

# 基于融合注意力机制的多标签文本分类方法研究

龚静, 刘现芳, 范江豪

湖南环境生物职业技术学院公共基础课部, 湖南 衡阳

收稿日期: 2024年11月1日; 录用日期: 2024年11月29日; 发布日期: 2024年12月9日

## 摘要

针对传统多标签文本分类方法忽略标签相关性及文本特征的深层语义信息, 导致分类精度和泛化能力不足的问题, 提出一种基于融合注意力机制的多标签文本分类方法。首先, 构建深度学习模型, 通过算法设计与融合注意力机制加强了模型对文本关键信息的提取能力。在模型优化方面, 结合标签重要性加权和多任务学习, 通过损失函数设计进一步提升了模型性能。为验证所提方法的有效性, 在Reuters-21578和RCV1-v2两个数据集上进行实验, 将所提模型与现有模型进行比较。实验结果表明, 基于融合注意力机制的多标签文本分类方法相较于传统多标签文本分类方法在准确率、召回率、F1值等指标上有显著提升, 能够更好地处理文本中的复杂语义和标签之间的依赖关系, 展示了更高的分类精度和泛化能力, 为多标签文本分类研究提供了新的视角和解决方案。

## 关键词

融合注意力机制, 多标签文本分类, 文本表示, 自然语言处理

# Research on Multi Label Text Classification Method Based on Fusion Attention Mechanism

Jing Gong, Xianfang Liu, Jianghao Fan

Department of Public Basic Courses, Hunan Polytechnic of Environment and Biology, Hengyang Hunan

Received: Nov. 1<sup>st</sup>, 2024; accepted: Nov. 29<sup>th</sup>, 2024; published: Dec. 9<sup>th</sup>, 2024

## Abstract

A multi label text classification method based on fusion attention mechanism is proposed to address

文章引用: 龚静, 刘现芳, 范江豪. 基于融合注意力机制的多标签文本分类方法研究[J]. 计算机科学与应用, 2024, 14(12): 90-96. DOI: 10.12677/csa.2024.1412243

the problem of insufficient classification accuracy and generalization ability caused by traditional multi label text classification methods ignoring the deep semantic information of label relevance and text features. Firstly, a deep learning model was constructed to enhance its ability to extract key information from text through algorithm design and the integration of attention mechanisms. In terms of model optimization, the combination of label importance weighting and multi task learning has further improved the model performance through loss function design. To verify the effectiveness of the proposed method, experiments were conducted on two datasets, Reuters-21578 and RCV1-v2, to compare the proposed model with existing models. The experimental results show that the multi label text classification method based on fusion attention mechanism has significant improvements in accuracy, recall, F1 value and other indicators compared to traditional multi label text classification methods. It can better handle the complex semantics and dependency relationships between labels in the text, demonstrate higher classification accuracy and generalization ability, and provide new perspectives and solutions for multi label text classification research.

## Keywords

Fusion Attention Mechanism, Multi Label Text Classification, Text Representation, Natural Language Processing

Copyright © 2024 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

在信息爆炸时代, 文本数据呈指数级增长, 如何对海量文本进行分类成了一个亟待解决的关键问题。多标签文本分类作为一项自然语言处理技术, 其应用场景广泛, 涵盖新闻分类、情感分析、商品标注等领域, 但其也面临诸多挑战, 例如标签复杂关系、文本多样性和语义模糊性等[1]。目前学界针对多标签文本分类展开了深入研究, 王嫒(2024)等提出基于编码器-解码器结构的标签语义交互 Seq2Seq 模型, 得到类别理解纠正的文本表示[2]; 张红(2024)引入 BERT 模型搭建自适应分类框架, 实现了多标签文本的自适应分类[3]。这些方法虽取得部分突破, 但未能有效捕捉复杂语义关系和标签相关性, 导致分类准确率不高, 无法满足精准分类需求。随着深度学习技术的发展, 融合注意力机制逐渐被引入多标签文本分类中, 其能够聚焦文本关键信息, 从而提高分类的准确性和效率, 因此为多标签文本分类提供了新的思路[4]。鉴于此, 提出一种基于融合注意力机制的多标签文本分类方法, 通过对模型的进一步优化, 能够克服传统方法的局限性, 为实际场景中的多标签文本分类提供更可靠的解决方案。

## 2. 理论基础

### 2.1. 多标签文本分类

在单标签分类中, 每个文本仅被赋予一个类别, 与其相比, 多标签文本分类中的每个文本可能同时属于多个类别, 其特点在于标签间可能存在相关性, 这样便增加了分类复杂性。例如, 关于智能产品的文章可能同时涉及“智能设备”“人工智能”“大数据技术”等多个标签, 难点在于准确捕捉文本与多个标签之间的复杂关系, 以及处理标签的失衡和不确定性。多标签文本分类在信息检索领域有着广泛应用, 例如搜索引擎可根据多标签分类为用户提供更精准的搜索结果[5]; 在情感分析中, 不仅要判断文本的整体情感倾向, 还要识别出喜悦、愤怒、悲伤等所有涉及的情感元素。此外, 在内容推荐系统中, 通过对用户兴趣的多标签分类, 可为用户推荐更符合其多样化需求的内容。

## 2.2. 融合注意力机制

融合注意力机制能够通过动态调整文本不同部分的关注权重，从而更精准地捕捉到与不同标签相关的关键信息。例如，在处理艺术文献时，对于与特定技巧标签相关的段落给予更高的权重，从而提高对该标签的分类准确性。在面对多维数据时，融合注意力机制可以有效地整合和筛选来自不同数据源和维度的信息[6]。例如，同时处理文本内容、作者信息、发布时间等多维度数据，能够根据任务需求灵活地分配注意力权重，从而更好地适应复杂的数据特征和结构。

## 3. 模型构建

### 3.1. 词向量表示

词向量使用 word2vec 方法表示，并通过 skip-gram 算法进行实现，在语料库上构建词共现矩阵进行词向量训练，可捕捉单词之间的语义特性，从而得到词向量表示。在词向量表示时，每个样本仅被赋予一个唯一标签，给定样本集  $S = \{x_1, x_2, \dots, x_n\}$ ，对于样本标签  $x_i$ ，其文本语义表示为  $x_j \in C$ ，其中  $C$  是一个无限的语义集合。

### 3.2. 标签注意力

常见的注意力机制包括软注意力机制与硬注意力机制，软注意力机制通过计算注意力分布来对输入信息进行加权平均，硬注意力机制直接选择某个特定的输入向量，实现方式通常基于最大采样或随机采样。为更好结合本方法考虑标签的相关性，将处理后的文本表示与标签表示相结合进行交互注意力计算。首先，针对序列中的文本标签  $x_i$  与文本语义表示  $x_j$ ，通过不同的线性层得到 Query  $Q_i$ 、Key  $K_j$  和 Value  $V_j$ ：

$$\begin{cases} Q_i = W^Q x_i \\ K_j = W^K x_j \\ V_j = W^V x_j \end{cases} \quad (1)$$

其中， $W^Q$ ， $W^K$ ， $W^V$  是不同的权重矩阵，随后计算 Query  $Q_i$  和所有 Key  $K_j$  之间的相似度，此时使用点积来实现：

$$\text{score}(i, j) = Q_i \cdot K_j^T \quad (2)$$

其中， $\cdot$  表示点积， $K_j^T$  表示 Key 的转置。接着，使用 softmax 函数对相似度得分进行归一化，得到注意力权重：

$$a_{ij} = \frac{\exp(\text{score}(i, j))}{\sum_k \exp(\text{score}(i, k))} \quad (3)$$

最后，使用注意力权重对 Value 进行加权求和，得到基于文本内容的标签表示  $\text{Output}_i$ ：

$$\text{Output}_i = \sum_j a_{ij} \cdot V_j \quad (4)$$

在模型构建中，随着注意力的进一步扩展，这样  $\text{Output}_i$  就能较为全面地结合文本内容考虑标签与文本的关系及标签间的相关性，由此便于模型运行。

### 3.3. 融合策略

融合注意力机制能够综合不同注意力机制的优点，提高对复杂文本信息的捕捉能力，适应不同类型和特点的文本数据。本文采用并行融合策略，当获得文本语义表示  $x_j$  与基于标签的文本表示  $\text{Output}_i$  后，

使用并行融合策略将二者结合，通过加权求和得到最终的文本表示，这样可进一步提取有效语义信息，从而提升模型泛化能力。

使用不同权重矩阵  $\omega_1$  与  $\omega_2$  确定提取上述两种文本表示，权重矩阵由 Tanh 函数获得：

$$\begin{cases} \omega_1 = \text{Tanh}(x_j \cdot \beta_1) \\ \omega_2 = \text{Tanh}(\text{Output}_i \cdot \beta_2) \end{cases} \quad (5)$$

其中， $\beta_1$ 、 $\beta_2$  分别表示可训练参数矩阵，可将最终预测的第  $n$  个标签的文本表示为：

$$F_n = \omega_1 + \omega_2 \quad (6)$$

## 4. 模型优化

为优化模型性能，将多标签分类的层次结构扁平化处理，并把融合后的文本语义信息进行转换，从而得到预测结果：

$$P = \text{Tanh}(\beta_3 \cdot \delta(\beta_4 F_n)) \quad (7)$$

其中， $\beta_3$ 、 $\beta_4$  分别表示可训练参数矩阵， $\delta$  为激活函数。随后，为进一步衡量模型预测结果与真实标签之间的差距，并指导学习过程，需选择多分类交叉熵损失函数来优化模型性能，可表示为

$$L_{\text{MCE}} = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^C y_{ij} \log(\hat{y}_{ij}) \quad (8)$$

其中  $C$  表示类别集合， $y_{ij}$  表示样本  $i$  属于类别  $j$  的真实标签， $\hat{y}_{ij}$  是样本  $i$  属于类别  $j$  的预测概率。在应用中，可根据数据集特点和模型结构对损失函数进行调整。如果数据集存在类别失衡问题，可为不同样本赋予不同权重，以减少失衡对模型训练的影响。此外，还可引入正则化项  $L1$  防止模型过拟合：

$$R_{L1}(\theta) = \lambda \sum_{j=1}^M |\theta_j| \quad (9)$$

其中  $\theta$  是模型参数， $M$  是参数数量， $\lambda$  是正则化参数。将损失函数与正则化项结合，可得到最终的优化目标函数：

$$J = L + \lambda R \quad (10)$$

其中  $L$  是损失函数， $R$  是正则化项。

## 5. 实验设置

### 5.1. 参数设置

模型训练需设置实验参数，如学习率、迭代次数、批大小等。学习率根据实验调整，学习率较小会导致训练收敛缓慢，而学习率较大会导致模型无法收敛；迭代次数取决于模型性能，当损失不再下降可认为已收敛，此时停止训练；批大小会影响模型的训练效率和稳定性，批大小较小会导致梯度估计方差过大，批大小较大则会占用过多内存。在具体设置方面，学习率设置为 0.001，迭代次数为 500，批大小为 64。使用 word2vec 将文本转换为 300 维的词向量。

### 5.2. 数据集

为有效训练模型，选择多标签文本分类数据集 Reuters-21578 和 RCV1-v2 进行实验。Reuters-21578 数据集主题广泛，包括 721,738 篇人工分类的新闻，共有 109 个标签；RCV1-v2 数据集则具有丰富的文本特征和多样的标签类别，包括 56,320 篇摘要与对应的 62 个标签。在预处理数据集时，首先进行数据清

洗，去除噪声和无效数据。随后使用分词工具对文本进行分词和词干提取，并通过 GloVe 嵌入技术将文本转换为词向量表示。此外，为应对类别失衡问题，采用过采样来平衡各类别的样本数量，以便模型处理。

### 5.3. 模型比较

为验证本方法相比传统模型在性能上的优势及其对多标签文本分类预测性能的提升，将其与已有模型进行比较实验。参与比较的模型分为两种：一种是基于决策树的多标签分类模型与基于支持向量机的多标签分类模型；第二种则是其他融合注意力机制的简单叠加模型。

### 5.4. 评价指标

为评估模型性能，本实验选择准确率(Precision)、召回率(Recall)和 F1 值(F1-score)作为衡量多标签文本分类模型性能的评估指标，其中 TP 表示模型正确预测为正类的样本数量，FP 表示模型错误预测为正类的样本数量，FN 表示模型错误预测为负类的正类样本数量。

$$\text{Precision} = \frac{TP}{TP + FP} \tag{11}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

$$\text{F1-Score} = \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 2 \tag{13}$$

## 6. 结果分析

为综合评估本模型与常用文本分类模型对多标签文本分类识别的效果，分别在 Reuters-21578 和 RCV1-v2 两个数据集上进行实验，同时选用准确率、召回率与 F1 值和作为评价指标。表 1 与表 2 分别展示了本模型与对照模型在两个数据集上的表现情况。

**Table 1.** Comparison of results on Reuters-21578 (%)

**表 1.** 在 Reuters-21578 上的结果比较(%)

模型	准确率	召回率	F1 值
基于决策树的模型	72.15	65.32	34.28
基于支持向量机的模型	78.36	70.21	37.03
其他融合注意力机制的简单叠加模型	82.69	75.33	39.42
本模型	89.78	83.46	43.25

**Table 2.** Comparison of results on RCV1-v2 (%)

**表 2.** 在 RCV1-v2 上的结果比较(%)

模型	准确率	召回率	F1 值
基于决策树的模型	80.15	80.32	40.12
基于支持向量机的模型	84.21	83.26	41.87
基于注意力机制的简单叠加模型	81.38	82.27	40.92
本模型	87.56	85.23	43.19

由上述实验可知，本模型在两个数据集上相对对照模型均取得最优结果。在 Reuters-21578 上，基于

决策树的模型在三个指标上的整体表现最差，这是因为其仅考虑文本语义信息，未考虑标签相关性与不同单词的贡献度，导致模型学习能力显著降低；基于注意力机制的简单叠加模型则在对照模型中表现最佳，其通过注意力机制计算词与标签的匹配得分，考虑到标签与文本的影响，但与其他两个模型一样，均未能更深入地学习标签间的相关性以及标签对文本的表示，所以相较本模型较差。

在 RCV1-v2 上，所有对照模型与本模型均表现较为良好，但本模型相比对照模型在三个指标上仍有提升，且都得到最优结果。参照模型在性能上表现最不理想的是基于注意力机制的简单叠加模型，而表现最好的则为基于支持向量机的模型，主要是因为基于注意力机制的简单叠加模型更注重标签与文本的语义关联，而 RCV1-v2 的类别非常明确，且次数较少，对挖掘深层的文本语义与标签关联信息易出现过拟合问题，由此导致分类精度下降，基于支持向量机的模型和基于决策树的模型相对更关注文本语义理解，因此在 RCV1-v2 上表现出更高的学习效率。观察两个数据集的实验结果，各参照模型的表现均存在局限性，而本模型均可获得最优效果，可知本模型的鲁棒性较好。

为进一步验证方法有效性，以 Reuters-21578 数据集为例进行消融实验，选用 Micro-F1、HMicro-F1、Macro-F1、HMacro-F1 这 4 个 F1 指标进行评测，Micro-F1 和 Macro-F1 通常用于评估模型在所有类别上的整体性能，而 HMicro-F1 和 HMacro-F1 则更侧重于考虑类别的样本分布，对于不平衡数据集的性能评估更为敏感。如表 3 所示，本模型分类效果显著优于其他模型，其在 Macro-F1 与 HMacro-F1 上分别提升 3.37% 与 3.69%。

**Table 3.** Results of ablation experiment of Reuters-21578 data set (%)

**表 3.** Reuters-21578 数据集消融实验结果(%)

模型	Micro-F1	HMicro-F1	Macro-F1	HMacro-F1
基于决策树的模型	37.26	36.82	35.77	34.96
基于支持向量机的模型	40.22	39.74	39.15	38.27
基于注意力机制的简单叠加模型	39.85	39.17	38.28	37.58
本模型	42.75	42.45	41.65	41.27

## 7. 结语

综上所述，本文提出了一种融合注意力机制的多标签文本分类方法，取得了显著成果。本方法使用的融合策略能够较好地适应各种数据集，可聚焦文本关键局部信息，从而更好地捕捉语义的复杂性和多样性。在处理多标签间的相关性方面，还可通过学习不同标签之间的关联，提高分类的准确性。在融合上，本方法能够更好地综合不同注意力机制的优点，避免了简单叠加可能带来的信息冗余和冲突。在结构上，注意力机制与文本特征提取部分能够更加紧密地结合，提高了模型的整体性能和泛化能力。在后续研究中，可探索更高效的注意力机制算法，以降低计算复杂度，提高模型的训练和预测速度。同时引入神经网络架构，从而进一步提升模型对复杂语义的处理能力。

## 基金项目

湖南环境生物职业技术学院南岳学者项目：多标签文本分类算法的研究(项目编号：NY2023-01)。

## 参考文献

- [1] 陆佳丽. 基于 Bert-TextCNN 的开源威胁情报文本的多标签分类方法[J]. 信息安全研究, 2024, 10(8): 760-768.
- [2] 王嫒, 胡鹏, 鄢艳玲, 等. 基于标签交互 Seq2Seq 模型的多标签文本分类方法[J]. 传感器与微系统, 2024, 43(8):

155-159.

- [3] 张红. 语义相似度与 BERT 模型融合的多标签文本自适应分类方法[J]. 微型电脑应用, 2024, 40(3): 49-52.
- [4] 刘杰, 唐宏, 杨浩澜, 等. 融合注意力机制的多标签文本分类[J]. 微电子学与计算机, 2023, 40(12): 26-34.
- [5] 文勇军, 刘随缘, 崔志豪. 基于标签语义匹配融合的多标签文本分类方法[J]. 湘潭大学学报(自然科学版), 2024, 46(3): 82-93.
- [6] 李建东, 傅佳, 李佳琦. 融合双向注意力和对比增强机制的多标签文本分类[J]. 计算机工程与应用, 2024, 60(16): 105-115.