第 39 卷 第 1 期
2024 年 2 月

天津科技大学学报
Journal of Tianjin University of Science & Technology

Vol. 39 No. 1
Feb. 2024

DOI:10.13364/j.issn.1672-6510.20230044

基于标签概念的多标签文本分类方法

汪乐乐，张贤坤
（天津科技大学人工智能学院，天津 300457）

摘　要：多标签文本分类是自然语言处理中重要且具有挑战性的任务之一。现有的方法注重文本表示学习，关注文本内部信息预测所属标签，忽略了属于某一标签的全体实例中共享的关键信息。鉴于此，本文提出一种基于标签概念的多标签文本分类方法：利用词频和潜在狄利克雷分布（latent Dirichlet allocation，LDA）方法从训练集全体实例中抽取各标签所对应的关键词，接着采取与文本编码相同方式对关键词编码，获得标签概念表示。在训练和预测过程中，检索与文本表示最相似的标签概念辅助分类，增加标签概念表示与文本表示的对比损失，使文本编码过程中能充分学习全局的标签概念信息。将本文方法嵌套在常用的多标签文本分类模型上进行实验，结果表明该方法有效提高了相应模型的性能。

关键词：标签概念；全局关键信息；对比损失；多标签文本分类
中图分类号：TP391.1　　**文献标志码**：A　　**文章编号**：1672-6510（2024）01-0073-08

Multi-Label Text Classification Method Based on Label Concept

WANG Lele，ZHANG Xiankun
（College of Artificial Intelligence，Tianjin University of Science & Technology，Tianjin 300457，China）

Abstract：Multi-label text classification is one of the important and challenging tasks in natural language processing. The existing methods pay attention to text representation learning，focus on the information inside the text to predict the label，but ignore the key information shared in all instances belonging to a certain label. In view of this，in this article we propose a multi-label text classification method based on the label concept. In our proposed method，word frequency and latent Dirichlet allocation（LDA）method are used to extract the key words corresponding to each tag from all the examples of the training set，and then the key words are encoded in the same way as text encoding to obtain the label concept representation. In the process of training and prediction，the auxiliary classification of tag concept that is most similar to text representation is retrieved，and the loss of comparison between tag concept representation and text representation is increased，so that the global tag concept information can be fully learned in the process of text coding. Experimental results showed that integrating our proposed method into commonly used multi-label text classification models significantly improved the performance of the respective models.

Key words：label concept；global key information；contrast loss；multi-label text classification

引文格式：
汪乐乐，张贤坤. 基于标签概念的多标签文本分类方法[J]. 天津科技大学学报，2024，39（1）：73-80.
WANG L L，ZHANG X K. Multi-label text classification method based on label concept[J]. Journal of Tianjin university of science & technology，2024，39（1）：73-80.

文本分类是自然语言处理领域中的典型任务之一。根据单个文本所对应标签数量的不同，可以将其

收稿日期：2023-03-01；**修回日期**：2023-06-04
基金项目：天津市科技计划项目（21ZYQCSY00050）
作者简介：汪乐乐（1994—），男，安徽阜南人，硕士研究生；**通信作者**：张贤坤，教授，zhxkun@tust.edu.cn

划分为单标签文本分类和多标签文本分类两种类型，其中多标签文本分类在实际生活中的应用更为广泛，例如主题分类[1]、情感分析[2]和标签推荐[3]等。

相较于单标签分类，多标签分类更准确地反映了文本的多样性和复杂性，但同时也面临着更多的挑战。首先是数据的稀疏性问题，每个文本实例可能涉及多个标签，这会导致标签组合的数量过于庞大，因此在训练数据中，很多标签组合出现较少甚至没有，这增加了模型训练的难度。其次是标签间可能存在依赖关系，也就是说某些标签的存在或缺失会对其他标签的预测产生影响，这种依赖关系会增加建模的复杂度。此外，每条文本都对应多个标签，即包含多个标签特征。因此，如何寻找与每个标签最相关且最具有辨别力的标签特定特征，利用这些特征提升模型的性能和效率具有重要意义。

在以往的多标签文本分类工作中，研究人员通常先利用各种方法学习文本表示，然后进行分类，其中不同的方法在文本处理阶段对于词权重的处理方式不同。一些前馈神经网络的方法，例如 DAN[4]和 FastText[5]，主要思想是将文本中的每个词映射为相应的向量，并通过计算这些向量的平均值获得文本表示；基于卷积神经网络（CNN）的方法利用不同大小的卷积核捕获文本的字、词或短语的局部特征；基于 Attention 机制的方法，通过向量点积或余弦相似性等计算方式衡量文本中词向量的相似性，并将其归一化获得权重，从而突出文本中不同词的重要性[6]，使模型能够在表示学习过程中关注重要信息。然而，上述方法都局限于某一文本实例内部，不能从全局判断文本中的关键信息。此外，使用领域专家知识或特定领域的词典、术语表等外部资源能够丰富文本的特征表示[7]，然而这种方法并不针对特定任务进行优化，因此可能无法充分表达特定任务所需的专业知识。

属于某一标签关键信息的聚合，相对于具体的文本实例可以更具体、更精准地表示标签信息。在具体推理中，如果某一文本实例与某一标签概念最相似，那么该文本实例大概率属于这一标签。利用这些更细粒度的全局关键信息具有重要意义。

为充分利用标签的全局概念信息，本文提出一种基于标签概念的多标签文本分类方法。根据词频和潜在狄利克雷分布（latent Dirichlet allocation，LDA）主题模型提取标签关键词，并使用与文本编码相同的方法对这些关键词进行编码，以获得标签概念。在学习和预测阶段使用 K 近邻（K-nearest neighbor，KNN）机制检索概念集，获取与当前文本表示最接近的 k 个标签概念作为预测结果，并融合原模型预测得到最终预测结果。为了评估 KNN 的预测结果，本文引入对比损失进行辅助训练。在对比学习的过程中，将文本所对应的标签概念视为正例，同时将其余标签概念视为负例，以拉近文本表示与对应正例之间的距离，并推算与负例之间的距离。

本文的主要贡献如下：（1）提出一种基于标签概念的多标签文本分类方法，从全局语料中获取标签概念，并使用对比损失优化的 KNN 辅助预测；（2）本文方法可作为通用框架与其他多标签文本分类模型结合，提高模型性能；（3）在 AAPD 和 RCV1-V2 两个数据集上进行实验，验证了本文方法的有效性。

1　相关工作

多标签文本分类任务方法主要分为两类：一类是基于机器学习的方法，另一类是基于深度学习的方法。

基于机器学习的方法侧重特征提取与分类器的设置，并可更细化地分为问题转换方法和算法自适应方法。问题转换方法的主要思路是将多标签分类问题转换为一个或者多个单标签分类问题[8]，如 Label Powerset[9]方法将标签的组合视为一个新的标签，将多标签分类问题转换为多分类问题。与问题转换类方法不同，算法自适应类方法的主要思想是改变原有算法使其能适应多标签分类需求。对于每一个新样本，ML-KNN[10]方法首先考虑距离该新样本最近的 k 个原样本的标签集合，再通过计算最大后验概率，预测该新样本的标签集合。基于机器学习的文本分类方法取得了一定成效，但因需要人工提取特征，忽略了文本数据中的顺序结构和上下文信息而陷入瓶颈。

近年来，神经网络和深度学习在多标签文本分类领域快速发展。由于 CNN 能够使用不同大小的卷积核对文本序列进行滑动窗口操作，从而有效捕获文本序列中词或短语的局部特征，因此很快被用于多标签文本分类任务。Kim[11]首先提出 TextCNN 模型，该模型用一层卷积捕获局部特征，并通过全局池化操作将这些特征整合为文本表示，最后使用全连接层进行分类。XML-CNN[12]设计动态的最大池化，相较于普通的最大池化，其补充了文档的不同区域更细粒度的特征，同时在池化层和输出层之间增加隐藏瓶颈层，用于学习紧凑的文本表示，既轻量化了模型，又提高了

分类性能。HFT-CNN[13]利用微调技术，将上层标签信息传递到下层，缓解了层级标签的数据稀疏性。相比基于 CNN 的方法，基于循环神经网络（RNN）的方法可以通过时间步的传递捕获长程相关性。此外，长短期记忆网络（LSTM）、门控循环单元（GRU）等 RNN 变体解决了梯度消失和梯度爆炸问题，因此得到了广泛应用。SHO-LSTM[14]方法使用了能解决非线性优化问题能力的斑点鬣狗优化器，通过优化 LSTM 的初始权重提升模型性能。注意力机制的引入，进一步提高了模型的分类性能。在文本分类任务中，注意力分数可以被理解为是一个重要的权重向量[15]。SGM[16]模型将多标签文本分类任务看作序列生成问题，考虑标签间的相互关联，并引入 Attention 自动获取输入文本的关键信息。Xiao 等[17]提出一种将两组基于历史信息的注意力机制应用于 seq2seq 模型的方法，其中一组注意力机制考虑历史上下文词汇以增强预测能力，另一组注意力机制考虑历史标签信息以缓解错误传播问题。Transformer 是一种基于多头注意力设计的模型，它克服了 RNN 不能并行计算、CNN 捕获全文的上下文信息效率低下的缺点。HG-Transformer[18]模型将文本建模为图，并将多层 Transfomer 结构用在单词、句和图级别捕获局部特征，最后依据标签的层次关系生成标签的表示形式。LightXML[19]是一种轻量级的深度学习模型，采用端到端训练和动态负标签采样。该方法使用生成合作网络对标签进行召回和排序，其中标签召回部分生成负标签和正标签，标签排序部分将正标签与这些标签区分开来。在标签排序的训练阶段中，通过将相同的文本表示作为输入，动态采样负标签，以提升模型的分类性能。MAGNET[20]模型利用图注意力网络学习标签之间的依赖关系，利用特征矩阵和相关系数矩阵生成分类器。该模型通过在图中传播信息和捕捉关系，有效解决了语义稀疏性问题，提升了分类性能。LiGCN[21]方法提出了一个可解释标签的图卷积网络模型，将词元和标签建模为异构图中的节点，解决多标签文本分类问题，通过这种方式，能够考虑包括词元级别关系在内的多个关系。然而，基于图神经网络的方法存在过平滑的问题，无法捕获深度依赖特征。

2　基于标签概念的多标签文本分类方法

本文提出一种基于标签概念的多标签文本分类方法，模型框架如图 1 所示。本文方法主要包括 3 个部分：标签概念获取、对比学习和综合预测。标签概念获取部分将训练集按照标签划分，依据词频和 LDA 主题模型抽取标签的全局关键词，随后采用与编码文本相同的方式对标签关键词进行编码，以获得标签概念的向量表示；对比学习阶段引入对比损失，使文本表示与对应标签概念之间的距离尽可能小，与其他标签概念之间的距离尽可能大；综合预测阶段将 KNN 预测结果与基础模型预测结果的加权和作为最终预测结果。

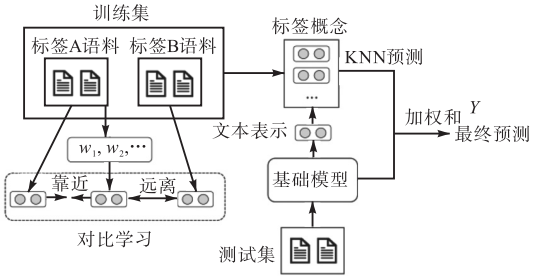


图 1　模型框架

Fig. 1　Model framework

2.1　问题陈述

数据集 $D=\{(x_i,y_i)\}_{i=1}^{N}$，$N$ 为文本总数，x_i 为其中一条文本，对应标签 $y_i\in\{0,1\}^L$，L 为标签总数。每条文本由一组预训练词向量序列构成，即 $x_i=\{w_1,w_2,\ldots,w_n\}$，$w_i\in R^k$，$k$ 为词嵌入维度。多标签文本分类任务旨在训练一个预测模型，使其能对全新的未分类文本进行准确分类。

2.2　标签概念获取

为获得具有全局关键信息的标签概念，本文提出了一种方法，首先基于词频和 LDA 主题模型抽取标签关键词，然后将这些关键词编码成标签概念。在抽取标签关键词前，随机抽取属于某一标签的文本语料，构建该标签的全局语料库。综合考虑本实验使用的数据集和实验效率，标签语料库的数量上限设定为 10 000。

从概率角度分析，显然标签语料中词频越高的词与该标签越相关，故对属于标签 y_i 的全体文本进行词频统计，取前 k 个词构成标签关键词集 Key_{pi}。

$$Key_{pi}=\{key_{pi,1},key_{pi,2},\cdots,key_{pi,k}\} \tag{1}$$

重复是提高词影响力的一种方式，故利用 LDA 主题模型进一步区分关键词。将 y_i 标签的全局语料主题数设置为 2，原因如下：每条文本具有 2 个及以上标签，若将标签看作主题，则该语料库所包含主题

必有 y_i 标签主题，剩余标签主题可统一视为另一主题，且这两个主题各包含 $k/2$ 个主题关键词。极端情况下，词频关键词包含 y_i 标签主题的关键词，且与剩余主题的关键词完全不相干，这样更关键的词就被筛选出来。采用 LDA 主题模型提取影响力前 k 个主题词得到标签关键词集 Key_{li}。

$$Key_{li}=\{key_{li,1},key_{li,2},\cdots,key_{li,k}\} \tag{2}$$

在频数关键词集 Key_{pi} 的基础上添加两子集的交集，得到标签 y_i 的最终关键词句，即

$$S_i=\{Key_{pi}+(Key_{li}\cap Key_{pi})\} \tag{3}$$

采取与文本相同的编码方式得到标签 y_i 的概念表示，即 $C_i=Encoder(S_i)$。

2.3　综合预测

在预测阶段，首先将文本 x_i 输入到基础模型，得到基础模型预测结果 y_b 和期间的文本表示 $Encoder(x_i)$。其次，度量该文本表示与所有标签概念之间的距离，取前 p 个最近的标签概念所对应的标签辅助预测。具体 KNN 预测过程为

$$\alpha_j=\frac{e^{-(d(Encoder(x_i),C_j))}}{\sum_{j=1}^{p}e^{-(d(Encoder(x_i),C_j))}} \tag{4}$$

$$y_{\mathrm{KNN}}=\sum_{j=1}^{p}\alpha_j*y_j \tag{5}$$

为防止梯度爆炸或者梯度消失，距离 $d(a,b)$ 为经 Z-score 标准化后的 a 与 b 间的欧氏距离。α_j 表示第 j 个标签的权重，与文本表示距离越近的标签概念权重越大。最后，将 KNN 预测结果与基础模型预测结果 y_b 作加权和，其中 KNN 预测权重为 λ_{KNN}，得到最终预测结果，即

$$y=\lambda_{\mathrm{KNN}}*y_{\mathrm{KNN}}+(1-\lambda_{\mathrm{KNN}})*y_b \tag{6}$$

2.4　损失函数

对于多标签文本分类，通常使用二元交叉熵损失作为损失函数，对于一批含有 M 个文本的数据来说，其损失 L_{BCE} 为

$$L_{\mathrm{BCE}}=-\frac{1}{M}\sum_{i=1}^{M}[y_i\log(\hat{y}_i)+(1-y_i)\log(1-\hat{y}_i)] \tag{7}$$

其中：y_i 为第 i 个文本的真实标签，$\hat{y}_i$ 为第 i 个文本的预测标签值。

此外，为了评估 KNN 预测的质量、推动文本编码时考虑全局的标签概念特征，引入标签概念和文本表示的对比损失[22]辅助训练模型。对于文本 x_i，计算文本表示与所有标签概念表示之间的距离，对比损失为

$$L_C=-\log\left(\frac{e^{-d(Encoder(x_i),C_i)}}{\sum_{j\in C,j\neq C_i}e^{-d(Encoder(x_i),C_j)}}\right) \tag{8}$$

其中：$Encoder(x_i)$ 为文本 x_i 的编码表示，C_i 为其真实标签概念。对比损失 L_C 与文本表示和真实标签概念表示之间的距离 $d(Encoder(x_i),C_i)$ 成正比，因此在模型训练的过程中，文本编码会考虑真实标签概念的影响，使他们之间的距离变得更近；对比损失 L_C 与其他标签概念间的距离成反比，则会被优化得更远。综合考虑两种损失，模型总体损失函数为

$$L=L_{\mathrm{BCE}}+\bar{L}_C \tag{9}$$

其中 $\bar{L}_C$ 为 M 个文本的平均对比损失。

3　实　验

3.1　实验数据

本文在两个广泛使用的数据集上对所提出的方法进行评估，数据集相关信息见表 1。

表 1　数据集相关信息

Tab. 1　Dataset related information

数据集	样本总数	标签数	平均标签数	样本平均长度
AAPD	55 840	54	2.41	163.42
RCV1-V2	804 414	103	3.24	123.94

AAPD[16]是 arXiv 上计算机科学领域的论文摘要数据集，共包含 55 840 篇摘要、54 个主题，其中每篇摘要对应多个领域主题。

RCV1-V2[23]是一个大型数据集，包含超过 80 万条路透社提供的新闻报道和 103 个主题，其中每篇报道对应多个新闻主题。

3.2　实验指标

实验选取常用多标签分类评价指标 Micro-F1、Macro-F1 和 Micro-Recall 评估实验性能。Micro-Recall 反映了预测为某一类的样本中，预测正确的比例，Micro-F1 综合考虑了预测的总体精确率和召回率，Macro-F1 则是 Micro-F1 全体的平均。此外，Micro-F1 指标倾向样本，任意样本权重相同；Macro-F1 指标倾向类别，任意类别权重相同。

3.3　基础模型

为了验证方法的有效性，选取以下模型作为基础模型：

FastText[5]：用词向量的平均作为文本表示再进行分类。

TextRNN[24]：使用循环神经网络捕获文本的长序

列特征信息的分类模型，神经元采用 GRU。

TextCNN[11]：使用不同大小的最大池化卷积核捕获文本特征的分类模型，侧重局部特征。

SGM[16]：将多标签文本分类任务看作生成输入文本的标签序列任务。

3.4　实验参数

词嵌入使用 Glove[25]预训练的 300 维词向量，并随机初始化词表以外的单词。KNN 预测标签数 p 在 AAPD 数据集上设定为 2，在 RCV1-V2 数据集上设为 3，抽取标签关键词数为 10，KNN 预测权重设定为 0.5。模型使用 Adam 优化器；模型批次大小为 64；初始学习率设置为 0.000 1，连续 5 个 epoch 实验性能未提升衰减学习率，衰减率为 0.1，连续 10 个 epoch 性能未提升则停止训练。模型均使用 PyTorch 框架实现，其余参数与原模型一致。

3.5　实验及结果分析

详细实验结果见表 2。结果表明：该方法应用在 4 个基础模型上均能提升模型性能，其中在 TextCNN 上效果最好，在 AAPD 数据集上 Micro-Recall、Micro-F1 和 Macro-F1 指标分别提升 3.67%、1.66%、4.07%，在 RCV1-V2 数据集上则分别提升 1.86%、1.02%、2.79%。应用在 TextCNN 上效果最好的原因可能是 KNN 预测部分补充了 TextCNN 最大池化过程中丢失的信息。

表 2　该方法应用在两个数据集上的实验性能对比

Tab. 2　Experimental performance comparison of the method applied to two data sets

模型	AAPD 数据集			RCV1-V2 数据集		
	Micro-Recall	Micro-F1	Macro-F1	Micro-Recall	Micro-F1	Macro-F1
FastText	0.652 036	0.710 322	0.518 887	0.849 199	0.867 742	0.685 727
FastText + 本文方法	0.676 167	0.718 928	0.537 805	0.856 638	0.873 804	0.693 392
TextCNN	0.585 047	0.665 273	0.443 848	0.798 689	0.838 817	0.622 202
TextCNN + 本文方法	0.621 809	0.681 948	0.484 554	0.817 387	0.849 035	0.650 176
TextRNN	0.635 688	0.701 618	0.504 978	0.830 500	0.852 855	0.647 634
TextRNN + 本文方法	0.638 579	0.705 614	0.520 817	0.835 311	0.859 016	0.656 273
SGM	0.647 738	0.698 237	0.502 841	0.846 261	0.864 285	0.642 263
SGM + 本文方法	0.651 169	0.704 285	0.518 676	0.853 935	0.871 126	0.651 837

此外，在调整好局部权重的前提下，本文补充了部分实验，进一步验证本文方法仍能通过学习全体实例中的关键信息提升模型性能。在 TextCNN、TextRNN 模型上引入 Attention 机制作为基础模型，再结合本文方法进行实验对比，实验结果见表 3。

通过对表 3 中实验结果与表 2 中带有 Attention 机制的 SGM 相关实验结果进行综合分析，可以发现，各基础模型添加本文方法后，实验性能均得到一定程度的提升，这表明本文方法能在文本编码过程中学习标签的全局关键信息，并与 Attention 机制具有互补关系。

表 3　引入 Attention 机制的 TextCNN、TextRNN 实验性能对比

Tab. 3　Experimental performance comparison of TextCNN and TextRNN with the introduced attention mechanism

模型	AAPD 数据集			RCV1-V2 数据集		
	Micro-Recall	Micro-F1	Macro-F1	Micro-Recall	Micro-F1	Macro-F1
TextCNN + ATT	0.641 305	0.697 905	0.503 779	0.829 286	0.847 152	0.637 600
TextCNN + ATT + 本文方法	0.653 186	0.704 368	0.518 936	0.833 171	0.850 204	0.653 375
TextRNN + ATT	0.655 101	0.707 781	0.513 325	0.836 414	0.858 426	0.650 985
TextRNN + ATT + 本文方法	0.659 117	0.711 964	0.526 396	0.840 500	0.861 214	0.657 748

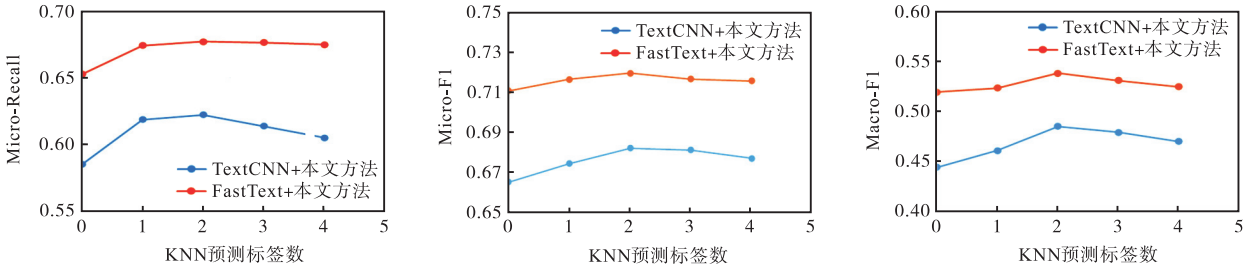
3.6　关键参数分析

本文方法中有 3 个关键参数对实验结果产生重要影响，分别是 KNN 预测的标签数、获取标签概念时的候选标签关键词数以及 KNN 预测权重。因此，结合 TextCNN 和 FastText 两个基础模型，在 AAPD 数据集上进行参数调节，分析参数对实验结果的影响。每组实验固定两个参数，对剩余参数进行调节。在默认设置下，KNN 预测标签数为 2，标签关键词数为 10，KNN 预测权重为 0.5。

3.6.1　KNN 预测标签数

KNN 预测标签数 p 对实验的影响如图 2 所示。实验结果表明，模型在 p 为 2 时性能最好。起始点 KNN 预测标签数量为 0，即为基础模型实验结果，性能较起始点的提升证明了该方法的有效性。分析数据集，AAPD 数据集的平均标签数为 2.4，其中标签数为 2 的数据占比为 69.4%，因此推理 KNN 预测标

签数应小于样本最大标签数，设为该数据集文本对应标签数的众数比较合理。此外，p 设为 3、4 时，各指标性能持续下降，说明预测多余标签反而会对结果造成干扰。



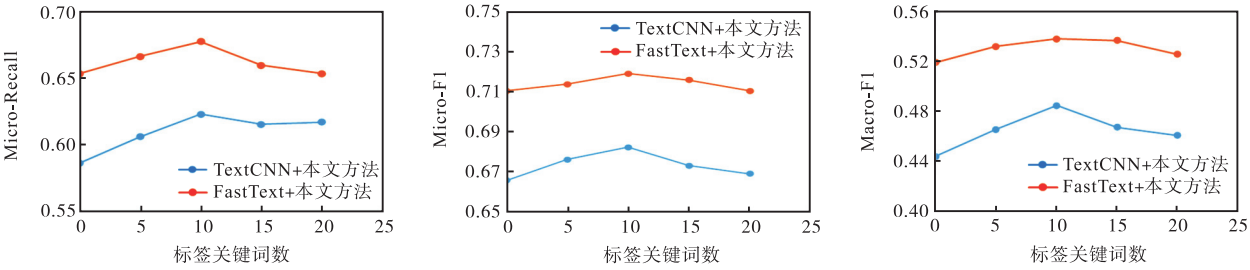
(a) Micro-Recall　(b) Micro-F1　(c) Macro-F1

图 2　KNN 预测标签数对实验的影响

Fig. 2　Effect of the number of labels predicted by KNN on the experiment

3.6.2　标签关键词数

标签关键词数 k 对实验结果的影响如图 3 所示。当标签关键词数量为 10 时，模型表现最佳。这说明即使面向全局，标签关键信息也是有限的，相对少量的关键词所编码的标签概念更具区分度。然而，随着标签关键词的增加，对应标签概念的普适性降低，进而导致实验性能下降。



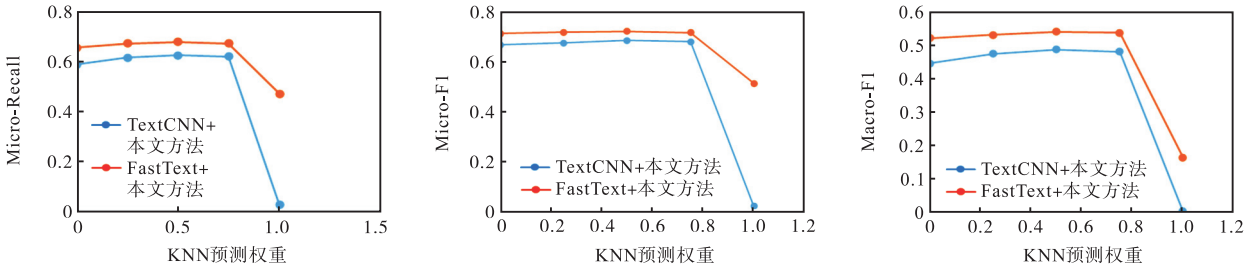
(a) Micro-Recall　(b) Micro-F1　(c) Macro-F1

图 3　标签关键词数对实验的影响

Fig. 3　Effect of the number of label key words on the experiment

3.6.3　KNN 预测权重

KNN 预测权重 λ_{KNN} 对实验结果的影响如图 4 所示。模型性能先增长再降低，性能先增长说明包含全局关键信息的 KNN 预测结果能弥补文本局部内权重分配不合理的缺陷。当 λ_{KNN} 取 1 时，即完全采取 KNN 预测时，本文方法结合 FastText 模型性能大幅度降低，结合 TextCNN 模型的性能几乎为零。这说明每条文本所含有的关键信息是稀疏的，该方法不具备独立预测能力。



(a) Micro-Recall　(b) Micro-F1　(c) Macro-F1

图 4　KNN 预测权重对实验的影响

Fig. 4　Effect of KNN prediction weight on the experiment

4　结　语

本文提出了一种基于标签概念的多标签文本分类方法。首先，为获取某一标签共享在全体实例中的关键信息并显式表达，先利用词频和 LDA 主题模型提取标签关键词，再将其编码为标签概念。其次，本文方法引入了对比损失，减小文本表示与所对应标签

概念之间的距离，从而在文本编码过程中能够充分学习文本所对应标签的全局关键信息。同时，本文方法具有良好的可移植性，可以嵌入现有的多标签文本分类模型中。在 AAPD 和 RCV1-V2 两个数据集上的实验结果表明，本文方法能有效提升基础模型的性能。此外，本文还对几个关键参数设定的原因和影响进行了讨论。然而，本文方法仍存在一些不足之处。由于综合预测和对比学习阶段需将每一条文本与所有的标签概念进行对比，对于具有大型标签集的数据而言，这两个阶段的时间开销较大，因此不适用于极限多标签文本分类任务。在后续的研究中，这将成为重点关注和讨论的内容。

参考文献：

[1] HUAN J L，SEKH A A，QUEK C，et al. Emotionally charged text classification with deep learning and sentiment semantic[J]. Neural computing and applications，2022，34：2341–2351.

[2] 王颖洁，朱久祺，汪祖民，等. 自然语言处理在文本情感分析领域应用综述[J]. 计算机应用，2022，42(4)：1011–1020.

[3] 徐鹏宇，刘华锋，刘冰，等. 标签推荐方法研究综述[J]. 软件学报，2022，33(4)：1244–1266.

[4] IYYER M，MANJUNATHA V，BOYD-GRABER J，et al. Deep unordered composition rivals syntactic methods for text classification[C]//ACL. Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing. Kerrville：ACL，2015：1681–1691.

[5] JOULIN A，GRAVE E，BOJANOWSKI P，et al. FastText zip：compressing text classification models[EB/OL]. [2023–01–01]. https://doi.org/10.48550/arXiv.1612.03651.

[6] DENG J，CHENG L，WANG Z. Attention-based BiLSTM fused CNN with gating mechanism model for Chinese long text classification[J]. Computer speech & language，2021，68：101182.

[7] MINAEE S，KALCHBRENNER N，CAMBRIA E，et al. Deep learning-based text classification：a comprehensive review[J]. ACM Computing surveys，2021，54(3)：1–40.

[8] 郝超，裘杭萍，孙毅，等. 多标签文本分类研究进展[J]. 计算机工程与应用，2021，57(10)：48–56.

[9] TSOUMAKAS G，KATAKIS I. Multi-label classification：an overview[J]. International journal of data warehousing and mining，2007，3(3)：1–13.

[10] ZHANG M L，ZHOU Z H. ML-KNN：a lazy learning approach to multi-label learning[J]. Pattern recognition，2007，40(7)：2038–2048.

[11] KIM Y. Convolutional neural networks for sentence classification[EB/OL]. [2023–01–01]. https://doi.org/10.48550/arXiv.1408.5882.

[12] LIU J，CHANG W C，WU Y，et al. Deep learning for extreme multi-label text classification[C]//ACM. Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. New York：ACM，2017：115–124.

[13] SHIMURA K，LI J，FUKUMOTO F. HFT-CNN：Learning hierarchical category structure for multi-label short text categorization[C]//ACL. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Kerrville：ACL，2018：811–816.

[14] MARAGHEH H K，GHAREHCHOPOGH F S，MAJIDZADEH K，et al. A new hybrid based on long short-term memory network with spotted hyena optimization algorithm for multi-label text classification[J]. Mathematics，2022，10(3)：488.

[15] LIU Y，LI P，HU X. Combining context-relevant features with multi-stage attention network for short text classification[J]. Computer speech & language，2022，71：101268.

[16] YANG P，SUN X，LI W，et al. SGM：sequence generation model for multi-label classification[EB/OL]. [2023–01–01]. http://arxiv.org/pdf/1806.04822.

[17] XIAO Y，LI Y，YUAN J，et al. History-based attention in Seq2Seq model for multi-label text classification[J]. Knowledge-based systems，2021，224：107094.

[18] GONG J，TENG Z，TENG Q，et al. Hierarchical graph transformer-based deep learning model for large-scale multi-label text classification[J]. IEEE Access，2020，8：30885–30896.

[19] JIANG T，WANG D，SUN L，et al. LightXML：transformer with dynamic negative sampling for high-performance extreme multi-label text classification[C]//AAAI. Proceedings of the AAAI Conference on Artificial Intelligence. California：AAAI，2021，35(9)：7987–7994.

[20] PAL A，SELVAKUMAR M，SANKARASUBBU M. Multi-label text classification using attention-based graph neural network[EB/OL]. [2023–01–01]. https://doi.org/10.48550/arXiv.2003.11644.

[21] LI I，FENG A，WU H，et al. LiGCN：Label-interpretable

graph convolutional networks for multi-label text classification[EB/OL]. [2023-01-01]. https://doi.org/10.48550/arXiv.2103.14620.

[22] HE K, FAN H, WU Y, et al. Momentum contrast for unsupervised visual representation learning[C]//IEEE. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE, 2020: 9729-9738.

[23] LEWIS D D, YANG Y, RUSSELL-ROSE T, et al. RCV1: a new benchmark collection for text categorization research[J]. Journal of machine learning research, 2004, 5: 361-397.

[24] LIU P, QIU X, HUANG X. Recurrent neural network for text classification with multi-task learning[EB/OL]. [2023-01-01]. https://doi.org/10.48550/arXiv.1605.05101.

[25] PENNINGTON J, SOCHER R, MANNING C D. Glove: global vectors for word representation[C]//ACL. Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Kerrville: ACL, 2014: 1532-1543.

责任编辑：郎婧

（上接第 21 页）

[27] 王艳红, 田少君, 张争全, 等. 大豆分离蛋白-黄原胶-茶多酚复合物的制备及其乳液性质表征[J]. 中国油脂, 2021, 46(4): 38-42.

[28] JIANG Y, LIU L, WANG B, et al. Cellulose-rich oleogels prepared with an emulsion-templated approach[J]. Food hydrocolloids, 2018, 77: 460–464.

[29] PATEL A R, CLUDTA N, SINTANG M D B, et al. Edible oleogels based on water soluble food polymers: preparation, characterization and potential application [J]. Food & function, 2014, 5: 2833–2841.

[30] LI J, XI Y, WU L, et al. Preparation, characterization and in vitro digestion of bamboo shoot protein/soybean protein isolate based-oleogels by emulsion-templated approach[J]. Food hydrocolloids, 2023, 136: 108310.

[31] MENG Z, QI K, GUO Y, et al. Macro-micro structure characterization and molecular properties of emulsion-templated polysaccharide oleogels[J]. Food hydrocolloids, 2018, 77: 17–29.

[32] LAREDO T, BARBUT S, MARANGONIA G. Molecular interactions of polymer oleogelation[J]. Soft matter, 2011, 7: 2734.

[33] WAND S J, ZHENG M G, YU J L, et al. Insights into the formation and structures of starch-protein-lipid complexes[J]. Journal of agricultural and food chemistry, 2017, 65: 1960–1966.

[34] LUPI F R, GRECO V, BALDINO N, et al. The effects of intermolecular interactions on the physical properties of organogels in edible oils[J]. Journal of colloid and interface science, 2016, 483: 154–164.

责任编辑：郎婧