

doi:10.6041/j.issn.1000-1298.2021.07.026

# 基于 BERT - LEAM 模型的食品安全法规问题多标签分类

郑丽敏<sup>1,2</sup> 乔振铎<sup>1</sup> 田立军<sup>1</sup> 杨璐<sup>1</sup>

(1. 中国农业大学信息与电气工程学院, 北京 100083; 2. 食品质量与安全北京实验室, 北京 100083)

**摘要:** 在食品安全法规问答系统中, 食品安全法规问题的单标签文本分类不能完全概括问题所包含的有效信息, 为了改进单标签文本分类效果, 根据问题所涉及食品安全角度和层次的不同, 提出一种基于 BERT - LEAM (Bidirectional encoder representational from transformers - label embedding attentive model) 的多标签文本分类方法。采用多角度、分层次的多标签标注方法将单个问题文本赋予多个标签, 并引入 BERT 预训练语言模型表示上下文特征信息, 通过 Attention 机制学习标签与文本的依赖关系, 进行 Word embedding 的聚合, 将标签应用到文本分类过程中。实验表明, 在粗粒度多标签数据集上的分类效果明显优于细粒度多标签数据集上的分类效果, BERT 进行文本特征表示的方法优于 Word2Vec 方法, 采用 BERT - LEAM 模型的分类方法在粗粒度多标签数据集与细粒度多标签数据集的  $F_{1-w}$  值分别为 93.35% 和 79.81%, 其分类效果优于其他分类模型。

**关键词:** 食品安全法规; 多标签分类; BERT; BERT - LEAM

中图分类号: TP391; TP311

文献标识码: A

文章编号: 1000-1298(2021)07-0244-07

OSID:



## Multi-label Classification of Food Safety Regulatory Issues Based on BERT - LEAM

ZHENG Limin<sup>1,2</sup> QIAO Zhenduo<sup>1</sup> TIAN Lijun<sup>1</sup> YANG Lu<sup>1</sup>

(1. College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

2. Beijing Laboratory of Food Quality and Safety, Beijing 100083, China)

**Abstract:** Effective classification of food safety regulatory issues is the key to the realization of the food safety regulatory question and answer system. In order to improve the effect of single label text classification, a multi-label text classification method based on bidirectional encoder representational from transformers - label embedding attentive model (BERT - LEAM) was proposed according to the different food safety perspectives and levels involved in the problem. A multi-angle and hierarchical multi-label labeling method was used to assign multiple labels to a single question text, and the pre-training language model of BERT was introduced to represent the context feature information. The dependency between the label and the text was learned by attention mechanism, the word was processed by embedding aggregation, and the tag was applied to the text classification process. The experimental results showed that the classification effect on the coarse-grained multi-label data set was better than that on the fine-grained multi-label data set. The method of text feature representation by BERT model was better than that of Word2Vec. The  $F_{1-w}$  values of coarse-grained multi-label data set and fine-grained multi-label data set were 93.35% and 79.81%, respectively, which was better than other classification methods model. The problem classification based on food safety regulations question answering system was realized effectively by using the method of BERT - LEAM classification, which laid the foundation for the implementation of the follow-up question answering system.

**Key words:** food safety regulations; multi-label classification; BERT; BERT - LEAM

收稿日期: 2020-09-29 修回日期: 2020-10-30

基金项目: 国家重点研发计划项目(2017YFC1601803)

作者简介: 郑丽敏(1962—), 女, 教授, 主要从事计算机视觉与人工智能研究, E-mail: zhenglimin@cau.edu.cn

## 0 引言

近年来,食品安全问题频发,为了更好地维护消费者合法权益,人们越来越重视食品安全法规问题<sup>[1]</sup>,因此,研究开发食品安全法规自动问答系统具有重要应用价值。然而,食品安全法规问题涉及管理对象、法规内容、法规约束以及法规管理项目等,内容丰富且复杂<sup>[2-3]</sup>,所以,法规问题的有效分类成为实现法规问答系统的关键<sup>[4]</sup>。

在传统的分类方法中,每个样本示例只属于一个类别标记,即单标记学习<sup>[5]</sup>。文献[6]提出 TextCNN 模型,将卷积神经网络(Convolutional neural network, CNN)加入到自然语言处理任务中。随后,深度学习<sup>[7-12]</sup>的方法被广泛应用于单标签文本分类任务中,取得了非常好的效果。但国内外关于多标签文本分类问题尚处于研究阶段<sup>[13]</sup>。

在问答系统的问题分类中,文献[14-15]为深度学习在农业领域的文本分类提供了可行性的依据,但多为单标签分类,且对食品安全法规领域的研究与应用尚显不足。针对食品安全法规的相关问题,不同角度的分类往往涉及多个层次,每个层次同时包含多个类别。文献[16]提出 LEAM 网络模型,将文本分类的标签转换为语义标签向量,从而使分类任务转换为向量匹配任务。受该模型的启发,根据食品安全法规问题特点引入 BERT<sup>[17]</sup> 预训练模型,该模型以大量语料为基础,采用自监督学习方法,可以得到较好的特征表示。针对具体问题,使用 BERT 特征可以明确地表达该问题的词嵌入特征,因此, BERT 可作为一个供其它方法迁移学习的模型,即特征提取模式。

针对目前单标签文本分类方法存在的局限性,为了直观反映食品安全法规问题的多义性,本文提出一种基于 BERT-LEAM 的多标签文本分类方法。首先,采用多角度、分层次的多标签标注方法将单个问题文本赋予多个标签;其次,引入 BERT 迁移学习模型表示上下文特征信息,通过 Attention 机制学习标签与文本的依赖关系,进行 Word embedding 的聚合,将标签应用到文本分类过程中;最后,将常用方法与本文方法进行比较,得到验证结果。以期实现食品安全法规问题的多标签分类,为设计准确、高效的问答系统奠定基础。

## 1 数据与方法

### 1.1 数据获取

为了准确有效地分析用户对于食品安全法规的需求<sup>[18]</sup>,本文筛选出如表 1 所示业内公认的规模较

大、可信度较高的 5 个公共法律咨询服务网站以及常用的公共问答社区——“百度知道”。利用爬虫技术爬取网站中有关食品安全法规的问答数据,共计 7 936 条。具体数据信息分布情况如表 1 所示。

表 1 数据来源  
Tab. 1 Data source

网站名称	信息数量/条
12348 中国法网	1 802
律师 365	1 340
找法网	501
110 法律咨询	1 106
法律快车	756
百度知道	2 431

### 1.2 数据预处理

本文直接爬取上述网站的食品安全法规相关问题数据,其中包含一部分重复问题数据,且具有与文本中心思想无关的标点、乱码、语气词和特殊符号,为方便运算,上述实验数据仍需进行清洗。此外,问题样本数据中仍包含食品安全相关名词的简称,如“QS(企业食品生产许可)”,与文本语义相关,为突出此类文本特征,仍需进行标准化操作,将其替换为汉语全称。数据清洗和标准化具体操作为:①去除无关数字、字母、特殊符号。②去除标点及语气词。③替换名词简称。④去除重复问题数据。经过预处理后,最终得到实验数据为 7 782 条。

预处理完成,获得有效数据,将其分为 2 部分,一部分作为 BERT 向量嵌入模型的输入数据,无需进行分词处理,另一部分用于验证 BERT 特征表示模型效果,需要采用分词操作。本文选用 BERT 预训练模型作为文本分类模型的向量嵌入层,它是基于字符级别的文本特征表示方法,所以该实验无需对数据集进行分词。为了验证 BERT 模型的特征表示效果,选取了 MIKOLOV 等<sup>[19]</sup>在 2013 年提出的预训练词向量工具 Word2Vec 作为对比方法。Word2Vec 是基于词级别的向量嵌入模型,为了确保对照实验的效果最优,对数据集进行了分词处理,将分词之后得到的文本作为对照实验的数据集。为了保证分词结果的准确率,避免食品安全相关名词被切分,在分词过程中加载了食品安全法规词典,该词典由法规名称、部门名称、法规关键词等名词构成。

### 1.3 数据分析及问题分类

对食品安全法律法规相关问题进行分析,分析示例如图 1 所示,其虚线框中为用户提出的问题:“我在超市购买到过期饮品维权的话适用食品法 96 条还是适用消费者保护法 55 条”。



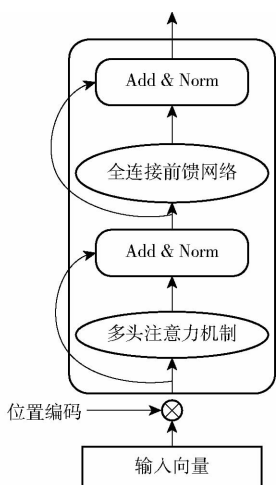


图 3 Transformer 模型示意图

Fig. 3 Schematic diagram of Transformer model

子序列的能力,因此,BERT 模型在数据预处理前加入位置编码,与输入的向量数据求和得到每个字的相对位置。除了位置嵌入特征外,BERT 模型对于输入的文本还有 WordPiece 嵌入和分割嵌入 2 个嵌入特征。此外,BERT 与传统模型不同点在于,预训练任务中,随机遮盖或替换一句话中的字词,计算被遮盖处的损失函数,此种操作便于捕捉词语级别的表征。实验选用了 Google 官方提供的 12 层 Transformer 的 BERT 模型,并选取 BERT 模型的 1 到 12 层输出特征,设置问题文本最大长度为 100,当问题内容相对较少时,在长度不足的位置补 0。

#### 1.4.2 LEAM 网络模型

传统文本分类流程中,直接将 Word 的 embedding 嵌入转变为定长的向量表示,而 LEAM 模型在每个步骤均使用标签信息(图 2)。如图 2 中流程  $f_0$  所示,模型学习了 Label 的 embedding 对 Word 的 embedding 影响,利用 Label 与 Word 之间的相关性进行 embedding 的聚合计算,计算公式为

$$G = (C^T V) \oslash \hat{G} \quad (4)$$

式中  $G$ ——聚合计算结果

$\hat{G}$ —— $K \times L$  阶归一化矩阵

$K$ ——Label 的类别数

$L$ ——Word 长度

$\oslash$ ——同位元素对应相乘

$C$ ——Label - embedding 矩阵

$V$ ——Word - embedding 矩阵

$\hat{G}$  中每个元素为第  $c$  个标签向量嵌入和第  $l$  个单词向量嵌入的  $l_2$  正则化的乘积,公式为

$$\hat{g}_{kl} = \|c_k\| \|v_l\| \quad (5)$$

式中  $\hat{g}_{kl}$ —— $\hat{G}$  中元素

$c_k$ ——第  $c$  个标签向量嵌入

$v_l$ ——第  $l$  个单词向量嵌入

为了更好地获取连续的 Word 之间的空间信息,模型引入卷积和 ReLU 激活函数,如图 2 所示。用  $G_{l-r:l+r}$  衡量中心在  $l$ 、长度为  $2r+1$  的 Label-Phrase 的相关性,第  $l$  个 Phrase 和  $k$  个 Label 的相似度向量计算公式为

$$u_l = \text{ReLU}(G_{l-r:l+r} W_1 + b_1) \quad (6)$$

$$(W_1 \in \mathbf{R}^{2r+1}, b_1 \in \mathbf{R}^k)$$

式中  $G_{l-r:l+r}$ ——以第  $l$  个词为中心词的片段

$W_1, b_1$ ——需要学习的参数

$u_l$ ——相似度向量

$\mathbf{R}$ ——标签与单词的共享向量空间

利用 max\_pooling 得到最大相关系数,即  $m_l = \max\_pooling(u_l)$ 。其中  $m_l$  是长度为  $L$  的向量  $m$  中第  $l$  个元素。在获得向量  $m$  的基础上,经过 softmax 函数后得到 Attention 的系数,如图 2 中的系数  $\beta$ 。文本最终表示  $z$  通过基于标签的 Attention score 加权词嵌入得到,公式为

$$z = \sum_l \beta_l v_l \quad (7)$$

式中  $\beta_l$ ——第  $l$  个元素的 Attention 系数

针对本文中所研究的多标签分类问题,模型将其拆解为  $M$  个单标记问题。训练目标公式为

$$\min_{f \in F} \frac{1}{NM} \sum_{n=1}^N \sum_{m=1}^M CE(y_{nm}, f_2(z_{nm})) \quad (8)$$

式中  $CE(y_{nm}, f_2(z_{nm}))$ ——2 个概率向量的交叉熵

$N$ ——类别数  $M$ ——标签数

$y_{nm}$ ——类别标签的概率表示

$z_{nm}$ ——类别标签的文本特征表示

为了标签起到增加类别判断权重的作用,即相同类别的文本表示之间的距离小于不同类别的文本表示之间的距离。模型引入一个标签的正则化项,公式为

$$\min_{f \in F} \frac{1}{M} \sum_{m=1}^M CE(y_m, f_2(c_m)) \quad (9)$$

#### 1.5 对比模型

(1) BERT - softmax 模型:对 BERT 模型进行 Fine-Tune,通过 softmax 激活函数计算标签与样本的差距,并输出分类结果。

(2) TextCNN 模型:文献[6]提出 TextCNN 模型,将 CNN 在图像领域中取得的成就应用到自然语言处理 NLP 任务中,与传统机器学习方法相比,神经网络会自己构建、选择特征并分类,是端到端的训练,可以达到一种联动的效果,实现全局优化。

(3) VDCNN 模型:文献[20]提出的非常深度卷积神经网络模型(Very deep convolution neural network, VDCNN),利用小尺度的卷积核池化操作,包含了 29 个

卷积层,可以从简单到复杂地提取文本特征。

(4) Self-Attention 模型:文献[21]提出 Self-Attention 模型,自身进行 Attention 操作,改进了 CNN 模型只能捕捉局部文本特征,RNN 模型难以并行化,且两者缺乏对全局特征的理解等缺点。Self-Attention 模型中每个词与文本中所有词均进行 Attention 计算,可以捕获长距离依赖关系。

(5) CapsuleNet 模型:文献[22]提出了 CapsuleNet 胶囊网络,该网络引入动态路由,保留了文本特征的位置信息,解决了卷积神经网络在识别实验对象子结构特点的缺陷,从而保证句子语义不变。

(6) LEAM 模型:参数较少,收敛速度快,其学习过程只涉及一系列的基本代数运算,因此,模型可解释性较强,有效保证准确率的同时还降低了时间成本。

以上述模型为基础,实验对比 Word2Vec 与 BERT 2 种向量嵌入层的文本分类结果。

## 1.6 评价指标

采用  $F_1$  值和汉明损失<sup>[23]</sup> (Hamming loss) 作为性能评价指标。由于数据集中每个标签之间的数据量不均衡,因此,将计算得出的  $F_1$  值取加权平均值  $F_{1-w}$ 。计算公式为

$$F_1 = \frac{2RP}{R+P} \quad (10)$$

$$F_{1-w} = \frac{M \prod_{i=1}^N F_{1_i}}{\sum_{i=1}^N F_{1_i}} \times 100\% \quad (11)$$

$$H = \frac{1}{S} \sum_{i=1}^N |Y_i \Delta Y_i^*| \quad (12)$$

式中  $P$ ——精确率  $R$ ——召回率  
 $S$ ——样本数  $H$ ——汉明损失值  
 $Y_i$ ——真实标签集合  
 $Y_i^*$ ——预测标签集合  
 $\Delta$ ——异或运算符

$F_1$  值作为精确率( $P$ )与召回率( $R$ )的加权平均值,当出现精确率与准确率其中一个值为 1 而另一个值为 0 时这种极端情况时, $F_1$  值可以很好地避免这种情况对实验结果的影响。 $F_{1-w}$  值作为加权平均值,值越大多标签分类效果越好。而  $H$  是衡量样本中误分标签平均数量的指标, $H$  值越小,说明误分标签数量越少,分类模型的效果越好。

## 2 问题分类实验与结果分析

### 2.1 实验设计

#### 2.1.1 数据集划分

实验选取上述食品安全法规问题数据为研究对

象,采用如图 1 所示的分析方法,对经过预处理的所有数据依次进行层次和类别分析,并依据分析结果赋予问题文本单个或多个标签,层次分析后的数据作为粗粒度多标签数据集,类别分析后的数据作为细粒度多标签数据集。2 类数据集标签情况见表 3,其中粗粒度多标签为表 2 中的划分层次,细粒度多标签为表 2 中的类别标签。

表 3 数据集标签信息

标签类型	标签类别数	数据集标签总数	最多标签数	最少标签数
粗粒度多标签	8	15 382	4	1
细粒度多标签	83	16 589	7	1

表 4 为以上 2 种数据集中不同数量标签的问题分布情况。本实验中,2 种数据集均将多标签视为标签组合类,采用 StratifiedKFold<sup>[24]</sup> 分层采样方法对相同标签组合下的数据进行训练集与测试集划分,确保划分结果中各标签组合类之间的比例与原始数据集相同。训练集与测试集的划分比例为 4:1,训练过程中在训练集中随机抽取 10% 进行验证。

表 4 2 种数据集的问题分布情况

标签类型	1	2	3	4	5	6	7
粗粒度多标签	2 976	2 420	1 978	408			
细粒度多标签	2 567	2 473	2 113	474	109	26	20

#### 2.1.2 参数设置

为防止模型过拟合,设定全连接层单元丢弃比例为 0.5。伴随训练轮次增加,梯度下降无法收敛至全局最优点,因此,设定学习率衰减系数为 0.9,学习率衰减步长为 1 000,即每间隔 1 000 次训练,学习率乘以衰减系数为新的学习率。

实验选用 Top\_K\_Categorical\_Accuracy<sup>[25]</sup> 性能评估函数作为训练时评价模型准确率的参数,即测试文本的真实值标签在所有类别标签预测准确率中排前  $k$  名。为保证训练模型的最优化,实验测试了选取不同  $k$  时,测试准确率产生的变化,结果见表 5。由表 5 可知,当  $k=5$  时,分类效果最佳。分类

表 5  $k$  值的确定

$k$	准确率/%	模型训练时间/s
2	79.45	2 710
3	79.07	3 284
4	79.27	3 952
5	79.91	4 365

模型训练可以离线完成,因此,模型应更加关注结果的准确率。

选用 Adam<sup>[26]</sup> 算法进行模型的优化,该优化算法是对随机梯度下降算法的扩展,计算效率高,占用内存少,适合解决例如文本分类这种特征较为稀疏的问题。

## 2.2 结果分析

本文所有实验均在主频为 2.9 GHz 的 i5-9400F 处理器和显卡为 NVIDIA GeForce GTX1660 super 的计算机上进行,采用的神经网络模型使用版本为 2.2.4 的 Keras 进行搭建。实验以 2.1.1 节所述的粗、细粒度多标签数据集为基础,选取 1.5 节所述的模型与本文模型进行对比。细粒度标签数据集实验结果如图 4 和表 6 所示。

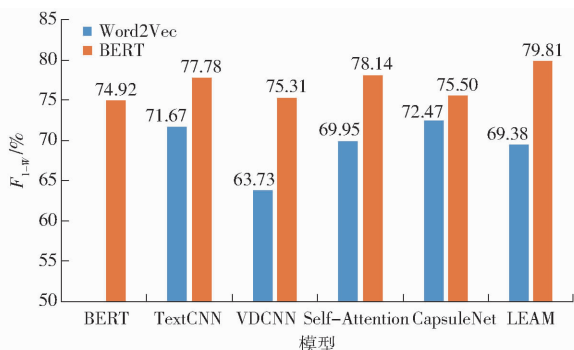


图 4 细粒度多标签  $F_{1-w}$  值对比

Fig. 4 Comparison of fine-grained multi-label  $F_{1-w}$  values

表 6 细粒度多标签  $H$

Tab. 6 Fine-grained multi-label  $H$

嵌入类型	BERT	TextCNN	VDCNN	Self-Attention	CapsuleNet	LEAM
Word2Vec	0.572 6	0.726 6	0.626 6	0.621 1	0.632 3	
BERT	0.363 3	0.302 9	0.432 3	0.296 5	0.422 1	0.288 9

由图 4 和表 6 可知,BERT-LEAM 模型  $F_{1-w}$  值为 79.81%, $H$  为 0.288 9,分类对比效果表明,BERT 作为向量嵌入层进行文本特征表示的方法明显优于 Word2Vec,尤其对 LEAM 模型分类效果的提升更为显著。与 BERT 对接的其他网络相比,LEAM 模型将 BERT 的  $F_{1-w}$  值提高了 4.89 个百分点, $H$  减小了 0.343 4,其优势更为明显。与直接对 BERT 进行 Fine-Tune 的方法相比,其他对接分类网络的方法均提升了分类效果。

实验方法首先将多标签问题转换成单标签多分类问题,标签数量较多,导致样本数据极不均衡,从而导致实验结果  $F_{1-w}$  与  $H$  效果较差,实验选取粗粒度多标签数据集进行了对比,来验证 BERT-LEAM 模型方法的鲁棒性,实验结果如图 5 和表 7 所示。

由图 5 和表 7 可知,BERT-LEAM 模型  $F_{1-w}$  值

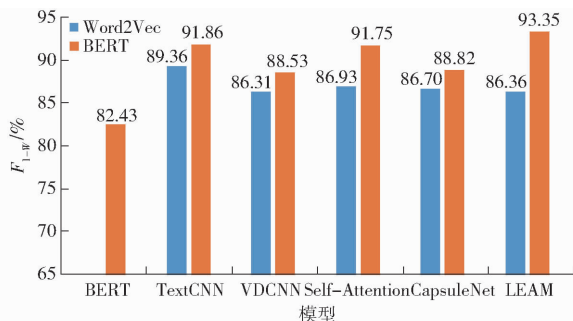


图 5 粗粒度多标签  $F_{1-w}$  值对比

Fig. 5 Comparison of coarse-grained multi-label  $F_{1-w}$  values

表 7 粗粒度多标签  $H$

Tab. 7 Coarse-grained multi-label  $H$

嵌入类型	BERT	TextCNN	VDCNN	Self-Attention	CapsuleNet	LEAM
Word2Vec	0.202 4	0.268 8	0.248 3	0.265 5	0.289 2	
BERT	0.277 1	0.154 4	0.234 6	0.150 1	0.202 6	0.122 6

为 93.35%, $H$  为 0.122 6,对粗粒度多标签数据集的分类效果也明显优于其他模型。实验结果表明,无论是基于粗粒度多标签数据集还是细粒度多标签数据集,本文采用的 BERT-LEAM 模型均获得了分类的最优效果。

由于分类模型的训练可以离线进行,实验对比了 5 种采用不同嵌入类型的模型在 1 556 条粗标签测试数据上的响应时间,对比结果如表 8 所示。

表 8 模型测试响应时间

Tab. 8 Model test response time

嵌入类型	TextCNN	VDCNN	Self-Attention	CapsuleNet	LEAM
Word2Vec	8.75	14.95	8.13	10.99	8.16
BERT	26.98	33.32	24.71	28.72	25.29

由表 8 可知,无论分类网络选择何种嵌入类型,与其他 3 种模型对比,Self-Attention 模型和 LEAM 模型响应时间均较低,其中 Self-Attention 模型主要计算为自身进行 Attention,结构相对简单,LEAM 模型的参数较少,收敛速度快,时间较短的同时仍能保证分类准确率。但是,综合细、粗粒度多标签数据集上的表现,BERT-LEAM 模型分类效果更好。响应时间较长的 TextCNN 模型、VDCNN 模型和 CapsuleNet 模型 3 种模型中,一维卷积层的 TextCNN 模型响应时间明显优于 VDCNN 模型和 CapsuleNet 模型,VDCNN 模型为深层的卷积神经网络,卷积层数较多,因此,测试响应时间最长。CapsuleNet 模型神经网络神经元输出由标量变为矢量,保留了子结构的位置特征,模型计算量较为复杂,因此测试响应时间相对较长。

除分类网络不同以外,预训练模型的选取也会影响模型的测试响应时间。实验选用 BERT 向量嵌

入模型,该模型为动态向量表征方法,能够根据文本上下文语义特征生成句向量,输出向量为768维,且计算参数较多,分类模型的响应时间相对较长。对照实验选用 Word2Vec 预训练向量工具,该工具为固态向量表征方法,当分类模型在训练或测试过程中,加载预训练完成的词向量,且该嵌入层输出向量为300维,计算量较小,因此时间成本较少。但是,综合两种嵌入模型在细、粗粒度多标签数据集上的分类表现,BERT模型明显优于 Word2Vec 模型。与 Word2Vec 模型相比,BERT在 LEAM 分类网络上将  $F_{1-w}$  值分别提高了10.43、6.99个百分点,在 Self-Attention 分类网络上将  $F_{1-w}$  值提高了8.19、4.82个百分点。因 Word2Vec 为固态向量表征模式,无法解决一词多义问题,导致分类效果较差,而 BERT 模型动态生成句向量,可有效规避词语多义性对分类效果的影响。

实验分析了细粒度多标签分类效果相对于粗粒度多标签分类效果降低的原因,统计了测试集中不同量级标签的分布情况,共计83种细粒度标签类别,样本数量大于100的标签为9种,数量处于50~100之间的标签有7种,数量在10~50之间的标签有23种,包括50%以上的细粒度多标签样本

数量在10个以下,其中大部分为食品类别标签,该类别标签所含样本数量和特征较少,导致实验结果不均衡,进而造成测试结果的  $F_1$  值在50%以下。另一个重要原因是,细粒度多标签分类的类别明显多于粗粒度分类,从而使得细粒度分类中每一类的样本数相对较少,进而导致细粒度标签数据集样本分布不均,分类效果较差。

### 3 结论

(1)提出了针对食品安全法规问题的多标签标注方法,并应用到文本分类任务中,有效实现了食品安全法规问题多角度、多层次的信息挖掘,切实改善了传统单标签问题分类数据特征单一、信息表达不完整的问题。

(2)提出了基于 BERT-LEAM 模型的多标签分类方法,该方法能够满足问答系统的实际需求,可有效对食品安全相关问题进行自动分类。基于 BERT-LEAM 模型的分方法在细粒度标签数据集和粗粒度标签数据集上  $F_{1-w}$  值分别达到了79.81%和93.35%,无论数据量还是数据特征,其分类效果均优于其他分类模型,有效实现了食品安全相关问题的多标签文本分类。

### 参 考 文 献

- [1] 魏宏兵. 食品安全问题与食品质量管理方式之探讨[J]. 食品安全导刊, 2019(36):12.
- [2] 编写组. 中华人民共和国食品安全法实用问答[M]. 北京:中国法制出版社, 2009.
- [3] 吴崇其, 李冀宁. 食品安全法律知识问答[M]. 北京:中国协和医科大学出版社, 2011.
- [4] 黄薇屹. 基于知识图谱的深度法律内容问答模型[D]. 深圳:中国科学院大学(中国科学院深圳先进技术研究院), 2020.
- HUANG Weiyi. In-depth legal content question and answer model based on knowledge graph [D]. Shenzhen: University of Chinese Academy of Sciences (Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences), 2020. (in Chinese)
- [5] 肖琳, 陈博理, 黄鑫, 等. 基于标签语义注意力的多标签文本分类[J]. 软件学报, 2020, 31(4):1079-1089.
- XIAO Lin, CHEN Boli, HUANG Xin, et al. Multi-label text classification based on label semantic attention [J]. Journal of Software, 2020, 31(4): 1079-1089. (in Chinese)
- [6] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv Preprint:1408.5882, 2014.
- [7] KIM Y, JERNITE Y, SONTAG D, et al. Character-aware neural language models[J]. arXiv Preprint:1508.06615, 2015.
- [8] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification[C]//Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015:2267-2273.
- [9] ZHOU C, SUN C, LIU Z, et al. A C-LSTM neural network for text classification[J]. Computer Science, 2015, 1(4):39-44.
- [10] LIU P, QIU X, HUANG X. Recurrent neural network for text classification with multi-task learning[J]. arXiv Preprint: 1605.05101, 2016.
- [11] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification[C]//Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2017.
- [12] JOHNSON R, ZHANG T. Deep pyramid convolutional neural networks for text categorization[C]//Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017.
- [13] 王敏蕊, 高曙, 袁自勇, 等. 基于动态路由序列生成模型的多标签文本分类方法[J]. 计算机应用, 2020, 40(7):1884-1890.
- WANG Minrui, GAO Shu, YUAN Ziyong, et al. Sequence generation model with dynamic routing for multi-label text classification[J]. Journal of Computer Applications, 2020, 40(7): 1884-1890. (in Chinese)
- [14] 金宇, 赵春江, 吴华瑞, 等. 基于 BiGRU\_MulCNN 的农业问答问句分类技术研究[J/OL]. 农业机械学报, 2020, 51(5):199-206.
- JIN Ning, ZHAO Chunjiang, WU Huarui, et al. Classification technology of agricultural questions based on BiGRU\_MulCNN [J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(5):199-206. [http://www.j-csam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20200522&journal\\_id=jcsam](http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20200522&journal_id=jcsam). DOI:10.6041/j.issn.1000-1298.2020.05.022. (in Chinese)



- [26] ZHAO B, FENG J, WU X, et al. A survey on deep learning-based fine-grained object classification and semantic segmentation [J]. *International Journal of Automation and Computing*, 2017, 14(2): 119 – 135.
- [27] 刘小刚, 范诚, 李加念, 等. 基于卷积神经网络的草莓识别方法[J/OL]. *农业机械学报*, 2020, 51(2): 237 – 244.  
LIU Xiaogang, FAN Cheng, LI Jianian, et al. Identification method of strawberry based on convolutional neural network [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2020, 51(2): 237 – 244. [http://www.jcsam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=2020026&flag=1](http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?file_no=2020026&flag=1). DOI: 10.6041/j.issn.1000-1298.2020.02.026. (in Chinese)
- [28] 汪传建, 赵庆展, 马永建, 等. 基于卷积神经网络的无人机遥感农作物分类[J/OL]. *农业机械学报*, 2019, 50(11): 161 – 168.  
WANG Chuanjian, ZHAO Qingzhan, MA Yongjian, et al. Crop identification of drone remote sensing based on convolutional neural network [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2019, 50(11): 161 – 168. [http://www.jcsam.org/jcsam/ch/reader/view\\_abstract.aspx?file\\_no=20191118&flag=1](http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?file_no=20191118&flag=1). DOI: 10.6041/j.issn.1000-1298.2019.11.018. (in Chinese)
- [29] TAN C, SUN F, KONG T, et al. A survey on deep transfer learning [C] // *International Conference on Artificial Neural Networks*, 2018: 270 – 279.
- [30] PAN S J, YANG Q. A survey on transfer learning [J]. *IEEE Transactions on Knowledge & Data Engineering*, 2010, 22(10): 1345 – 1359.
- [31] LIN T, ROYCHOWDHURY A, MAJI S. Bilinear CNN models for fine-grained visual recognition [C] // *International Conference on Computer Vision*, 2015: 1449 – 1457.
- [32] SZEGEDY C, IOFFE S, VANHOUCKE V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C] // *National Conference on Artificial Intelligence*, 2016: 4278 – 4284.
- [33] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C] // *Computer Vision and Pattern Recognition*, 2016: 2818 – 2826.
- [34] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C] // *Computer Vision and Pattern Recognition*, 2016: 770 – 778.
- [35] QIAN Q, JIN R, YI J, et al. Efficient distance metric learning by adaptive sampling and mini-batch stochastic gradient descent (SGD) [J]. *Machine Learning*, 2015, 99(3): 353 – 372.
- [36] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C] // *International Conference on Learning Representations*, 2015: 1 – 14.

~~~~~

(上接第 250 页)

- [15] 赵明, 董翠翠, 董乔雪, 等. 基于 BIGRU 的番茄病虫害问答系统问句分类研究[J/OL]. *农业机械学报*, 2018, 49(5): 271 – 276.  
ZHAO Ming, DONG Cuicui, DONG Qiaoxue, et al. Question classification of tomato pests and diseases question answering system based on BIGRU [J/OL]. *Transactions of the Chinese Society for Agricultural Machinery*, 2018, 49(5): 271 – 276. [http://www.jcsam.org/jcsam/ch/reader/view\\_abstract.aspx?flag=1&file\\_no=20180532&journal\\_id=jcsam](http://www.jcsam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=20180532&journal_id=jcsam). DOI: 10.6041/j.issn.1000-1298.2018.05.032. (in Chinese)
- [16] WANG G, LI C, WANG W, et al. Joint embedding of words and labels for text classification [C] // *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics*, 2018.
- [17] DEVLIN J, CHANG M, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [J]. *arXiv Preprint*: 1810.04805, 2018.
- [18] 张馨月. 食品安全知识图谱构建可视化研究[D]. 北京: 中国农业大学, 2020.  
ZHANG Xinyue. Research on visualization of building food safety knowledge graph [D]. Beijing: China Agricultural University, 2020. (in Chinese)
- [19] MIKOLOV T, CHEN K, CORRADO G, et al. Efficient estimation of word representations in vector space [J]. *arXiv Preprint*: 1301.3781, 2013.
- [20] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for text classification [J]. *arXiv Preprint*: 1606.01781, 2017.
- [21] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [J]. *arXiv Preprint*: 1706.03762, 2017.
- [22] SABOUR S, FROSST N, HINTON G. Dynamic routing between capsules [J]. *arXiv Preprint*: 1710.09829, 2017.
- [23] TSOUMAKAS G, KATAKIS I. Multi-label classification: an overview [J]. *International Journal of Data Warehousing and Mining*, 2007, 3(3): 1 – 13.
- [24] KOMER B, BERGSTRA J, ELIASMITH C. Hyperopt – Sklearn: automatic hyperparameter configuration for Scikit-Learn [J/OL]. *Icml Automl Workshop*, 2014: 32 – 37. DOI: 10.25080/Majora-14bd3278-006.
- [25] RAMASUBRAMANIAN K, SINGH A. Deep learning using Keras and TensorFlow [M] // *Machine Learning Using R*. 2019.
- [26] KINGMA D, BA J. Adam: a method for stochastic optimization [J]. *International Conference on Learning Representations*. *arXiv Preprint*: 1412.6980, 2014.