

基于BERT阅读理解框架的司法要素抽取方法

黄辉¹, 秦永彬^{1,2}, 陈艳平^{1,2}, 黄瑞章^{1,2}

1. 贵州大学计算机科学与技术学院, 贵州 贵阳 550025;

2. 公共大数据国家重点实验室, 贵州 贵阳 550025

摘要

司法要素抽取是司法智能化辅助应用的重要基础,其目的是判别裁判文书涉及的关键案情要素。以往司法要素抽取通常采用多标签分类方法进行建模,模型主要依赖于裁判文书文本特征,忽略了要素标签的语义信息。同时,由于司法数据集存在样本分布不均衡的情况,分类方法会因负例过多而导致模型性能不佳。针对上述问题,提出基于BERT阅读理解框架的司法要素抽取方法。该方法引入标签信息和法律先验知识构造辅助问句,利用BERT机器阅读理解模型建立辅助问句和裁判文书之间的语义联系。同时,在问句中标签所在位置前后增加特殊标识符以增强模型的学习能力。实验结果表明,该方法在CAIL2019要素抽取公开数据集上性能得到显著提升,在婚姻家庭、劳动争议、借款合同3种案由上分别提升F1值2.7%、11.3%、5.6%。

关键词

要素抽取; 机器阅读理解; 神经网络; BERT

中图分类号: TP391.1

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2021057

Legal element extraction method based on BERT reading comprehension framework

HUANG Hui¹, QIN Yongbin^{1,2}, CHEN Yanping^{1,2}, HUANG Ruizhang^{1,2}

1. School of Computer Science and Technology, Guizhou University, Guiyang 550025, China

2. State Key Laboratory of Public Big Data, Guiyang 550025, China

Abstract

Extraction of legal elements is an important basis for judicial intelligent auxiliary applications, and its purpose is to identify the key elements involved in the judgment document. In the past, extracting legal elements usually used multi-label classification methods for modeling. These methods mainly relied on the text features of the judgment document, thereby ignoring the label features. Besides, due to the imbalanced data problem in judicial data sets, the classification method will lead to poor model performance because of too many negative examples. To solve the above problems, a legal element extraction method based on BERT reading comprehension framework was proposed. This method constructed auxiliary questions with label information and legal prior knowledge, and used the machine reading comprehension model based on BERT to establish the semantic associations between question and judgment document. And this method added special

tokens before and after the label in the question to enhance the learning ability of the model. Experiments were conducted on the legal element extraction data sets of the CAIL2019. Experiment results show that the performance is improved significantly, and the F1 value has been increased by 2.7%, 11.3%, and 5.6% respectively on the data sets of marriage and family case, labor dispute case, and loan contract dispute case.

Key words

element extraction, machine reading comprehension, neural network, BERT

1 引言

随着智慧法院建设的提出,大数据、人工智能与司法实践的融合开始不断推进。司法人工智能的本质是基于已有案例的审判工作,提示和辅助司法人员进行审判决策。因此,为司法人员提供可解释的决策依据是司法智能化建设的重点。将端到端的深度学习应用于司法领域通常会存在可解释性差的问题,无法为司法人员提供模型决策依据。案情要素体系是司法领域专家根据法院审判流程总结、建立的体系,主要由司法人员在决策过程中关注的案情要点构成。利用深度学习模型识别裁判文书中的关键案情要素,可以为案件判决预测提供依据,在一定程度上实现可解释性。并且,司法要素是判定案件之间关联性的重要依据之一,具备相同案件要素的案件拥有类似的案情和判决结果,结合司法要素可以有效地实现具有可解释性的类案推送。因此,司法要素抽取工作是司法人工智能中必不可少的一环。

司法要素抽取任务需要根据既定案情要素体系,判定裁判文书案情内容包含哪些具体案情要素。例如,给定句子“原告认为被告违法解除与原告的劳动关系,故原告起诉至法院,要求判决:被告向原告支付赔偿金186 600元”,根据语义信息可以判定该句子包含既定司法要素中的“解除劳动关系”“支付经济补偿金”。以往的司法要

素抽取采用自然语言处理(natural language processing, NLP)中的多标签分类方法来完成,然而多标签分类方法只利用了句子的文本信息,从上述例子中可以看出,司法要素和句子之间有很强的关联性,如何充分利用要素标签信息是本文研究的重点。

此外,司法要素抽取数据集中通常会存在样本分布不均衡的问题。在劳动争议案由中,大量裁判文书包含“解除劳动关系”要素,但涉及“经济性裁员”和“有调解协议”两种要素的案件非常少。在借款合同案由中也是如此,“债务人转让债务”“约定利率不明”等要素极少在案件中出现。训练样本分布极度不均衡导致分类模型的推理能力较差。

综上,本文提出基于BERT(bidirectional encoder representations from transformer)阅读理解框架的司法要素抽取方法。该方法基于BERT抽取式机器阅读理解框架,通过引入要素标签信息和法律先验知识构造相应的辅助问句,建立辅助问句和文本之间的语义联系,为模型提供语义完整的标签信息和先验知识。同时,在问句中标签所在位置前后增加特殊标识符以增强模型的学习能力。实验证明本文方法不同于传统分类方法,不再受限于样本,可被很好地应用于分布不平衡的数据集。

2 相关工作

在司法领域,一段案情描述往往涉及

多个标签,因此,司法要素抽取工作大多以多标签分类的形式开展。当前解决多标签文本分类任务的方法可以分为两类:机器学习方法和神经网络方法。传统机器学习方法需要对特征进行筛选,Elisseeff A等人^[1]采用类似于学习系统的支持向量机(support vector machine, SVM)来解决多标签问题;Ghamrawi N等人^[2]将条件随机场应用于该任务;Li C等人^[3]提出了基于条件伯努利混合的多标签分类方法,通过引入多个特征来捕获标签依赖,这些特征的处理效果将对模型的分​​类结果造成很大的影响。

近年来,神经网络在自然语言处理领域取得了巨大的成功。神经网络方法可以根据训练样本进行特征筛选、提取,因此现有的文本分类模型多基于神经网络实现。卷积神经网络(convolutional neural network, CNN)具有较强的表征学习能力,自Kim Y^[4]提出文本卷积神经网络(TextCNN)模型以来,卷积神经网络在文本分类任务中的应用愈加广泛。Conneau A等人^[5]提出了VDCNN模型,模型堆叠了高达29层的卷积层,并且在大部分文本分类数据集上证明了深度卷积网络可以有效地抽取文本特征。卷积神经网络具有很强的局部特征提取能力,但是文本中的词序依赖关系也很重要。因此,Sun X等人^[6]将TextCNN和长短期记忆(long short term memory, LSTM)网络进行结合,提出TC-LSTM模型,有效地改善了分类任务的精度。Lin J Y等人^[7]则将膨胀卷积和注意力机制进行结合,从而有效地提取单词的语义信息。随着语言模型的出现,多标签文本分类任务开始使用微调方式。ELMo^[8]、GPT^[9]、BERT^[10]、XLNet^[11]等预训练语言模型使用了大规模语料进行预训练,并且相较于之前的网络模型具有很强的特征抽取能力,因此极大地提升了多标签分类任务的性能。上述方

法聚焦于文本信息,如果采用这些方法进行司法要素抽取,模型易过度依赖数据样本,进而导致过拟合问题。

目前,与司法要素抽取任务相关的研究较少,与其任务场景相似的罪名预测和法条推荐已具有一定的研究方案。Luo B F等人^[12]通过提取案情文本特征对法条和罪名预测进行联合训练。Zhong H X等人^[13]在此基础上提出多任务拓扑依赖学习模型TOP JUDGE,对法律审判的多个子任务(如适用法律条款、指控、罚金、刑期)进行建模。陈文哲等人^[14]在犯罪行为链^[15]的研究基础上,结合裁判文书文本特征和犯罪行为序列特征,增强了法条推荐效果。然而,这些研究都在文本分类的框架下进行,对文本特征的依赖程度很高,存在泛化能力不足的问题。同时,文本分类框架下的方法没有充分利用标签信息,缺乏文本和标签之间的语义联系。

抽取式机器阅读理解任务通过捕获问题和段落之间的联系,推理出答案在段落中的位置。机器阅读理解模型包含很强的交互层,如R-NET^[16]、BiDAF^[17]、BERT^[10]等模型在内部利用注意力机制进行问句和段落的语义交互,从而在多数机器阅读理解任务上具有很好的表现。Levy O等人^[18]和McCann B等人^[19]将关系提取作为一项单轮问答任务。在此基础上,Li X Y等人^[20]将实体关系抽取任务转换成多轮机器阅读理解问答形式开展,利用问题查询向模型提供重要信息。刘奕洋等人^[21]将实体识别转化为问答任务,通过卷积神经网络捕捉文档级文本上下文信息,通过分类器实现答案预测。

BERT^[10]由Google团队于2018年提出,其凭借大规模语料的预训练和强大的语义特征提取能力,极大地提升了多项NLP任务的表现。BERT模型的嵌入层包含词、位置、句子类型3种特征,然后

经过多层双向Transformer^[22]层得到输入的特征表示。Transformer完全抛弃了CNN和循环神经网络(recurrent neural network, RNN)结构,采用自注意力网络实现了长距离依赖。结合司法要素抽取任务的现实需求,本文以BERT为主要模型结构,引入要素标签信息构造辅助问句作为模型输入,使用机器阅读理解框架完成司法要素抽取任务。

3 司法要素抽取方法

本文将司法要素抽取任务转换成阅读理解模式来解决,机器阅读理解模型输入包括问句和段落,因此首先需要针对司法要素标签进行问句构造;其次,采用BERT抽取式机器阅读理解模型对问句和裁判文书进行深度交互,得到答案位置的概率分布,从而判别句子中是否包含涉及的要素标签。

3.1 基于标签信息增强的辅助问句构造方法

由于裁判文书中存在的标签数量不确定,本文对每一个标签都进行问题构造,以此判定文本是否包含此要素。预定义问题模板为“句子中是否包含_element_、_knowledge_。是或者否?”其中,_element_为可替换标签占位符,不同的要素标签可被替换成相应的标签文本描述;_knowledge_为与要素相关的先验知

识;“是”和“否”为需要抽取的答案。3种案由的辅助问句构造样例见表1。对于每一个句子,使用所有标签的辅助问句进行多轮问答,以确定是否包含相应的要素标签。

在构造的辅助问句中,标签文本信息极为重要,为了增强模型的学习效果,在问句中要素文本描述的前后位置插入特殊标识符。对于BERT模型,可以使用预留的“[unused*]”tokens作为特殊标识符。将原始问句表示为:

$$Q=(t_1, \dots, e_1, \dots, e_k, \dots, t_n) \quad (1)$$

其中, t_1, \dots, t_n 表示除要素标签外的token, e_1, \dots, e_k 表示要素标签的token。在问句中添加特殊标识符后的问句 Q_+ 表示为:

$$Q_+=(t_1, \dots, u_1, e_1, \dots, e_k, u, \dots, t_n) \quad (2)$$

其中, u 为插入的特殊标识符。

3.2 基于BERT的司法要素抽取模型

将裁判文书句子表示为 $P=(p_1, p_2, \dots, p_m)$, 辅助问句表示为 $Q=(q_1, \dots, u, e_1, \dots, e_k, u, \dots, q_n)$, m, k, n 分别表示裁判文书、要素标签、辅助问句的token个数。本文将BERT作为模型主体,提取句子和问题编码特征。如图1所示,将 P 和 Q 与BERT既定标识符CLS、SEP进行拼接,作为模型的输入序列。其中CLS置于首位,用于标识整个输入的语义;SEP用于分割问句和段落的字符输入。在嵌入层,模型将输入字符序列映射为字符嵌入(token embedding)、位置嵌入(position embedding)、句子类型嵌入(segment embedding),并将3种嵌入表示相加得到

表1 3种案由的辅助问句构造样例

要素	先验知识	构造问句
不动产分割	不动产包括房屋、土地等	句子中是否包含不动产分割,不动产包括房屋、土地等。是或者否?
夫妻共同财产	夫妻共同财产包括夫妻关系存续期间的工资、生产投资收益等	句子中是否包含夫妻共同财产,包括夫妻关系存续期间的工资、生产投资收益等。是或者否?
特殊工种	特殊工种包括有害工种,如采煤工、锅炉工、下井挖掘等	句子中是否包含特殊工种,特殊工种包括有害工种,如采煤工、锅炉工、下井挖掘等。是或者否?

裁判文书句子和问句的特征表示 \mathbf{X} 。

BERT模型主要由多个Transformer的编码器堆叠而成，每一层的输入来自上一层的输出：

$$\mathbf{H}_1 = \text{Transformer}(\mathbf{X}) \quad (3)$$

$$\mathbf{H}_i = \text{Transformer}(h_{i-1}), \forall i \in [2, n] \quad (4)$$

其中， \mathbf{H}_i 表示第*i*层Transformer的输出，*n*表示编码器的Transformer层数。

通常机器阅读理解模型使用Softmax得到答案开始和结束位置的概率分布，这种模式主要针对答案为片段类型的数据集。在本文构造的辅助问句中，定义的候选答案（“是”和“否”）都只占一个位置，因此只需要预测一个位置概率分布。将BERT输出特征 \mathbf{H}_n 中间问句字符所在位置对应的向量表示 $\mathbf{H}_n^q = \{h_{q_1}, h_{q_2}, \dots, h_{q_n}\}$ 输入全连接层，得到位置概率分布 \mathbf{a} ：

$$\mathbf{a} = \text{Softmax}(\mathbf{H}_n^q \mathbf{W} + \mathbf{b}) \quad (5)$$

其中， \mathbf{W} 为全连接层的权重参数矩阵， \mathbf{b} 为偏置向量。

在推理阶段，通过比较辅助问句中“是”和“否”所在位置的概率大小，判定裁判文书句子是否包含问句指向的要素标签：

$$y = \begin{cases} 1, & a_{\text{yes}} > a_{\text{no}} \\ 0, & a_{\text{yes}} < a_{\text{no}} \end{cases} \quad (6)$$

4 实验及结果分析

4.1 数据集

本实验建立在2019中国“法研杯”司

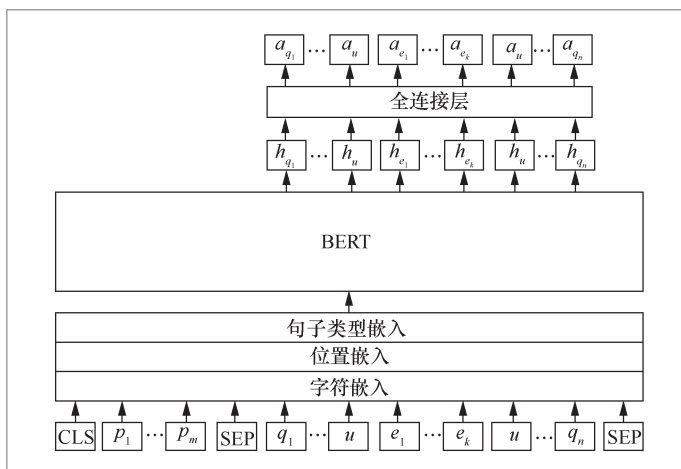


图1 基于BERT 阅读理解框架的司法要素抽取模型

法人工智能挑战赛(CAIL2019)要素抽取任务数据集上，实验分别使用赛题第一、第二阶段提供的训练集作为训练集和测试集。数据涉及婚姻家庭(divorce)、劳动争议(labor)和借款合同(loan)3种案由，每种案由都包含20类要素标签。数据集文档及句子分布统计具体见表2。

在数据集中，每个案由样本分布都具有长尾分布的特点。劳动争议案由的要素标签样本正例分布如图2所示，从图2可以看出，标签分布极度不均衡，这将对模型训练产生极大的影响。

4.2 评价指标

本文采用F1值的宏平均(macro average)和微平均(micro average)两种评价指标(即F1-macro和F1-micro)共同对模型进行评估。宏平均是每一个标签的性能指标的算术平均值，而微平均是每一

表2 数据集分布统计

对比项	婚姻家庭		劳动争议		借款合同	
	训练集	测试集	训练集	测试集	训练集	测试集
文档/个	869	400	685	151	485	150
句子/个	25 235	11 685	26 048	5 682	17 567	5 124

个样本示例的性能指标的算术平均值。因此,宏平均更加关注数据集中样本较少的类别,微平均更加关注数据集中样本较多的类别。在本文的实验数据集上,由于样本不均衡,模型的宏平均会低于微平均。

4.3 参数设置

实验对多标签分类方法和本文方法进行了对比,具体参数见表3。其中,分类方法涉及传统神经网络模型和语言模型,括号内的参数为传统神经网络模型的参数值。

4.4 实验结果与分析

为了全方位地对本文提出的模型进行评估,设计了3组对比实验。实验A使用本

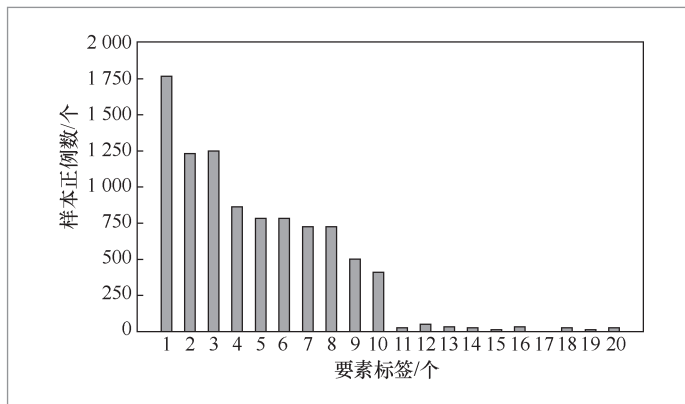


图2 劳动争议案由要素标签样本正例分布

表3 实验参数设置

参数名称	多标签分类方法	本文方法
学习率	2×10^{-5} (1×10^{-4})	2×10^{-5}
优化器	AdamW	AdamW
权重衰减指数	0.01	0.01
最大文本长度	128	128
最大问句长度	—	32
迭代轮次/轮	9(30)	3
随机种子	42	42
批次大小	32	32

文方法与多标签分类方法、文本匹配方法进行对比实验;实验B设计消融实验验证阅读理解框架和特殊标识符的效果;实验C对本文方法在不同案由上的标签进行迁移分析。

(1) 实验A: 模型对比实验

多标签分类模型包括以下几种。

- 循环递归神经网络 (TextRNN)^[23]: 基于传统的循环神经网络,针对句子内单词之间的序列依赖关系进行建模。

- 基于注意力机制的长短期记忆网络 (Att-LSTM)^[24]: 基于传统的LSTM模型结构,结合注意力机制对所有时序特征进行加权求和。

- TextCNN^[4]: 将文本当作一维图像,利用一维卷积神经网络来捕捉临近词之间的关联。

- BERT^[10]: 基于双向Transformer的预训练语言模型,采用微调方式进行多标签分类。

- ALBERT^[25]: 采用参数共享的方式解决BERT参数过多的问题,并改进BERT预训练方式。

- ERNIE^[26]: 在BERT的基础上对先验语义知识单元进行建模,增强了模型语义表示能力。

- RoBERTa-WWM-zh^[27]: 在BERT基础上,结合中文词特性,采用全词mask的方式在更多的语料上进行预训练。

为了验证辅助问句和机器阅读理解框架的有效性,将本文模型与文本匹配模型进行了对比。实验使用BERT句子匹配模型进行对比,其中模型输入构造为“[CLS]标签[SEP]文本[SEP]”。

模型对比实验结果见表4,其中F1值为宏平均和微平均的平均值,而F1-search为多标签分类方法经过阈值搜索后的F1值,JRC (judicial reading comprehension)指本文方法,MATCH指文本匹配方法,RBT指RoBERTa-WWM-zh。其中TextRNN、

Att-LSTM、TextCNN属于传统神经网络模型，ALBERT、BERT、ERNIE、RBT属于预训练语言模型。

实验结果显示，在多标签分类框架下，预训练语言模型在婚姻家庭和劳动争议案由上的表现大多远超过传统神经网络模型。这主要是因为预训练语言模型在大规模通用数据上进行过预训练，模型具备很好的先验知识，因此在下游任务上收敛更快、效果更好。但是由于数据集中存在一部分标签的样本可以根据局部关键词进行判定，而CNN具有很好的局部特征提取能力，非常适用于此类数据；而且预训练语言模型使用的通用数据与司法数据分布相差较大，大规模数据预训练的优势在司法数据集上也不如其他数据集明显，因此TextCNN模型在借款合同案由上取得了高于预训练模型的结果。

与多标签分类框架下的模型(TextRNN、Att-LSTM、TextCNN、ALBERT、BERT、ERNIE、RBT)相比，本文方法在3个案由上F1值分别至少提高2.7%、11.3%、5.6%，显著超过了多种多标签分类模型。为了进一步证明本文方法的优越性，对多标签分类方法增加了阈值搜索操作。实验表明，本文方法依然优于经过阈值搜索后的多标

签分类模型。文本匹配方法和本文方法都使用了标签信息和文本信息，但本文构造的辅助问句中除了标签信息，还包括法律先验知识，能够为模型提供更多信息。此外，语言模型在预训练阶段会学习到语言之间的相关性，为模型提供高质量、具有完备语义信息的辅助问句，从而提升模型的推理能力。因此，相对于文本匹配方法，本文方法在3种案由上F1值分别提升1.1%、4.2%、0.4%。

(2) 实验B：消融实验分析

本文使用RoBERTa-WWM-base预训练权重进行了消融实验。对分类方法(CLS)和机器阅读理解方法(MRC)进行了对比，并且验证标签、标识符、法律先验知识3种特征的效果，具体见表5。

从表5可以看出，融入标签信息的机器阅读理解方法(RBT-MRC)相比多标签分类方法(RBT-CLS)具有明显提升，尤其是F1-macro在3种案由上分别提升5.1%、21.7%、13.8%。这说明引入标签信息后，通过机器阅读理解模型对标签信息和文本信息进行语义交互，可以指导模型更好地进行推理。其次，模型增加了标识符后，在3种案由数据集上F1值都得到了提升，这进一步说明了标签信息在辅助问句中占据主导地位。将BERT最后

表4 模型对比实验结果

模型	婚姻家庭		劳动争议		借款合同	
	F1	F1-search	F1	F1-search	F1	F1-search
TextRNN	58.4%	61.5%	48.6%	50.1%	54.6%	55.1%
Att-LSTM	69.8%	71.0%	59.2%	61.2%	61.2%	62.2%
TextCNN	71.9%	74.0%	59.0%	60.7%	64.6%	65.3%
ALBERT	69.2%	73.3%	60.0%	60.6%	59.5%	59.7%
BERT	76.4%	78.1%	61.5%	63.3%	61.8%	64.7%
ERNIE	76.2%	78.1%	61.8%	62.3%	60.4%	60.8%
RBT	75.9%	78.0%	61.4%	62.3%	62.7%	63.8%
MATCH(RBT)	78.0%	—	68.9%	—	69.8%	—
JRC(RBT)	79.1%	—	73.1%	—	70.2%	—

一层Transformer的多头注意力权重进行可视化,如图3所示,在未增加标识符时,要素标签部分token的注意力主要集中在“补偿金”以及“除”,并且权重分布比较散乱。增加了标识符后,注意力便集中在“解除劳动合同”,这更符合人类判定的关注点。最后,在辅助问句中增加法律先验知识,进一步提升了司法要素抽取的性能,提升分值仅次于标签信息。这说明辅助问句中引入更多的先验信息能够使模型更好地拟合目标任务,但是该方式非常依赖引入先验知识的质量。在实验中,由于借款合同案由数据中添加的先验知识质量不如其他两种案由,因此提升并不明显。

(3) 实验C: 标签迁移分析

机器阅读理解模型通过构造的辅助问句可以指导模型根据不同的问句判定文本

中是否包含要素标签,因此本文方法在不同的标签之间具有较好的迁移性。此外,本文在构造辅助问句过程中引入了标签信息和人工构造的先验知识,也能够提升模型在标签迁移上的效果。为了证明本文方法的标签迁移性,选定使用其中一个案由训练集训练的模型,测试该模型在其他案由上的效果,具体结果见表6(多标签分类的训练方式需要固定标签,因此无法进行标签迁移)。

从表6可以看出,使用本文方法训练的模型可以在不同案由数据下进行标签迁移,而迁移的结果主要取决于训练和测试数据之间的相似性。因此,具有标签信息和先验知识的辅助问句可以帮助模型将从正样本多的标签数据中学习到的推理知识迁移至正样本少的标签上,从而提升模型在标签分布不均衡条件下的表现。从图4可以看出,在

表 5 消融实验结果

模型	特征			婚姻家庭		劳动争议		借款合同	
	标签	标识符	法律先验知识	F1-micro	F1-macro	F1-micro	F1-macro	F1-micro	F1-macro
RBT-CLS	—	—	—	84.2%	67.9%	80.4%	42.4%	80.1%	45.3%
	✓	—	—	84.6%	72.1%	81.9%	59.3%	81.0%	57.3%
RBT-MRC	✓	✓	—	84.9%	72.3%	81.9%	62.2%	81.1%	57.7%
	✓	✓	✓	85.1%	73.0%	82.0%	64.1%	81.4%	59.1%

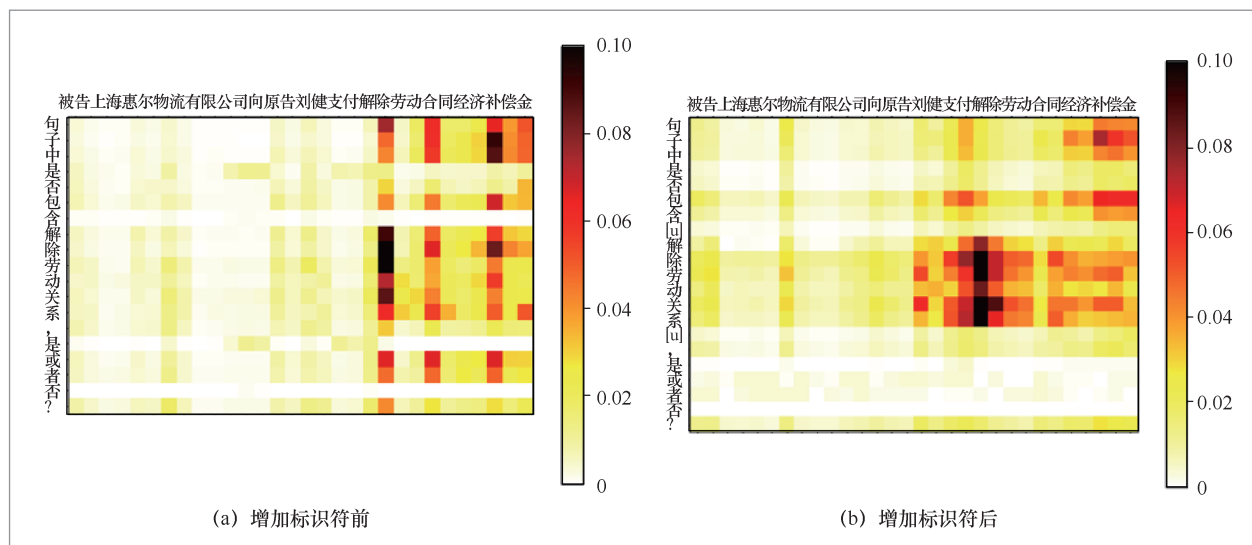


图 3 BERT 注意力权重可视化

劳动争议案由下,本文提出的方法在正样本越少的标签上相对于多标签分类方法提升效果越明显,这也说明了标签迁移性可以有效地解决样本分布不均衡的问题。

5 结束语

司法要素抽取任务通常采用多标签分类方法建模,模型只从案情文本的维度进行分析预测,没有利用要素标签的语义信息。并且,司法要素抽取数据存在类别分布不均衡的问题,导致多标签分类模型表现较差。针对这两个问题,本文提出基于BERT阅读理解框架的司法要素抽取方法,将司法要素抽取任务转换成抽取式机器阅读理解任务来解决。通过构造包含要素标签信息和法律先验知识的辅助问句,为模型提供推理线索。同时,利用BERT机器阅读理解模型实现问句和案情描述之间的深层次语义交互,充分利用了标签信息。此外,实验证明该模型具有良好的标签迁移性,可以解决数据分布不均衡的问题。

目前,本文方法没有考虑司法要素之间的依赖性。在未来工作中,笔者将探索如何利用机器阅读理解框架对标签依赖进行建模,并且将本文方法在更多案由上进行实践、验证,以推进司法智能化建设。

参考文献:

- [1] ELISSEEFF A, WESTON J. A kernel method for multi-labelled classification[C]//Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Cambridge: MIT Press, 2001: 681-687.
- [2] GHAMRAWI N, MCCALLUM A. Collective multi-label classification[C]//Proceedings of the 14th ACM International

表6 标签迁移实验结果

训练集	测试集	F1-micro	F1-macro
婚姻家庭	劳动争议	50.6%	30.0%
	借款合同	39.0%	26.8%
劳动争议	婚姻家庭	37.9%	32.3%
	借款合同	45.3%	28.4%
借款合同	婚姻家庭	25.2%	22.1%
	劳动争议	58.3%	34.3%

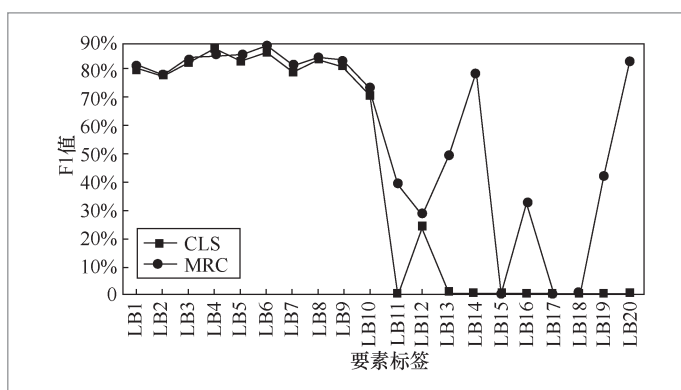


图4 RBT-CLS和RBT-MRC在劳动争议案由各标签上的性能

- Conference on Information and Knowledge Management. New York: ACM Press, 2005: 195-200.
- [3] LI C, WANG B, PAVLU V, et al. Conditional bernoulli mixtures for multi-label classification[C]//Proceedings of the 2016 International Conference on Machine Learning. [S.l.:s.n.], 2016: 2482-2491.
 - [4] KIM Y. Convolutional neural networks for sentence classification[J]. arXiv preprint, 2014, arXiv:1408.5882.
 - [5] CONNEAU A, SCHWENK H, BARRAULT L, et al. Very deep convolutional networks for text classification[J]. arXiv preprint, 2016, arXiv:1606.01781.
 - [6] SUN X, MA X H, NI Z W, et al. A new LSTM network model combining TextCNN[C]//Proceedings of the 2018 International Conference on Neural Information Processing. Heidelberg: Springer, 2018: 416-424.
 - [7] LIN J Y, SU Q, YANG P C, et al. Semantic-unit-based dilated convolution

- for multi-label text classification[C]// Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018.
- [8] PETERS M E, NEUMANN M, IYYER M, et al. Deep contextualized word representations[J]. arXiv preprint, 2018, arXiv:1802.05365.
- [9] RADFORD A, NARASIMHAN K, SALIMANS T, et al. Improving language understanding by generative pre-training[Z]. 2018.
- [10] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint, 2018, arXiv:1810.04805.
- [11] YANG Z L, DAI Z H, YANG Y M, et al. XLNet: generalized autoregressive pretraining for language understanding[J]. arXiv preprint, 2019, arXiv:1906.08237.
- [12] LUO B F, FENG Y S, XU J B, et al. Learning to predict charges for criminal cases with legal basis[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2017: 2727-2736.
- [13] ZHONG H X, GUO Z P, TU C C, et al. Legal judgment prediction via topological learning[C]//Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2018: 3540-3549.
- [14] 陈文哲, 秦永彬, 黄瑞章, 等. 基于犯罪行为序列的法律条文预测方法[J]. 计算机工程与应用, 2019, 55(22): 245-249, 264.
- CHEN W Z, QIN Y B, HUANG R Z, et al. Legal text prediction method based on criminal behavior sequence[J]. Computer Engineering and Applications, 2019, 55(22): 245-249, 264.
- [15] 秦永彬, 冯丽, 陈艳平, 等. “智慧法院”数据融合分析与集成应用[J]. 大数据, 2019, 5(3): 35-46.
- QIN Y B, FENG L, CHEN Y P, et al. “Intelligent Court” data fusion analysis and integrated application[J]. Big Data Research, 2019, 5(3): 35-46.
- [16] Natural Language Computing Group, Microsoft Research Asia. R-NET: machine reading comprehension with self-matching networks[R]. 2017.
- [17] SEO M, KEMBHAVI A, FARHADI A, et al. Bidirectional attention flow for machine comprehension[J]. arXiv preprint, 2016, arXiv:1611.01603.
- [18] LEVY O, SEO M, CHOI E, et al. Zero-shot relation extraction via reading comprehension[J]. arXiv preprint, 2017, arXiv:1706.04115.
- [19] MCCANN B, KESKAR N S, XIONG C M, et al. The natural language decathlon: multitask learning as question answering[J]. arXiv preprint, 2018, arXiv:1806.08730.
- [20] LI X Y, YIN F, SUN Z J, et al. Entity-relation extraction as multi-turn question answering[J]. arXiv preprint, 2019, arXiv:1905.05529.
- [21] 刘奕洋, 余正涛, 高盛祥, 等. 基于机器阅读理解的中文命名实体识别方法[J]. 模式识别与人工智能, 2020, 33(7): 653-659.
- LIU Y Y, YU Z T, GAO S X, et al. Chinese named entity recognition method based on machine reading comprehension[J]. Pattern Recognition and Artificial Intelligence, 2020, 33(7): 653-659.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]// Proceedings of the 31st Conference on Neural Information Processing Systems. [S.l.:s.n.], 2017: 5998-6008.
- [23] SOCHER R, PERELYGIN A, WU J, et al. Recursive deep models for semantic compositionality over a sentiment treebank[C]//Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2013: 1631-1642.
- [24] ZHOU P, SHI W, TIAN J, et al. Attention-based bidirectional long short-term memory networks for relation classification[C]// Proceedings of the 54th Annual Meeting

- of the Association for Computational Linguistics (Volume 2: Short Papers). Stroudsburg: Association for Computational Linguistics, 2016: 207–212.
- [25] LAN Z Z, CHEN M D, GOODMAN S, et al. ALBERT: a lite BERT for self-supervised learning of language representations[J]. arXiv preprint, 2019, arXiv:1909.11942.
- [26] SUN Y, WANG S H, LI Y K, et al. ERNIE: enhanced representation through knowledge integration[J]. arXiv preprint, 2019, arXiv:1904.09223.
- [27] CUI Y M, CHE W X, LIU T, et al. Pre-training with whole word masking for Chinese BERT[J]. arXiv preprint, 2019, arXiv:1906.08101.

作者简介



黄辉 (1994–), 男, 贵州大学计算机科学与技术学院硕士生, 主要研究方向为自然语言处理、智能问答。



秦永彬 (1980–), 男, 博士, 贵州大学计算机科学与技术学院教授、院长, 主要研究方向为大数据处理、云计算、文本挖掘。



陈艳平 (1980–), 男, 博士, 贵州大学计算机科学与技术学院副教授, 主要研究方向为人工智能、自然语言处理。



黄瑞章 (1979–), 女, 博士, 贵州大学计算机科学与技术学院副教授, 主要研究方向为信息检索、文本挖掘。

收稿日期: 2021-01-18

通信作者: 秦永彬, ybqin@foxmail.com

基金项目: 国家自然科学基金资助项目 (No.62066008); 贵州省科学技术基金重点项目 (No.[2020]1Z055)

Foundation Items: The National Natural Science Foundation of China (No.62066008), The Key Projects of Science and Technology of Guizhou Province (No.[2020]1Z055)