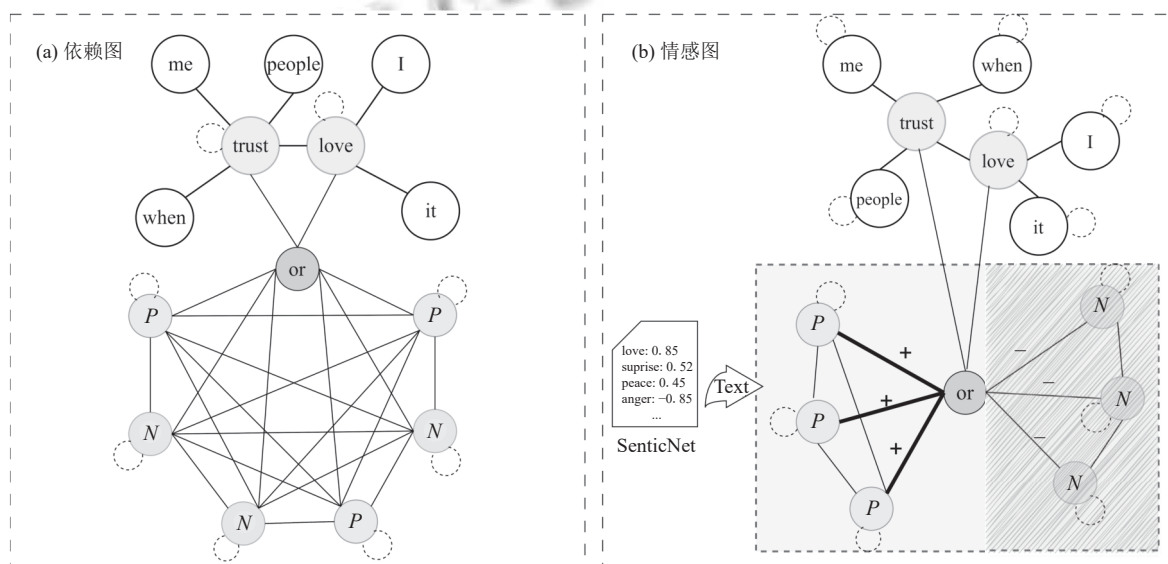


图4 构建依赖图和情感图

对于数据集中的每个句子,本文通过将句子与情感标签结合起来构建依存图,单词作为图的节点,单词间的连接则作为边.依存图的结构基于句子的树状层级.此外,本文利用情感词典和每个单词的情感价值权重,基于依存图生成情感图,并移除情感联系较弱的边.这些权重体现了单词与情感标签间的情感强度.通过结合依存图和情感图,模型能够更全面地理解句子中单词间的情感关系及其上下文,从而有助于模型得出更全面的情感分析结果.

2.4 特征融合

BERT 的输出分为两部分,一是将 BERT 的初始上下文表示与创建的情感图和依赖图一起传递给图卷积

网络.这些图被交替地传递给图卷积网络,以便每个节点的表示在给定的 GCN 层上根据从其邻域获得的隐藏表示进行更新.因此,每个节点从相邻节点的信息流和之前的输出表示中学习其权重:

$$g^l = \text{ReLU}(H \cdot \text{ReLU}(\text{Ag}^{l-1} W_a^l + b_a^l) W_d^l + b_d^l) \quad (7)$$

其中, g^{l-1} 是每个 GCN 层生成的潜在节点表示, H 是第 1 个图卷积层的输入,为 BERT 学习到的初始上下文表示. W_a^l 和 b_a^l 是图卷积网络第 l 层中的可学习的权重矩阵和偏置向量,为后续信息传递和节点表示更新做准备. W_d^l 和 b_d^l 是第 l 层第 2 组可学习的权重矩阵和偏置向量,综合考虑了之前的信息和图结构信息.二是

提取 BERT 输出的最后一层与情感标签相同的词对应的词向量, 并进一步引入一个前馈网络 (FFN), 该网络由一个具有 $\tanh$ 激活的非线性隐藏层 ($f_i(H_i)$) 和一个位置向量 p_i 组成, 用于计算 f_i 和 p_i 的输出之间的点积. 最后采用融合策略将二者特征信息 T 、 F 相结合, 由于情感分类任务涉及多标签情绪分类, 模型添加了一个 Sigmoid 激活来确定分类是否是正确的情绪标签:

$$T = W_o g^l + b_o \quad (8)$$

$$F = FFN(H_i) \quad (9)$$

$$\hat{y} = \text{Sigmoid}(T \oplus F) \quad (10)$$

其中, W_o 解释为可学习的权重矩阵, 训练中会调整其值以使 g^l 经线性变换更好地适应分类任务. b_o 解释为可学习的偏置向量, 为线性变换提供偏移量, 增加模型表现力和适应性.

2.5 损失函数

在实验中损失函数采用了标签相关感知损失 (LCA), 它以一个真实二进制标签向量 (y) 和一个概率向量 ($\hat{y}$) 作为输入:

$$\mathcal{L}_{\text{LCA}}(y, \hat{y}) = \frac{1}{|y^0||y^1|} \sum_{(p,q) \in y^0 \times y^1} \exp(\hat{y}_p - \hat{y}_q) \quad (11)$$

其中, y^0 表示负标签集, y^1 表示正标签集. $\hat{y}_p$ 表示向量 $\hat{y}$ 的第 p 个元素. 该损失函数通过扩大标签之间的差异来保留标签之间的依赖关系. 换句话说, 当模型预测到一对给定示例中不应该共存的标签时, 它应该受到惩罚.

3 实验结果

3.1 数据集

本文在两个多标签情绪分类数据集上进行了广泛的实验, 即 SemEval2018 Task-1C^[14] 和 GoEmotions^[15] 数据集, 数据集概况如表 1 所示.

表 1 数据集概况

Dataset	Train	Valid	Test	Total
SemEval2018 Task-1C	6838	886	3259	10983
GoEmotions	43410	5426	5427	54263

SemEval2018 Task-1C 数据集包含 2016–2017 年收集的 Tweets, 数据集中标注了 11 种情绪 (anger, anticipation, disgust, fear, joy, love, optimism, pessimism, sadness, surprise, and trust). 数据集提供了 3 种语言 (阿

拉伯语、西班牙语和英语) 用于情绪分类任务, 本文只使用了英语语言数据.

GoEmotions 是一个细粒度情绪数据集, 包含 58k 个来自 Reddit 评论的样本. 该数据集使用 28 种情绪 (27 种情绪类别及中性) 进行注释. 本文将 28 个标签映射到 Ekman 的 6 个类别 (anger, disgust, fear, joy, sadness and neutral). 并将数据集随机分为训练集 (80%)、开发集 (10%) 和测试集 (10%), 如文献[16]中的定义.

3.2 实验评价指标

本文的实验中, 使用 Micro-F1、Macro-F1、Jaccard 作为评价指标, 各个指标的具体计算公式如下:

$$\text{Micro-F1} = \frac{\sum_{i=1}^L 2TP_i}{\sum_{i=1}^L 2TP_i + FP_i + FN_i} \quad (12)$$

$$\text{Macro-F1} = \frac{1}{L} \sum_{i=1}^L \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (13)$$

$$\text{Jaccard} = \frac{1}{|D|} \sum_{t \in D} \frac{G_t \cap P_t}{G_t \cup P_t} \quad (14)$$

其中, i 表示标签集合中第 i 类标签, TP_i , FP_i , FN_i 分别代表真正例, 假正例, 假负例. G_t 表示句子 t 的真实标签集, P_t 表示句子 t 的预测标签集, 而 D 表示测试集中句子的总数.

3.3 参数设置

在本项实验中, 我们使用 Python 3.6 编程语言, 并基于 PyTorch 框架进行开发. 实验中选用了 hugging face 提供的 BERT-base-uncased 模型作为文本和标签嵌入的基础. 在模型配置上, 我们设定了 12 层的 BERT 隐藏层, 每层有 768 个单元, 同时采用了两层的图卷积网络 (GCN), 每层的维度同样为 768. 为了优化模型, 我们使用了 Adam 优化器, 并设置了 0.1 的 dropout 比率以减少过拟合的风险. 在这些设置下, 我们对模型进行了 12 轮的训练, 以确保模型性能达到最佳状态.

4 实验结果

针对模型效果, 本文从 3 个不同的实验进行分析验证, 分别是对模型整体性能的分析、消融实验分析和不同情感标签排序对结果的影响.

4.1 整体性能对比

由于本文模型需要对有语义的情感标签进行处理, 因此选择 SemEval2018 Task-1C 数据集和 GoEmotions 数据集来测试以证明该模型的有效性, 并选择了以下

多标签分类的方法作为对比实验。

(1) JBNN^[17]. 引入了一种联合二元神经网络, 该网络基于 Plutchik 的情绪轮理论^[2], 重点学习情绪之间的关系, 然后通过积分进行多标签情绪分类

(2) RERc^[18]. 定义了一种情绪相关损失排序方法, 其重点是将情绪关系纳入损失函数, 以提高情绪预测和相关情绪的排序。

(3) DATN^[19]. 提出了一种双重注意力转移网络, 通过情感分类来改进多标签情感分类

(4) NTUA^[13]. 是 SemEval2018 比赛中排名第一的模型, 因为它依赖于不同的预训练和微调策略。

(5) BERT+DK^[20]. 使用与本文相同的编码器, 考虑额外的领域知识 (DK)。

(6) SpanEmo^[21]. 旨在将多标签情感分类作为一个跨度预测问题, 特别是在输入句子中添加情绪类别, 这有助于模型学习特定情绪关联并提高其性能。

(7) LEM^[22]. 引入了一种潜在情绪记忆网络, 其中潜在情绪模块通过变分自编码器学习情绪分布, 而记忆模块捕获与每种情绪对应的特征。

表 2 中的数据清楚地展示了 MGE-BERT 模型在 SemEval2018 Task-1C 数据集上的性能表现。MGE-BERT 在 *Micro-F1* 和 *Macro-F1* 指标上均优于其他模型, 显示出其在多标签情感分类任务中的优越性。*Micro-F1* 得分为 0.714, 这一成绩意味着模型在综合考虑所有情绪标签时, 能够保持较高的精确度和召回率, 尤其是在处理大量标签共现的情况下。*Macro-F1* 得分 0.586, 比 BERT+DK 模型高出 3.7%, 这表明 MGE-BERT 在处理各个独立情绪标签时更为均衡, 即使是那些出现频率较低的情绪标签也能被有效识别。此外, *Jaccard* 指数 0.595 的得分, 显示了模型在预测情绪标签时的准确性, 即模型能够精确地识别出文本中实际存在的情绪标签, 而不是过度预测不相关的标签。

表 2 SemEval2018 Task-1C 数据集实验结果

模型名	<i>Micro-F1</i>	<i>Macro-F1</i>	<i>Jaccard</i>
JBNN	0.630	0.528	0.556
RERc	0.651	0.539	0.567
DATN	0.668	0.551	0.583
NTUA	0.701	0.528	0.588
BERT	0.695	0.520	0.570
BERT+DK	0.710	0.549	0.591
LEM	0.675	0.567	0.592
SpanEmo	0.713	0.578	0.601
MGE-BERT (本文)	0.714	0.586	0.595

(1) RoBERTa^[23]. 使用了更大规模的训练数据, 将训练过程中的某些超参数进行了优化, 能够更好地理解语言的含义和上下文信息, 在许多 NLP 任务上取得了比 BERT 更好的表现。

(2) Dim-RoBERTa^[24]. 结合了预训练模型 RoBERTa 和无监督的对比学习目标, 通过将正样本与负样本进行对比, 学习到样本之间的相似性和差异性, 从而达到更准确的语义表示。

根据表 3 的实验结果可以看出, MGE-BERT 在 *Micro-F1*、*Macro-F1* 和 *Jaccard* 得分分别为 0.715、0.628 和 0.678。这表明在文本特征提取方面 MGE-BERT 模型能够更准确地表达文本的语义含义, 并且具有更好的泛化能力。此外, 在标签相关性方面, MGE-BERT 模型通过引入注意力机制来建立标签之间的依赖关系和相关性。通过对标签的权重的调整, MGE-BERT 模型能够更好地捕捉标签之间的相关性, 从而提高多标签分类的准确性。尤为重要的是, 实验结果还证明了即使在标签集相对较少的情况下, MGE-BERT 模型依然能够提高准确度, 这一点在多标签情感分类任务中尤为重要, 因为它显示了模型在处理不同规模数据集时的鲁棒性和有效性。

表 3 GoEmotions 数据集实验结果

模型名	<i>Micro-F1</i>	<i>Macro-F1</i>	<i>Jaccard</i>
BERT	0.622	0.640	0.602
RoBERTa	0.691	0.618	0.659
Dim-RoBERTa	0.686	0.610	0.657
SpanEmo	0.710	0.611	0.672
MGE-BERT (本文)	0.715	0.628	0.678

综上所述, SemEval2018 Task-1C 的实验中, MGE-BERT 在 *Macro-F1* 指标上的优异表现可能是因为它能够处理情绪标签的复杂性。该模型利用情感标签排序和图卷积技术, 精确地识别了 Tweets 中的情绪细微差。对 GoEmotions 数据集, MGE-BERT 在 *Micro-F1* 和 *Jaccard* 上的优势可能源于其处理长文本的能力, 这些文本来自 Reddit, 包含更丰富的背景信息。模型融合了 BERT 的深度语义解析和 GCN 的细节特征提取, 更准确地判断了文本的情绪倾向。

4.2 消融实验分析

如表 4 所示, 在 SemEval2018Task-1C 数据集的消融实验中, 我们发现 MGE-BERT 模型的各个组成部分都对性能有积极作用。基础 BERT 模型的 *Micro-F1*、

Macro-F1 和 *Jaccard* 指数分别为 0.672、0.551 和 0.546。加入 GCN 后, 这些指标提升至 0.692、0.580 和 0.564, 显示出 GCN 在捕捉情绪标签间关系上的优势。引入情绪标签排序后, 指标进一步增至 0.710、0.571 和 0.594, 突出了排序在识别情绪共现上的效果。完整的 MGE-BERT 模型在所有指标上都实现了最佳性能, *Micro-F1*、*Macro-F1* 和 *Jaccard* 指数分别达到 0.715、0.586 和 0.601。这些结果证实了 GCN 和情感标签排序在提高模型精确性和召回率上的重要性, 以及它们在 MGE-BERT 模型中的协同效应。

表 4 在 SemEval2018 Task-1C 数据集和 GoEmotions 数据集上的消融实验

模型名	SemEval2018 Task-1C			GoEmotions		
	<i>Micro-F1</i>	<i>Macro-F1</i>	<i>Jaccard</i>	<i>Micro-F1</i>	<i>Macro-F1</i>	<i>Jaccard</i>
BERT	0.672	0.551	0.546	0.679	0.590	0.568
BERT+GCN	0.692	0.580	0.564	0.682	0.582	0.604
BERT+Label	0.710	0.571	0.594	0.706	0.620	0.664
MGE-BERT	0.715	0.586	0.601	0.715	0.631	0.678

4.3 不同情感标签顺序对结果的影响

在 BERT 模型的应用实践中, 我们发现句子中相同单词因位置差异可能承载不同含义。鉴于情感标签能依据情感强度排序, 并且具有相同情感极性 (正面或负面) 的标签更可能在同一文本中共现, 而相反极性的标签则较少同时出现, 本文通过在 SemEval2018 Task-1C 和 GoEmotions 数据集上设置 3 组对照实验来验证情感标签排序对模型性能的影响, 如图 6 和图 7 展示。实验中, “Sort1”表示情感标签按情感强度顺序排列, 而 “Disorder1”与 “Disorder2”则是情感标签的随机排列。

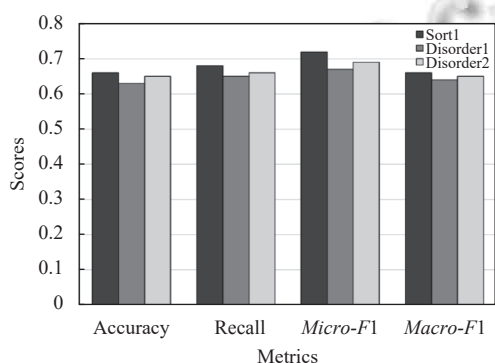


图 6 SemEval2018 Task-1C 数据集标签顺序对结果的影响

实验结果表明通过排序策略, 模型能够更精确地识别文本中的情感信息, 尤其是在处理包含多重情感的文本时更为有效。模型通过优先处理情感强度较高

的标签, 能更有效地识别和分类文本的主要情感, 从而提高情感分析的准确性。

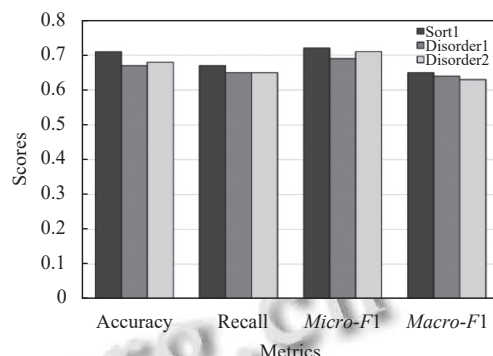


图 7 GoEmotions 数据集标签顺序对结果的影响

5 结论

本文提出了一种创新的深度学习模型 MGE-BERT, 专门用于多标签情感文本分类任务。该模型独特地关注情感标签之间的情感强度和情感层次关联性, 进行有序排列, 并使用 BERT 模型提取词级和句子级的序列文本特征。通过一种独特的图构建方法, 分别构建情感图和依存图, 并依据不同情感极性的标签分配不同的权重, 以揭示每个标签之间的潜在语义关系。随后, 将情感标签与文本信息相融合, 最后通过不同粒度的信息减少特征损失, 从而改进模型的分类性能。实验结果表明, 与现有模型相比, MGE-BERT 在 SemEval2018 Task-1C 数据集和 GoEmotions 数据集上均取得了更优异的性能, 显著提升了分类效果。

参考文献

- Ekman P, Oster H. Facial expressions of emotion. Annual Review of Psychology, 1979, 30(1): 527–554. [doi: 10.1146/annurev.ps.30.020179.002523]
- Plutchik R. A general psychoevolutionary theory of emotion. Emotion: Theory, research, and experience. 1980, 1: 3–33.
- Parrott WG. Emotions in social psychology: Essential readings. Psychology: Psychology Press, 2001.
- Price GF, Ogren M, Sandhofer CM. Sorting out emotions: How labels influence emotion categorization. Developmental Psychology, 2022, 58(9): 1665–1675. [doi: 10.1037/dev0001391]
- Elfaiik H, Nfaoui EH. Deep bidirectional LSTM network learning-based sentiment analysis for Arabic text. Journal of Intelligent Systems, 2020, 30(1): 395–412. [doi: 10.1515/jisys-2020-0021]

- 6 Kumar S, Kumar N, Dev A, *et al.* Movie genre classification using binary relevance, label powerset, and machine learning classifiers. *Multimedia Tools and Applications*, 2023, 82(1): 945–968. [doi: [10.1007/s11042-022-13211-5](https://doi.org/10.1007/s11042-022-13211-5)]
- 7 杨帆. 基于改进注意力机制的短文本情感分析研究 [硕士学位论文]. 武汉: 华中科技大学, 2019.
- 8 Zhou TT, Wang YT, Zheng X. Chinese text classification method using FastText and term frequency-inverse document frequency optimization. *Journal of Physics: Conference Series*, 2020, 1693(1): 012121. [doi: [10.1088/1742-6596/1693/1/012121](https://doi.org/10.1088/1742-6596/1693/1/012121)]
- 9 Baziotis C, Athanasiou N, Papalampidi P, *et al.* NTUA-SLP at SemEval2018 task 3: Tracking ironic tweets using ensembles of word and character level attentive RNNs. *Proceedings of the 12th International Workshop on Semantic Evaluation*. New Orleans: ACL, 2018. 613–621.
- 10 Ameer I, Bölücü N, Siddiqui MHF, *et al.* Multi-label emotion classification in texts using transfer learning. *Expert Systems with Applications*, 2023, 213: 118534. [doi: [10.1016/j.eswa.2022.118534](https://doi.org/10.1016/j.eswa.2022.118534)]
- 11 Pal A, Selvakumar M, Sankarasubbu M. Multi-label text classification using attention-based graph neural network. *arXiv:2003.11644*, 2020.
- 12 刘心惠. 基于改进 seq2seq 模型的多标签文本分类研究 [硕士学位论文]. 大连: 大连海事大学, 2020.
- 13 Baziotis C, Nikolaos A, Kolovou A, *et al.* NTUA-SLP at SemEval2018 task 2: Predicting emojis using RNNs with context-aware attention. *Proceedings of the 12th International Workshop on Semantic Evaluation*. New Orleans: ACL, 2018. 438–444.
- 14 Baziotis C, Pelekis N, Doukeridis C. DataStories at SemEval2017 task 4: Deep LSTM with attention for message-level and topic-based sentiment analysis. *Proceedings of the 11th International Workshop on Semantic Evaluation*. Vancouver: ACL, 2017. 747–754.
- 15 Demszky D, Movshovitz-Attias D, Ko J, *et al.* GoEmotions: A dataset of fine-grained emotions. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. ACL, 2020. 4040–4054.
- 16 Ahanin Z, Ismail MA, Singh NSS, *et al.* Hybrid feature extraction for multi-label emotion classification in English text messages. *Sustainability*, 2023, 15(16): 12539. [doi: [10.3390/su151612539](https://doi.org/10.3390/su151612539)]
- 17 Deng JW, Ren FJ. Hierarchical network with label embedding for contextual emotion recognition. *Research*, 2021, 2021: 3067943.
- 18 Fu RL, Wang H, Zhang XJ, *et al.* Decomposing complex questions makes multi-hop QA easier and more interpretable. *Proceedings of the 2021 Findings of the Association for Computational Linguistics*. Punta Cana: ACL, 2021. 169–180.
- 19 Vashishth S. Neural graph embedding methods for natural language processing. *arXiv:1911.03042*, 2019.
- 20 González-Carvajal S, Garrido-Merchán EC. Comparing BERT against traditional machine learning text classification. *arXiv:2005.13012*, 2020.
- 21 Alhuzali H, Ananiadou S. SpanEmo: Casting multi-label emotion classification as span-prediction. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*. Paola Merlo: ACL, 2021. 1573–1584.
- 22 Obrębski T, Stolarski M. UAM text tools-a flexible NLP architecture. *Proceedings of the 5th International Conference on Language Resources and Evaluation*. Genoa: ACL, 2006. 2259–2262.
- 23 Delobelle P, Winters T, Berendt B. RobBERT: A dutch RoBERTa-based language model. *Proceedings of the 2020 Findings of the Association for Computational Linguistics*. ACL, 2020. 3255–3265.
- 24 Ameer I, Bölücü N, Sidorov G, *et al.* Emotion classification in texts over graph neural networks: Semantic representation is better than syntactic. *IEEE Access*, 2023, 11: 56921–56934. [doi: [10.1109/ACCESS.2023.3281544](https://doi.org/10.1109/ACCESS.2023.3281544)]

(校对责编: 王欣欣)