

# 基于深度学习的民事案件判决结果分类方法研究

王立梅<sup>1</sup> 朱旭光<sup>2,3</sup> 汪德嘉<sup>3</sup> 张勇<sup>4</sup> 邢春晓<sup>4</sup>

1 中国政法大学刑事司法学院 北京 100088

2 中国政法大学网络法学研究院 北京 100088

3 江苏通付盾科技有限公司 江苏 苏州 215000

4 清华大学北京信息科学与技术国家研究中心 北京 100084

(limeiw@cupl.edu.cn)

**摘要** 裁判文书数量的快速增长对自动化分类提出了迫切要求,然而已有研究缺乏在民事案件这一细分领域下以判决结果为分类标准的方法的研究,无法实现对民事案件判决结果的准确分类。文中将深度学习技术应用于民事案件判决结果分类领域,通过横向对比多种深度学习模型得出了该领域下表现较好的模型,并依据裁判文书的数据特点对该模型进行了进一步的优化。实验结果证明,Transformer模型的判决结果分类的宏精准率、宏召回率和宏F1分数均高于其他模型。通过对数据预处理流程的优化和对Transformer模型位置嵌入方式的优化,模型的性能指标提升了1%~2%。

**关键词:** 大数据;司法数据;裁判文书;自然语言处理;分类;深度学习

中图法分类号 TP391.4

## Study on Judicial Data Classification Method Based on Natural Language Processing Technologies

WANG Li-mei<sup>1</sup>, ZHU Xu-guang<sup>2,3</sup>, WANG De-jia<sup>3</sup>, ZHANG Yong<sup>4</sup> and XING Chun-xiao<sup>4</sup>

1 School of Criminal Justice, China University of Political Science and Law, Beijing 100088, China

2 Institute of Cyber Law, China University of Political Science and Law, Beijing 100088, China

3 Jiangsu PayEgis Technology Co., Ltd., Suzhou, Jiangsu 215000, China

4 Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China

**Abstract** The rapid increase in the number of judgment documents puts forward an urgent need for automated classification. However, there is a lack of method in existing studies that use judgment results as the subject of classification in the subdivision of civil cases, and therefore they cannot achieve accurate classification of judgment results in civil cases. In this paper, we apply deep learning technology in the field of classification of judgment results of civil cases, and obtain a model with better performance in this field through horizontal comparison of multiple deep learning models. This model is further optimized based on the data characteristics of the judgment document. After experiments, the Transformer model's macro precision rate, macro recall rate and macro F1 score in the judgment result classification are all higher than other models. By adjusting the data preprocessing process and adjusting the position embedding method of the Transformer model, the performance index of the model is increased by 1%~2%.

**Keywords** Big data, Judicial data, Judgment documents, Natural language processing, Classification, Deep learning

### 1 引言

近年来,随着大数据技术的快速发展,社会数字化的进程也在快速推进。如何应用大数据技术为数字化的司法数据持续赋能,成为了当前司法数字化领域的热点研究话题之一,这也是“智慧法院”发展纲要的重要研究方向<sup>[1]</sup>。

裁判文书是司法数据的重要组成部分。作为诉讼结果的

载体,裁判文书是人民法院确定和分配当事人实体权利义务的唯一凭证,具有结构完整、逻辑严谨、要素齐全的特征。判决结果是对判决内容的提炼概括,使阅读者能够快速理解案件的最终裁判结果。

依据判决结果对裁判文书进行分类有着其独特的应用意义,通过对民事案件判决结果的分类,民众可以依据限定条件快速检索相关文件,为个人法律活动提供参考和指导;法院等

到稿日期:2021-03-12 返修日期:2021-05-21

基金项目:国家重点研发计划(2018YFC0831202)

This work was supported by the National Key R&D Program(2018YFC0831202).

通信作者:朱旭光(xuguangs@yeah.net)

司法机构则可以通过数字化的裁判文书数据为审判等执法流程提供理论依据,贯彻“类案类判”的判决方针;律师等法律从业人员也可以从数字化的司法文书中快速获得需要的司法信息。

当前对裁判文书的数字化分类面临诸多困难。一方面,判决书增长速度迅猛,根据裁判文书网的统计,当前数字化裁判文书总量约 1.14 亿篇,单日新增约 5 万篇。面对海量的历史数据和大量的新增数据,传统的人工标注判决结果的分类方式难以在司法领域得出广泛应用。另一方面,裁判文书中的诉讼结果作为裁判文书的核心部分,不能通过关键词的方式分类。具体原因是:案件本身存在多样性,原告的诉求往往数量繁多、形式多样、内容复杂,导致判决结果也较为复杂,难以通过简单的文字模板对判决结果进行快速分类。例如,在民事案件判决结果中,文书会对判决的执行内容进行详细描述,但不会给出原告“胜诉、败诉、多少条诉求被支持”等明确字样供机器在分类时进行参考,需要通过语义学习的方式对判决结果进行分类。

深度学习与自然语言处理技术的发展可以基于自然语义实现对非结构化数据的处理。非结构化数据指具有不完整或不规则数据结构,不具备标准的数据模型,也不能简单地使用数据库逻辑来存储数据。裁判文书便是典型的非结构化数据。以裁判文书为文本挖掘的对象,应用自然语言处理等技术可以实现对裁判文书数据价值的进一步挖掘。例如,通过基于深度学习的文本分类模型来实现对裁判文书的段落进行分类,通过情感分析对判决结果进行预测分类,以及根据判决书对案件罪名进行分类等,这些都进一步提升了裁判文书数字化的应用价值。

然而,已有工作使用深度学习技术对裁判文书进行分类时,往往未对裁判文书进行进一步的领域细分,分类标准也不是判决结果,无法实现对民事案件最终判决结果的准确分类。本文提出了一种基于深度学习技术的文本挖掘方法,以中国裁判文书网民事案件判决书为文本挖掘对象,利用深度学习模型,实现根据民事案件判决书中的判决部分对案件判决结果的自动分类。本文的主要贡献如下:

(1)首次将深度学习技术应用于基于裁判文书判决部分的民事判决结果分类中,具体为使用 Transformer, BERT, CNN, RNN, Bi-LSTM 等神经网络模型对民事案件判决结果进行自动分类,代替人工标注的方式,大幅提升了分类效率。

(2)使用真实的裁判文书数据集进行了可行性验证,横向对比了多种神经网络文本分类模型框架在真实民事裁判文书数据集上的表现,得出了在该领域的最优分类模型推荐。

(3)针对裁判文书细分种类多样的特征,添加了基于正则表达的数据预处理方式,提高了训练样本的质量;针对民事案件判决书用词精简、句长相对固定的特征,对 Transformer 模型进行了进一步的优化,使用绝对位置嵌入的方式代替正余弦函数表达位置信息的方式,提升了模型的精度。

## 2 相关工作

### 2.1 主流文本分类方法研究

Tang 等<sup>[2]</sup>提出了一种面向文本分类的 Transformer-cap-

sule 集成模型。该方法针对单模型文本分类算法不能很好地提取到文本序列多层次特征的问题,分别利用胶囊网络(capsule network)和 Transformer 来提取文本的局部短语特征和全局语义特征,通过集成的形式更全面地得到文本序列的多层次特征表示。

Zhu 等<sup>[3]</sup>提出了一种使用双向 LSTM 与 CNN 模型实现文本分类的方法。该方法应用注意力机制以及 aspect-level embedding 来实现文本的情感极性分析,根据文本情感实现了对文本的分类。

Prasad 等<sup>[4]</sup>提出了一种文本挖掘方法,通过一种混合颗粒(granular hybrid)方法实现文本挖掘过程中的文本分类。该方法可以在目标分类较多的情况下仍然保持较高的准确率。

Bugueno 等<sup>[5]</sup>探讨了文本分类中类别数量不均衡的问题。部分目标类别重要性很高但包含的样本数量少,从而导致重要的类别不能被学习器正确学习,可能使分类模型准确性虚高而使用性不高。他们还讨论了一种基于 Transformer, BERT 等模型框架下的类别合并方法,用于解决文本分类中类别数量不均衡的问题。

Zhang 等<sup>[6]</sup>提出了一种改进后的 BERT 模型进行分类,结合了深度学习技术以提升 BERT 模型的推理效率与准确率。Chen 等<sup>[7]</sup>提出了一种针对 BERT 模型进行 fine-tuning (精校)的方式,提升了 BERT 模型在文本分类中的应用效果。Yu 等<sup>[8]</sup>提出了一种 BERT 与 BiGRU 模型相结合的方式,提升了文本多分类问题的准确性。Lee 等<sup>[9]</sup>提出了一种使用 BERT 进行自动化文章主题提炼的方法。

本文应用上述文本分类方法,对比其在裁判文书分类领域的应用效果。

### 2.2 针对裁判文书的文本分类研究

针对裁判文书段落分类,Weng 等<sup>[10]</sup>通过 Text CNN, BERT, BiLSTM-CRF 模型,结合段落级别的上下文语义,对裁判文书的段落结构进行智能自动化分类,并通过准确率、召回率和 F1 分数 3 个指标对上述 3 个模型进行分类结果进行了横向比较。该文还提出了一种基于段落级别的上下文语义特征信息的序列标注方法,通过学习完整的裁判文书中段落标签的结构信息、段落上下文之间的联系,来实现良好的裁判文书结构化效果。该研究可以识别出裁判文书的判决部分,但并没有实现对判决内容的进一步分类。

针对裁判文书情感倾向分类,Wang 等<sup>[11]</sup>利用 BiGRU 模型,并根据情感分析结果预测判决结果的倾向性,实现了基于情感分析对判决结果进行预测的初步尝试。该研究可以实现对判决结果的情感倾向性分析,但不能实现对判决结果的准确分类。

针对裁判文书法条与判决结果分类,Cheng<sup>[12]</sup>利用 RNN 网络与 BiLSTM 模型,使用裁判文书网的真实文书数据进行实验,并结合注意力机制对模型的卷积层、池化层、Embedding 层进行设计调整,使模型达到较好的效果。所提模型能够较好地解决相似案例匹配中训练数据较少且文书间不易区分的问题。Radha 等<sup>[13]</sup>利用神经网络实现了对判决罪名的自动分类,并在网络犯罪领域对模型进行了可行性验证与优

化尝试。这两项研究成果实现了对判决书引用的法条和罪名的分类,但不能实现对最终判决结果的准确分类。

相比之前的研究,本文方法探索了 Transformer, BERT, textCNN, textRNN 等多种神经网络模型在民事案件裁判文书判决结果分类这一任务中的应用效果,并对多个模型进行了横向比较。通过使用真实的裁判文书数据,验证了不同模型在这一司法数据领域的应用效果,并对该领域下表现较好的模型提出了进一步的优化建议。

### 3 系统架构及算法介绍

本文用于民事判决结果分类的系统的主要架构图如图 1 所示,包括 3 个层次:数据层、算法层和应用层。

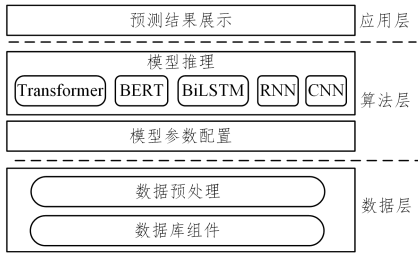


图 1 系统架构图

Fig. 1 System structure diagram

数据层主要包含数据库组件和数据预处理组件。数据库提供非结构化数据存储与快速数据检索功能;数据预处理组件实现对数据的初步筛选和格式转换,通过正则匹配的方式提取判决书中的判决部分,并对空值等异常值进行删除处理,将元数据转换为可被模型使用的 Pandas Dataframe 数据格式。

算法层包括模型参数配置模块和模型推理模块,模型推理模块根据模型参数配置选择对应的模型和参数来对数据进行推理,得出预测结果。算法层使用的模型框架包括 Transformer, Bert, CNN(用于 textCNN), RNN 以及 Bi-LSTM(用于 textRNN)。

应用层包括模型预测结果展示功能,提供预测结果的可视化展示,具体展示内容包括每一个类别的预测概率,以及每一个模型训练迭代(epoch)的精准率与召回率指标。

#### 3.1 Transformer

Transformer 是谷歌于 2017 年提出的一个深度学习模型框架<sup>[14]</sup>。它提出了一种基于注意力的特征抽取机制,大幅提升了模型的准确率和运算效率。该模型的核心机制——注意力的定义如下:

$$Attention(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

其中,  $Q$  表示 query, 对应需要表达的序列;  $K$  和  $V$  表示 key 和 value, 对应用来表达  $Q$  的序列,  $\text{softmax}$  是激活函数, 将多个神经元的输出映射到  $(0, 1)$  区间。通过独特的注意力机制, Transformer 模型可以使词向量拥有全局信息, 从而提升模型对全局信息的理解能力, 使该模型在文本分类、关系抽取、机器翻译等领域均有不错的表现。裁判文书具有结构内容复杂、文本前后关联性强的特征, Transformer 的注意力机制可以更好地处理这种结构, 需要大量前后关联的文本分类任务。

#### 3.2 BERT

BERT 即 Bidirectional Encoder Representations from Transformers, 同样由谷歌提出, 在 Transformer 的基础上增加了双向编码表征。双向编码表征可以使模型在处理一个词向量的同时考虑该词前后单词的含义, 实现对上下文语义的理解, 进而提升模型在复杂语义理解领域的表现能力<sup>[15]</sup>。裁判文书具有用语专业、用词紧凑的行文特征, BERT 的双向编码表征可以更好地理解专业司法陈述中的复杂语义。

#### 3.3 textCNN

卷积神经网络(Convolutional Neural Networks, CNN)是一种经典的神经网络模型,它具有表征学习的能力,能按照阶层之间的排列信息对输入的序列进行平移不变分类,在图像识别领域有着成熟的应用<sup>[16]</sup>。textCNN 和 CNN 的模型结构一致,但模型的输入变成了由词向量组成的句子矩阵,使其可以应用于文本分类任务<sup>[17]</sup>。textCNN 是文本分类任务中出现较早、框架较为成熟的模型,在短文本分类上取得了不错的成绩。针对裁判文书判决部分中较为精简紧凑的判决语句, textCNN 可能会有不错的表现。

#### 3.4 textRNN

循环神经网络(Recurrent Neural Networks, RNN)也是一种经典的神经网络模型。不同于传统的前馈神经网络, RNN 模型加入了一种定向循环机制,使其能够处理输入文本之间的前后关联关系。但是,传统的循环神经网络不能有效处理时间步长差距过大的关联信息,导致其在长句或是整段文本的上下文语义理解任务上表现不佳。textRNN 在原有模型基础上添加了长短记忆网络(Long Short-Term Memory, LSTM),大幅增强了模型在长文本分类领域的能力<sup>[18]</sup>。裁判文书的判决部分存在对判决执行方式的详细描述,属于典型的长文本, textRNN 中的长短记忆网络可以更好地识别这部分内容。

## 4 实验分析

#### 4.1 数据集与实验环境介绍

为了探究不同文本分类模型框架在真实司法数据上的表现,我们通过中国裁判文书网收集了 10000 份民事判决书,审判程序限定为民事一审和民事二审,其中民事一审 8734 件,民事二审 1266 件。

本实验通过 Python Tensorflow1.8 完成对模型框架的搭建、模型训练以及模型的推理计算。操作系统环境为 Windows10, 硬件环境为 4 核 CPU; AMD Ryzen 9 5900HX 3.30 GHz, 内存为 16 GB, GPU 为 Nvidia 3070, 显存为 8 GB。

#### 4.2 样本标签介绍

本次实验旨在对民事判决结果进行分类。根据一审、二审判决结果性质的不同,需要制定不同的分类标签,其中一审判决多为对原告诉求的支持判断,可以通过支持诉求的数量进行胜负标签的制定;二审判决是对原判的支持判断,结果主要集中在维持、撤销和改判这 3 个类别。通过对样本数据的分析、对大量裁判文书的调研以及与民事法律领域的专家的沟通学习,最终制定分类标签如下。

一审判决结果分类如表 1 所列。

表1 一审判决书标签

Table 1 First instance judgment label

类别标签名称	类别标签含义
一审原告胜诉	至少一个原告诉求得到支持
一审原告败诉	全部原告诉求被驳回
其他	不属于上述任何一个类别

二审判决结果分类如表2所列。

表2 二审判决书标签

Table 2 Second instance judgment label

类别标签名称	类别标签含义
二审维持原判	判决结果为维持一审判决
二审撤销判决	判决结果为撤销一审判决,但驳回全部二审原告诉求
二审改判	判决结果为撤销一审判决,至少一个二审原告诉求得到支持
其他	不属于上述任何一个类别

### 4.3 数据预处理

通过判决书名称可以直接区分出一审、二审判决,再通过一审、二审民事判决书的固定模板截取判决部分,对于不能通过模板判断判决部分的判决书,则保留全部内容。同时,删除因各种原因而不能公开的判决书(空判决书),并采集新的判决书进行数据补齐。

为了加快神经网络的收敛速度,提高模型的训练效率,嵌入(embedding)后的数据需要进行中心化处理,采用的方式是零均值化:

$$x = x - \mu \quad (2)$$

其中, $X$ 是数据集中的每一项, $\mu$ 是数据集的均值。

原始数据按照7:3的比例划分为训练集和验证集。

### 4.4 模型训练

实验采用标准Transformer, BERT, textCNN和textRNN模型进行训练。为了统一对比标准,不同模型的训练参数尽可能保持一致。模型的主要通用训练参数如表3所列。

表3 参数

Table 3 Parameters

参数名	参数值
失活系数(dropout)	0.5
嵌入尺寸(embedding size)	128
批尺寸(batch size)	32
学习率(learning rate)	$5 \times 10^{-5}$
迭代次数(epochs)	5

### 4.5 模型推理

使用训练好的模型对验证集数据进行预测,单次预测结果如图2所示。

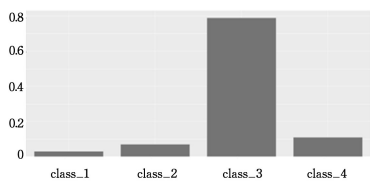


图2 预测结果示例图

Fig. 2 Forecast results example diagram

如图2所示,针对二审判决书,模型会给出4个类别的预测概率, $x$ 坐标中的class\_1至class\_4分别对应二审维持原判、二审撤销判决、二审改判和其他, $y$ 坐标轴为预测概率。系统会选择概率最高的一类作为模型预测的最终结果。以

图2为例,该次模型预测的最终结果为class\_3,即“二审改判”。

### 4.6 模型结果评估

实验对4个模型框架的表现进行了横向对比,对比指标主要包含以下3个。

(1)宏精准率(macro-P)

在二分类问题下,精准率为正确预测为正占全部预测为正的比率;在多分类问题下,宏精准率为多个混淆矩阵所计算出的精准率的均值。

(2)宏召回率(macro-R)

在二分类问题下,召回率为正确预测为正占全部正样本的比率;在多分类问题下,宏召回率为多个混淆矩阵所计算出的召回率的均值。

(3)宏F1分数

宏F1分数指宏精准率和宏召回率调和均值的2倍。

实验结果总结如下:

如图3和图4所示,在使用标准模型框架与模型参数的条件下,Transformer模型在民事案件判决分类数据集上的表现系统性地优于BERT, textRNN和textCNN这3种模型。在一审判决结果分类任务上,其宏精准率可以达到约0.96,宏召回率可以达到约0.95,宏F1分数可以达到约0.96;在二审判决结果分类任务上,其宏精准率可以达到约0.91,宏召回率可以达到约0.90,宏F1分数可以达到约0.91,系统性地高于其他3种模型的对应指标。

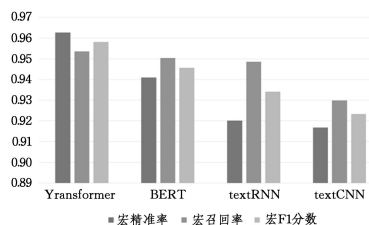


图3 模型指标对比图(一审)

Fig. 3 Model index comparison diagram (first judgement)

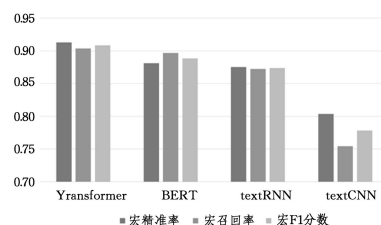


图4 模型指标对比图(二审)

Fig. 4 Model index comparison diagram (second judgement)

从实验结果来看,基于注意力的Transformer和BERT模型略优于传统的RNN/CNN结构的深度学习模型,这符合我们的实验预期。裁判文书的结构较长,判决部分会对原告的多个诉求进行详细说明,需要模型具备较高的上下文理解能力和段落理解能力,在这点上,注意力机制有着明显的优越性。尽管BERT在Transformer的基础上添加了双向编码表征,理论上增强了对词语前后词的理解能力,但其实际表现不如基本的Transformer模型,可能的原因是裁判文书的司法术语较多且用词精简,单一词语已经可以独立表意,强行关联前后词进行理解反而会影

在实验中,部分样本被4个模型错分。以某建设工程施工合同纠纷案判决书为例,该判决书为二审判决,判决中对一审判决中的多个分项分别进行了维持、撤销、改判处理,正确的分类应为二审改判。但在描述中,由于改判的部分较少(仅有一项)且只是对支付金额进行调整,没有出现“改判”“更改”等明显字样(但出现了“变更”字样),判决整体描述部分也没有对改判部分进行详细说明,而撤销的判决项较多,导致4个模型均把此判决书归类为“二审撤销判决”。该类错误说明训练样本的覆盖度不够,不能全面覆盖所有“二审撤销判决”的表述方式,后续可以通过增加训练样本数量来进行优化。

#### 4.7 模型优化

针对模型结果和裁判文书的数据特征,我们提出了一种对当前最优模型(即Transformer模型)的进一步优化方案,包括对数据预处理流程的优化和对Transformer位置编码方式的优化。

首先是对数据预处理流程的优化。裁判文书具有细种类多样的特征,具体种类包括判决书、裁定书、调解书、决定书和通知书等,除去常见的判决书之外,其余形式的裁判文书多为案件流程的中间结果,不在本次判决结果分类的分析目标之列。例如,样本中包含少量的裁定书,裁定结果是对案件管辖归属地的规定,不属于最终判决。这类裁判文书往往会被模型错误分类,从而影响最终模型的准确率。通过分析发现,各类裁判文书均具备固定的判决格式,可以通过正则表达式(regular expression)的方式进行过滤。例如,全部裁定书在裁定部分均包含“裁定如下”字样,可通过正则匹配过滤掉全部裁定文书。通过对样本的进一步过滤可以提升样本纯度,进而提高模型的判断准确率。

其次是对模型位置编码方式的调整。Transformer使用正余弦函数来表示词向量之间的绝对位置,通过正弦余弦公式可实现一个词位置信息被另一个词位置线性表示<sup>[14]</sup>,具体公式如下:

$$PE(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (3)$$

$$PE(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \quad (4)$$

其中, $d$ 是词向量的维度, $2i$ 表示偶数维度, $2i+1$ 表示奇数维度, $pos$ 是词在句子中的位置。使用正余弦函数表示词向量间绝对位置的优点是可以适应样本中未出现过的句子长度来防止过拟合,但代价是准确率也会有所损失。然而,裁判文书具有用词精简、句长相对固定的特征,模型过拟合风险较小,使用绝对位置编码代替正余弦函数可能有更好的效果。绝对位置编码通过网络来学习位置关系,具体实现方式为在每一个位置随机初始化位置嵌入信息,并将其添加到词嵌入中,作为参数在模型中进行训练,具体如图5所示。

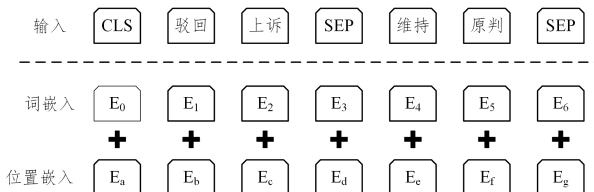


图5 Transformer位置嵌入修改

Fig. 5 Changing the position embedding of Transformer

通过数据预处理流程的优化和对Transformer模型位置编码方式进行调整后,模型的指标在一、二审判判决书的分类上均有了小幅上升。其中,一审判决宏精准率提升了0.01,达到了0.97,宏召回率提升了0.02,达到了0.97,宏F1分数提升了0.01,达到了0.97;二审判决宏精准率提升了0.02,达到了0.93,宏召回率提升了0.02,达到了0.92,宏F1分数提升了0.02,达到了0.93。

**结束语** 裁判文书是典型的司法数据,裁判文书的数字化有着重要的现实应用意义。对民事案件判决结果的分类可以指导民众的法律活动,并为执法机构提供参考依据。然而,裁判文书基于判决结果的分类面临着数量庞大、人工分类效率低、机器分类难以实现等诸多问题。已有工作往往未将判决结果作为分类标准,也缺乏对民事判决结果分类这一细分领域的判决结果进行分类研究,无法实现对民事案件最终判决结果的准确分类。本文提出了一种深度学习技术在民事案件判决结果分类领域的应用方式,通过横向比对多种现有深度学习模型框架,以真实的裁判文书数据集对模型进行了可行性验证,得出Transformer模型在该领域的优越性,根据裁判文书细种类多样的特征添加了新的数据预处理方式,通过正则表达式的方式过滤掉噪声样本,提高了训练样本的质量;根据民事案件判决书用词精简、句长相对固定的特征,使用绝对位置嵌入的方式代替Transformer模型正余弦函数位置嵌入的方式,提升了模型的精度。受研究周期和实验环境等因素的限制,本文提出的实验方法在数据采集的量级上还有不足,下一步计划扩充样本量级至全量民事案件判决书数据,以提升模型的稳定性与泛化能力。因为更高的数据量级能够提供更全的词汇表,可以通过优化词嵌入方式进一步提升模型的性能。

#### 参考文献

- [1] WANG L S. China's experience in the construction of smart courts and its path optimization—Based on the application of big data and artificial intelligence [J]. Inner Mongolia Social Sciences, 2021, 42(1): 104-114.
- [2] TANG Z, WANG Z S, ZHOU A, et al. Transformer-capsule ensemble model for text classification [J]. Computer Engineering and Applications, 2020, 56(24): 151-156.
- [3] ZHU Y, GAO X, ZHANG W, et al. A Bi-Directional LSTM-CNN Model with Attention for Aspect-Level Text Classification [J]. Future Internet, 2018, 10(12): 116.
- [4] PRASAD K M S, REDDY D T H. Text Mining: Classification of Text Documents using Granular Hybrid Classification Technique [J]. International Journal of Research in Advent Technology, 2019, 7(6): 1-8.
- [5] BUGUENO M, MENDOZA M. Learning to combine classifiers outputs with the transformer for text classification [J]. Intelligent Data Analysis, 2020, 24(1): 15-41.
- [6] ZHANG A, LI B, WANG W, et al. MII: A Novel Text Classification Model Combining Deep Active Learning with BERT [J]. Computers, Materials & Continua, 2020, 63(3): 1499-1514.
- [7] CHEN L, SHAH R, LINK T, et al. Bert model fine-tuning for

text classification in knee OA radiology reports[J]. *Osteoarthritis and Cartilage*, 2020, 28(1): S315-S316.

- [8] YU Q, WANG Z, JIANG K. Research on Text Classification Based on BERT-BiGRU Model[J]. *Journal of Physics: Conference Series*, 2021, 1746(1): 012019.
- [9] LEE T S, KANG S S. Automatic Text Summarization Based on Selective OOV Copy Mechanism with BERT Embedding[J]. *Journal of KIISE*, 2020, 47(1): 36-44.
- [10] WENG Y, GU S Y, LI J, et al. A Text Classification Algorithm for the Structured Large-scale Judgment Documents[J]. *Journal of Tianjin University (Natural Science and Engineering Technology Edition)*, 2021, 54(4): 418-425.
- [11] WANG N, LI S L, LIU T L, et al. Tendency analysis of BiGRU decision results based on attention mechanism[J]. *Computer Systems & Applications*, 2019, 28(3): 191-195.
- [12] CHENG H. Law Prediction and Similar Case Matching Research Facing Judicial Big Data[D]. Taiyuan: Shanxi University, 2020.
- [13] MOTHUKURI R, BASAVESWARARAO B, SUNEETHA B. Judgement Classification Using Hybrid ANN-Shuffled Frog Leaping Model on Cyber Crime Judgement Database[J]. *Rev. d'Intelligence Artif*, 34(4): 445-456.
- [14] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS' 17)*. 2017: 6000-6010.
- [15] LI K Y, CHEN Y, NIU S Z. BERT-based social e-commerce text

classification algorithm [J]. *Computer Science*, 2021, 48(2): 87-92.

- [16] NEBAUER C. Evaluation of convolutional neural networks for visual recognition[J]. *IEEE Transactions on Neural Networks*, 1998, 9(4): 685-696.
- [17] WU H Y, YAN J, HUANG S B, et al. CNN\_BiLSTM\_Attention Hybrid Model for Text Classification[J]. *Computer Science*, 2020, 47(11A): 24-27.
- [18] DU L, CAO D, LIN S Y, et al. Extraction and Automatic Classification of TCM Medical Record Text Based on BERT and BiLSTM Fusion Attention Mechanism [J]. *Computer Science*, 2020, 47(11A): 416-420.



**WANG Li-mei**, born in 1974, Ph.D, professor. Her main research interests include cyber law and so on.



**ZHU Xu-guang**, born in 1989, master. His main research interests include artificial intelligence and cyber law.