(19)国家知识产权局

(12)发明专利

(10)授权公告号 CN 115017917 B
(45)授权公告日 2022.10.28

(21)申请号 202210947007.9

(22)申请日 2022.08.09

(65)同一申请的已公布的文献号
申请公布号 CN 115017917 A

(43)申请公布日 2022.09.06

(73)专利权人 北京肇祺信息科技有限公司
地址 100020 北京市朝阳区东三环中路1号环球金融中心西塔802

(72)发明人 彭勃

(74)专利代理机构 北京高文律师事务所 11359
专利代理师 徐江华 李宝玉

(51)Int.Cl.
*G06F 40/30*(2020.01)
*G06F 40/284*(2020.01)
*G06K 9/62*(2022.01)
*G06N 3/04*(2006.01)
*G06N 3/08*(2006.01)

(56)对比文件
CN 113553856 A,2021.10.26
CN 113345574 A,2021.09.03
CN 113312916 A,2021.08.27
WO 2018113498 A1,2018.06.28
吕学强等.融合BERT 与标签语义注意力的文本多标签分类方法.《计算机应用》.2021,
孙弋等.基于BERT 和多头注意力的中文命名实体识别方法.《重庆邮电大学学报》.2021,

审查员 刘杉

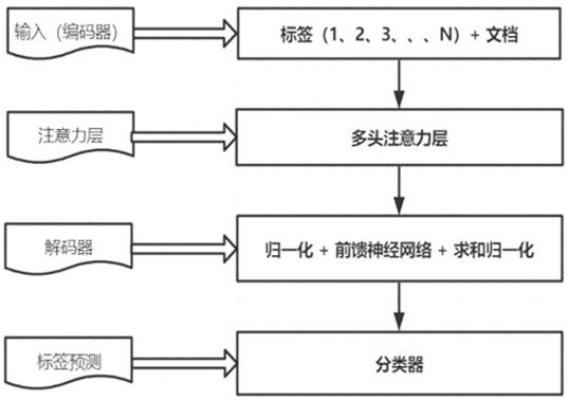
权利要求书2页 说明书9页 附图2页

(54)发明名称

基于多头注意力机制的裁判文书争议焦点识别方法

(57)摘要

本发明提供一种基于多头注意力机制的裁判文书争议焦点识别方法,包括以下步骤:S1:获取裁判文书数据集,对裁判文书数据集进行预处理;S2:裁判文书诉辩称文段语义表示学习;S3:争议焦点多标签语义表示学习;S4:融合多头注意力进行特征提取;S5:解码器解码;S6:争议焦点标签预测。本发明能够同步地对争议焦点多标签与裁判文书的关系,以及争议焦点多标签之间的关系进行建模,利用两者交互信息的同时避免误差累计。本发明使用共享语义参数的编码器提取争议焦点多标签和裁判文档的语义表示,以便减少其在建模语义相关性阶段的偏差问题。




CN 115017917 B

1.一种基于多头注意力机制的裁判文书争议焦点识别方法,包括以下步骤:

S1:获取裁判文书数据集,对裁判文书数据集进行预处理;

S2:裁判文书诉辩称文段语义表示学习,使用共享参数的BERT提取裁判文档和争议焦点多标签的语义表示,对诉辩段的文本进行向量化操作;

S3:争议焦点多标签语义表示学习,对争议焦点标签类型和诉辩称文段进行联合向量化操作;采用共享参数的BERT对标签和标签描述文本进行联合编码,利用 BERT 中的"[CLS]"符号为每个标签学习一个向量表示,得到多标签语义表示的运算如下:

$$A_k = \mathrm{BERT}(c_k)$$

$$A = [A_1, A_2, \ldots A_L]$$

$C_k$表示含有争议焦点的描述文本初始向量,$A_k$表示经过共享参数的BERT处理后的向量,A表示每个标签的语义表示,则 A 中的每一行为每个标签的语义表示;

S4:融合多头注意力进行特征提取,对争议焦点标签和裁判文档进行相关性操作;

S5:解码器解码,对经过特征提取后的多标签和文档进行输出层前的解码操作;解码器解码是对经过特征提取后的多标签特征进行输出层前的解码操作,由残差结构、前馈神经网络和层归一化组成,解码器解码运算如下:

$$C' = \mathrm{LN}(C)$$

$$C'' = \mathrm{FNN}(C')$$

$$M = \mathrm{LN}(C'' + C')$$

其中,C'表示对上文得到的多头注意力C进行层归一化操作,C''表示对C'进行基于前馈网络的非线性操作,LN为层归一化操作,FNN为前馈神经网络,最终得到的M为标签对应的文档表示, 得到C'的公式为层归一化操作,得到C''的公式为基于前馈网络的非线性操作,得到M的公式为层归一化操作;

S6:争议焦点标签预测。

2.根据权利要求1所述的基于多头注意力机制的裁判文书争议焦点识别方法,其特征在于:步骤S1中,裁判文书数据集D由N个裁判文档X和对应的争议焦点标签Y组成,L为标签总数,i表示序号,对应有:

$$D = \{(X_i, Y_i)\}_{i=1}^{N}$$ 。

3.根据权利要求1所述的基于多头注意力机制的裁判文书争议焦点识别方法,其特征在于:步骤S1中,预处理阶段包含对原始裁判文书诉称和辩称文段的抽取,具体为诉称文段取"原告诉称"至"被告辩称"之间的文段,辩称文段取"被告辩称"至"本院认为"之间文段。

4.根据权利要求1所述的基于多头注意力机制的裁判文书争议焦点识别方法,其特征在于:步骤S2中,裁判文书诉辩称文段语义表示学习,是通过对文书的诉称和辩称文段进行抽取,进而对此文段进行BERT向量化表示学习,使用共享参数的BERT提取裁判文档和争议焦点多标签的语义表示,以使两者处于同一语义空间中,对于第i篇裁判文档词元化后得到j个词元,将词元序列输入到 BERT 中进行编码,得到每个词元的语义H表示:

$$H = \mathrm{BERT}(w_1, w_2, \ldots, w_j)$$

$W_j$表示每篇文档中每个词的初始向量表示，BERT中隐藏层表示的维度为768，则H的每一行为每个词元的语义表示。

5.根据权利要求1所述的基于多头注意力机制的裁判文书争议焦点识别方法，其特征在于：步骤S3中，为使模型注意到标签中的层级信息，在上下级的争议焦点描述文本中间加入词元‘/’。

6.根据权利要求1所述的基于多头注意力机制的裁判文书争议焦点识别方法，其特征在于：步骤S4中，使用标签表示作为查询向量，标签表示和文档表示拼接后作为键向量和值向量，故多头注意力进行特征提取的运算如下：

$$\boldsymbol{K}=\mathrm{Concat}(\boldsymbol{A},\boldsymbol{H})$$

$$\boldsymbol{V}=\mathrm{Concat}(\boldsymbol{A},\boldsymbol{H})$$

$$\mathrm{Attention}(\boldsymbol{Q},\boldsymbol{K},\boldsymbol{V})=\mathrm{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^{T}}{\sqrt{d_k}})\boldsymbol{V}$$

$$\boldsymbol{head}_k=\mathrm{Attention}(\boldsymbol{Q}\boldsymbol{W}_k^{Q},\boldsymbol{K}\boldsymbol{W}_k^{K},\boldsymbol{V}\boldsymbol{W}_k^{V})$$

$$\boldsymbol{C}=\mathrm{Concat}(\boldsymbol{head}_1,\boldsymbol{head}_2,...,\boldsymbol{head}_h)\boldsymbol{W}^{O}$$

其中，W为可学习矩阵，K表示键向量，V表示值向量，A表示上文中每个标签的语义表示，H表示每个词元的语义，Q表示查询向量，h为注意力头数，k表示第k个注意力头，最终得到多头注意力C。

7.根据权利要求1所述的基于多头注意力机制的裁判文书争议焦点识别方法，其特征在于：步骤S6中，模型采取sigmoid函数进行预测，模型预测运算如下：

$$y_c=\sigma(\boldsymbol{M}_c\boldsymbol{W}_c^{y}+b_c)$$

其中，$y_c$表示第 c个类别的概率，$M_c$表示上文提到的标签对应的文档表示，$W_c$表示标签对应的可学习权重参数，y表示预测标签，$\boldsymbol{W}_c^{y}$表示第c个类别的权重参数，$b_c$表示可学习的偏置参数。

8.根据权利要求1所述的基于多头注意力机制的裁判文书争议焦点识别方法，其特征在于：步骤S1中，所述裁判文书争议焦点分为两大类别：费用项争议焦点和非费用项争议焦点。

# 基于多头注意力机制的裁判文书争议焦点识别方法

## 技术领域

[0001] 本发明涉及自然语言处理领域和法律行业裁判文书中争议焦点的智能识别技术领域，尤其是涉及一种基于多头注意力机制的裁判文书争议焦点识别方法。

## 背景技术

[0002] 随着人工智能和互联网的进步和司法程序的发展，司法信息呈现爆炸式增长。而如何从海量的司法文本中快速准确地挖掘出关键信息，成为了司法领域的关键问题之一。裁判文书是人民法院庭审过程中记载的案件诉辩双方观点、证词和结果等内容的总称，它是庭审中诉讼事件结果的载体，也是人民法院用于裁定和判定各当事人实体权利以及负担义务的凭证。裁判文书的重要性在于，它是整个诉讼程序的浓缩，是对于庭审过程最为客观、动态的记录；也是用于分析、排解矛盾纷争最为客观、真实的工具；同时，裁判文书更体现着庭审法官在该过程中对于自身审判权的运用方式。当前阶段，法院对于争议焦点的提取方法仍然停留在依靠法官人工阅读、整理、分析、归纳裁判文书中的双方陈词，使得这一步骤会耗费大量法官的时间精力资源。因此，通过人工智能算法自动提取争议焦点以便帮助法官整理争议焦点信息具有重大的意义。

[0003] 然而目前争议焦点提取的技术方法中，普遍存在争议焦点提取的精度低的缺点。故而如何提高争议焦点提取的精度成为目前技术人员需要探索的地方。本发明旨在提出一种精度较高的技术方法来提取争议焦点。

[0004] 在司法文档数据中，司法领域特有的争议焦点类型多标签分类与通用领域不同，采用通用的多标签分类技术提取效果不理想。文本分类是自然语言处理中重要且经典的问题。在传统的文本分类问题中，每个样本只有一个类别标签，并且各个类别标签之间相互独立，分类粒度比较粗略，称为单标签文本分类。随着文本信息日益丰富，分类粒度细化程度越来越高，一个样本与多个类别的标签相关，同时类别标签之间存在一定的依赖关系，称为多标签文本分类。比如一篇交通事故裁判书中可能存在“医疗费争议”、“误工费”等等争议焦点标签。多标签学习方法可以直观地反映出多义性对象所具有的多种语义信息。多标签文本分类方法已经逐渐取代单一标签文本分类方法，成为自然语言处理领域的一个研究课题，许多学者对此进行了广泛而深入的探索和研究。多标签文本分类方法主要分为两大类：传统机器学习方法和基于深度学习的方法。传统机器学习方法利用经典的机器学习算法来处理多标签问题，缺点在于需要手工提取复杂的特征以及准确率较低。基于深度学习的方法是利用各种深度神经网络模型来处理多标签文本分类问题，根据网络的结构将其分为基于卷积神经网络结构、基于循环神经网络结构和基于Transformer结构的多标签文本分类方法。由于基于深度学习的方法的准确率以及端到端的优势，此方法逐渐成为多标签分类的主流方法。

[0005] 在多标签文本分类任务上，目前的工作主要解决以下三个方面的问题：(1)探究如何从文本中捕捉文本的语义信息以便得到语义信息更丰富的文本表示，这是多标签文本分类的基本问题。(2)探索如何获取多标签的文本表示，即标签-文本关系的表示。文本本身是

一个复杂的语义集，导致文本的不同部分对于不同标签的判别的贡献存在差异。(3)探索如何利用标签之间的相关性，即标签-标签的关系的表示。标签与标签之间具有相关性，例如大部分多标签文本分类任务的标签之间有多层次结构。大部分相关工作在解决第一个方面问题的基础上，主要关注对后两者之一的探索，少量工作同时对两方面进行了探索。但这些模型在不同阶段建模两个相关性，造成了误差传递且没有利用到两者的交互信息。

[0006] 司法领域争议焦点多标签分类任务是指：给定一条法律文档，通过提取诉称和辩称文段进而判定存在争议焦点标签的文书包含的争议焦点的标签有哪些，即争议焦点标签的多标签分类任务。例如：一篇交通事故裁判书中可能存在事故责任认定争议、医疗费争议、误工费等等争议焦点标签。而争议焦点多标签分类任务需要做的是：识别出裁判文书中存在的上述争议焦点类型。

## 发明内容

[0007] 本发明提供了一种基于多头注意力机制的裁判文书争议焦点识别方法，解决裁判文书中争议焦点多标签分类精度较低的问题，其技术方案如下所述：

[0008] 一种基于多头注意力机制的裁判文书争议焦点识别方法，包括以下步骤：

[0009] S1：获取裁判文书数据集，对裁判文书数据集进行预处理；

[0010] S2：裁判文书诉辩称文段语义表示学习，对诉辩段的文本进行向量化操作；

[0011] S3：争议焦点多标签语义表示学习，对争议焦点标签类型和诉辩称文段进行联合向量化操作；

[0012] S4：融合多头注意力进行特征提取，对争议焦点标签和裁判文档进行相关性操作；

[0013] S5：解码器解码，对经过特征提取后的多标签和文档进行输出层前的解码操作；

[0014] S6：争议焦点标签预测。

[0015] 进一步的，步骤S1中，裁判文书数据集D由N个裁判文档i和对应的争议焦点标签Y组成，L为标签总数，对应有：

[0016] $$D=\{(X_i,\boldsymbol{Y}_i)\}_{i=1}^{N}$$

[0017] 进一步的，步骤S1中，预处理阶段包含对原始裁判文书诉称和辩称文段的抽取，具体为诉称文段取“原告诉称”至“被告辩称”之间的文段，辩称文段取“被告辩称”至“本院认为”之间文段。

[0018] 进一步的，步骤S2中，裁判文书诉辩称文段语义表示学习，是通过对文书的诉称和辩称文段进行抽取，进而对此文段进行BERT向量化表示学习，使用共享参数的BERT提取裁判文档和争议焦点多标签的语义表示，以使两者处于同一语义空间中，对于第i篇裁判文档词元化后得到j个词元，将词元序列输入到 BERT 中进行编码，得到每个词元的语义H表示：

[0019] $$\boldsymbol{H}=\mathrm{BERT}(w_1,w_2,...,w_j)$$

[0020] $W_j$表示每篇文档中每个词的初始向量表示，d为BERT中隐藏层表示的维度768，则H的每一行为每个词元的语义表示。

[0021] 进一步的，步骤S3中， 为使模型注意到标签中的层级信息，在上下级的争议焦点描述文本中间加入词元‘/’。

[0022] 进一步的，步骤S3中，采用共享参数的BERT对标签和标签描述文本进行联合编

码,,利用 BERT 中的“[CLS]”符号为每个标签学习一个向量表示,得到多标签语义表示的运算如下:

[0023] $$A_k = \mathrm{BERT}(c_k)$$
$$A = [A_1, A_2, \ldots A_L]$$

[0024] Ck表示含有争议焦点的描述文本初始向量,$A_k$表示经过共享参数的BERT处理后的向量,A表示每个标签的语义表示,则 A 中的每一行为每个标签的语义表示。

[0025] 进一步的,步骤S4中,使用标签表示作为查询向量,标签表示和文档表示拼接后作为键向量和值向量,故多头注意力进行特征提取的运算如下:

$$K = \mathrm{Concat}(A, H)$$
$$V = \mathrm{Concat}(A, H)$$

[0026] $$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}(\frac{QK^T}{\sqrt{d_k}})V$$
$$head_k = \mathrm{Attention}(QW_k^Q, KW_k^K, VW_k^V)$$
$$C = \mathrm{Concat}(head_1, head_2, \ldots, head_h)W^O$$

[0027] 其中,W为可学习矩阵,K表示键向量,V表示值向量,A表示上文中每个标签的语义表示,H表示每个词元的语义,Q表示查询向量,h为注意力头数,k表示第k个注意力头,最终得到C。

[0028] 进一步的,步骤S5中,解码器解码是对经过特征提取后的多标签特征进行输出层前的解码操作,由残差结构、前馈神经网络和层归一化组成,故解码器解码运算如下:

[0029] $$C' = \mathrm{LN}(C)$$

[0030] $$C'' = \mathrm{FNN}(C')$$

[0031] $$M = \mathrm{LN}(C'' + C')$$

[0032] 其中,C’表示对上文得到的多头注意力C进行层归一化操作,C’’表示对C’进行基于前馈网络的非线性操作,LN为层归一化操作,FNN为前馈神经网络,最终得到的M为标签对应的文档表示。得到C’的公式为层归一化操作,得到C’’的公式为基于前馈网络的非线性操作,得到M的公式为层归一化操作。

[0033] 进一步的,步骤S6中,模型采取sigmoid函数进行预测,模型预测运算如下:

[0034] $$y_c = \sigma(M_c W_c^y + b_c)$$

[0035] 其中,$y_c$表示第 c个类别的概率,$M_c$表示上文提到的标签对应的文档表示,$W_c$表示标签对应的可学习权重参数,y表示预测标签,$W_c^y$表示第c个类别的权重参数,$b_c$表示可学习的偏置参数。

[0036] 进一步的,步骤S1中,所述裁判文书争议焦点分为两大类别:费用项争议焦点和非费用项争议焦点。

[0037] 所述基于多头注意力机制的裁判文书争议焦点识别方法,能够同步地对争议焦点多标签与裁判文书的关系,以及争议焦点多标签之间的关系进行建模,利用两者交互信息的同时避免误差累计。本发明使用共享语义参数的编码器提取争议焦点多标签和裁判文档

的语义表示,以便减少其在建模语义相关性阶段的偏差问题。

## 附图说明

[0038] 图1是所述基于多头注意力机制的裁判文书争议焦点识别方法的模型结构图;

[0039] 图2是所述基于多头注意力机制的裁判文书争议焦点识别方法的流程示意图;

[0040] 图3是裁判文书争议焦点详细类别图。

## 具体实施方式

[0041] 在对本发明的任意实施例进行详细的描述之前,应该理解本发明的应用不局限于下面的说明或附图中所示的结构的细节。本发明可采用其它的实施例,并且可以以各种方式被实施或被执行。基于本发明中的实施例,本领域普通技术人员在没有做出创造性改进前提下所获得的所有其它实施例,均属于本发明保护的范围。

[0042] 以下通过特定的具体实例说明本发明的实施方式,本领域技术人员可由本说明书所揭露的内容轻易地了解本发明的其他优点与功效。本发明还可以通过另外不同的具体实施方式加以实施或应用,本说明书中的各项细节也可以基于不同观点与应用,在没有背离本发明的精神下进行各种修饰或改变。需要说明的是,以下实施例中所提供的图示仅以示意方式说明本发明的基本构想,在不冲突的情况下,以下实施例及实施例中的特征可以相互组合。

[0043] 其中,附图仅用于示例性说明,表示的仅是示意图,而非实物图,不能理解为对本发明的限制;为了更好地说明本发明的实施例,附图某些部件会有省略、放大或缩小,并不代表实际产品的尺寸;对本领域技术人员来说,附图中某些公知结构及其说明可能省略是可以理解的。

[0044] 本发明实施例的附图中相同或相似的标号对应相同或相似的部件。在本发明的描述中,需要理解的是,若有术语“上”、“下”、“左”、“右”、“前”、“后”等指示的方位或位置关系为基于附图所示的方位或位置关系,仅是为了便于描述本发明和简化描述,而不是指示或暗示所指的装置或元件必须具有特定的方位、以特定的方位构造和操作,因此附图中描述位置关系的用语仅用于示例性说明,不能理解为对本发明的限制,对于本领域的普通技术人员而言,可以根据具体情况理解上述术语的具体含义。

[0045] 首先需要说明的是,本发明的模型使用bert-large-uncased预训练模型作为编码器,bert-large-uncased预训练模型是一种bert预训练模型的变体,隐藏层维度为768,注意力头数为16,编码器外的参数全部随机初始化。本发明使用Adam优化器进行训练,初始学习率为0.00005,批大小为32。

[0046] 如图2所示,所述基于多头注意力机制的裁判文书争议焦点识别方法,包括以下步骤:

[0047] S1:获取裁判文书数据集,对裁判文书数据集进行预处理

[0048] 所述裁判文书数据集D由N个裁判文档i和对应的争议焦点标签Y组成,L为标签总数。争议焦点进行多标签文本分类的目标就是学习一个从裁判文书到对应的多个标签的映射:

[0049] $$D=\{(X_i,\boldsymbol{Y}_i)\}_{i=1}^{N}$$

[0050] 其中,i表示序号,无意义。

[0051] 获取裁判文本数据集D,需要构建裁判文本的训练数据集合,裁判文本数据集D来自专家的标注数据以及通过数据增强来获得,来自专家的标注数据是指业务专家标注的包含争议焦点标签的裁判文书,通过数据增强是指对标注数据的非关键词的随机删除、随机插入、同义词替换等方式的增强。

[0052] 然后对裁判文本数据集D进行预处理,预处理阶段包含对原始裁判文书诉称和辩称文段的抽取。具体为:诉称文段取“原告诉称”至“被告辩称”之间的文段,辩称文段取“被告辩称”至“本院认为”之间文段。

[0053] 所述裁判文本数据集D,对应的前期数据准备阶段包括裁判文书的数据,以及争议焦点的分类。本发明提出的裁判文书争议焦点共为两大类别:费用项争议焦点和非费用项争议焦点。例如:非费用项争议焦点包括:责任认定争议争议焦点、免责未告知争议焦点、自身原发性疾病争议焦点等等;费用项争议焦点包括:非医保扣减争议焦点、二次手术费争议焦点、误工费标准问题争议焦点等等。本发明的裁判文本数据集D为每种争议焦点业务专家标注的约10条,共有1910条标注数据。通过标注数据的非关键词的随机删除、随机插入、同义词替换等方式对数据进行增强至每种争议焦点100条数据,至此,数据集增强到19100条数据。其中,训练集和测试集按照7:3的比例划分。

[0054] S2:裁判文书诉辩称文段语义表示学习,目的在于对诉辩段的文本进行向量化操作;

[0055] 裁判文书诉辩称文段语义表示学习指的是通过对文书的诉称和辩称文段进行抽取,进而对此文段进行BERT向量化表示学习。具体来说,BERT通过掩码语言模型和上下句预测任务在大规模无监督语料上进行预训练,以此学习到文本的语义信息和语法信息,使用BERT作为编码器,可以获取到通过海量语料训练得到的包含丰富语义信息的表示。本发明模型使用共享参数的BERT提取裁判文书争议焦点多标签的语义表示,以使两者处于同一语义空间中。具体来讲,对于第i篇裁判文档词元化后得到j个词元,将词元序列输入到 BERT中进行编码,得到每个词元的语义H表示:

[0056] $$\boldsymbol{H}=\mathrm{BERT}(w_1,w_2,...,w_j)$$

[0057] $W_j$表示每篇文档中每个词的初始向量表示,d为BERT中隐藏层表示的维度768,则H的每一行为每个词元的语义表示。至此,输入的裁判文书诉辩称文段语义表示学习完成。

[0058] 所述共享参数的Bert指的是共享attentions和FFN层的参数,进而降低参数总量,所述共享参数分为三种形式,只是共享attentions、只是共享FFN、全部共享。共享的意思就是这部分结构只使用同样的参数,在训练的时候只需要训练这一部分的参数就可以。

[0059] S3:争议焦点多标签语义表示学习,目的在于对争议焦点标签类型和诉辩称文段进行联合向量化操作;

[0060] 争议焦点多标签语义表示学习指的是通过对每个裁判文书的争议焦点多标签的描述文本和争议焦点标签本身进行联合向量化操作。具体来说,对于L个争议焦点类别标签,每个争议焦点标签都有文本作为标签的描述,大部分工作都没有利用到这部分信息。在现实场景中‘/’常常作为具有上下位关系文本的分隔符,为了使模型注意到标签中的层级

信息，在上下级的争议焦点描述文本中间加入词元‘/’，并且把争议焦点描述文本处理后的内容记为$C_k$。

[0061] 利用争议焦点标签描述文本将标签文本与裁判文档同时输入到 BERT 中进行编码。这种方法有以下问题：(1)由于BERT是基于自注意力机制的，与文档同时输入到BERT时会互相干扰，难以利用预训练模型学习到的语法语义知识；(2)争议焦点多标签文本的标签类别数量通常较大，而包括BERT在内的大部分预训练模型有输入长度限制。这会导致输入标签文本后，文档文本不能全部输入到预训练模型中，损失很多信息。

[0062] 所以采用共享参数的BERT预训练模型对标签和标签描述文本进行联合编码，既可以充分利用预训练模型学习到的知识，又不受编码长度限制。这里利用 BERT 中的“[CLS]”符号为每个标签学习一个向量表示。得到多标签语义表示的运算如下：

[0063]
$$\boldsymbol{A}_k = \mathrm{BERT}(c_k)$$
$$\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots \boldsymbol{A}_L]$$

[0064] $C_k$表示含有争议焦点的描述文本初始向量，$A_k$表示经过共享参数的BERT处理后的向量，A表示每个标签的语义表示。得到Ak的公式为采用BERT对文本进行表示学习。得到A的公式为每个标签的语义表示。至此，争议焦点多标签语义表示学习完成。

[0065] S4：融合多头注意力进行特征提取，目的是对争议焦点标签和裁判文档进行相关性操作；

[0066] 融合多头注意力进行特征提取指的是对争议焦点标签和裁判文档进行相关性操作。具体来说，每个争议焦点标签关注的文档内容是不同的，可以通过文档和标签的语义相关性对其进行建模，同时标签的描述文本也蕴含了丰富的标签间关系。这一阶段使用多头注意力机制同时建模文档与标签的相关性和标签与标签的相关性。为了在得到标签特定的文档表示的同时获取到标签间关系，使用标签表示作为查询向量，标签表示和文档表示拼接后作为键向量和值向量。多头注意力进行特征提取的运算如下：

[0067]
$$\boldsymbol{K} = \mathrm{Concat}(\boldsymbol{A}, \boldsymbol{H})$$
$$\boldsymbol{V} = \mathrm{Concat}(\boldsymbol{A}, \boldsymbol{H})$$
$$\mathrm{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \mathrm{softmax}(\frac{\boldsymbol{Q}\boldsymbol{K}^T}{\sqrt{d_k}})\boldsymbol{V}$$
$$\boldsymbol{head}_k = \mathrm{Attention}(\boldsymbol{Q}\boldsymbol{W}_k^Q, \boldsymbol{K}\boldsymbol{W}_k^K, \boldsymbol{V}\boldsymbol{W}_k^V)$$
$$\boldsymbol{C} = \mathrm{Concat}(\boldsymbol{head}_1, \boldsymbol{head}_2, \ldots, \boldsymbol{head}_h)\boldsymbol{W}^O$$

[0068] 其中，W为可学习矩阵，K表示键向量，V表示值向量，A表示上文中每个标签的语义表示，H表示每个词元的语义，Q表示查询向量，h为注意力头数，k表示第k个注意力头，最终得到多头注意力C。得到K的公式为A向量和H向量的拼接操作，得到V的公式为A向量和H向量的拼接操作，得到Attention(Q,K,V)的公式为注意力分布计算操作，得到head的公式为计算输入信息的加权平均操作，得到C的公式为多头注意力拼接操作。至此，融合多头注意力对标签和文档进行特征提取完成。

[0069] 其中，后三行公式是经典的构造attention模型的计算公式，分别是注意力分布计算、计算输入信息的加权平均、多头注意力的拼接操作，第一行和第二行的键向量和值向量

来自于拼接向量A和向量H后的结果。

[0070] S5:解码器解码,目的在于对经过特征提取后的多标签和文档进行输出层前的解码操作;

[0071] 解码器解码指的是对经过特征提取后的多标签和文档进行输出层前的解码操作。具体来说,解码器受 transformer 结构启发,由残差结构、前馈神经网络和层归一化组成。解码器解码运算如下:

[0072] $$C' = \mathrm{LN}(C)$$

[0073] $$C'' = \mathrm{FNN}(C')$$

[0074] $$M = \mathrm{LN}(C'' + C')$$

[0075] 其中,C'表示对上文得到的多头注意力C进行层归一化操作,C''表示对C'进行基于前馈网络的非线性操作。LN为层归一化操作,FNN为前馈神经网络,最终得到的M为标签对应的文档表示。得到C'的公式为层归一化操作,得到C''的公式为基于前馈网络的非线性操作,得到M的公式为层归一化操作。至此,解码完成。

[0076] S6:争议焦点标签预测,目的在于对含有争议焦点的文书进行预测;

[0077] 最后一步进行争议焦点的预测推理阶段,模型主要采取sigmoid函数进行预测。模型预测运算如下:

[0078] $$y_c = \sigma(M_c W_c^y + b_c)$$

[0079] 其中,$y_c$表示第 c个类别的概率,$M_c$表示上文提到的标签对应的文档表示,$W_c$表示标签对应的可学习权重参数,y表示预测标签,$W_c^y$表示第c个类别的权重参数,$b_c$表示可学习的偏置参数,得到$y_c$的公式为对模型参数进行非线性操作。至此,整个基于多头注意力的争议焦点多标签分类模型流程完成。

[0080] 结合图1可见,本发明的模型结构包含于流程结构的步骤中,具体步骤为:

[0081] 步骤S1,多标签和文档的输入,并通过编码器进行共享参数BERT的编码,目的在于对多标签和文档进行联合向量化操作。

[0082] 步骤S2,多头注意力层进行多标签和文档的特征提取,目的在于对多标签和文档进行语义特征理解。

[0083] 步骤S3,解码器解码,目的在于对文档进行标签的解码。

[0084] 步骤S4,标签预测,目的在于对文档中多标签的预测。

[0085] 以上所述的具体实施方式,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施方式而已,并不用于限定本发明的保护范围,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

[0086] 本发明能够同步地对争议焦点多标签与裁判文书的关系,以及争议焦点多标签之间的关系进行建模,利用两者交互信息的同时避免误差累计。本发明使用共享语义参数的编码器提取争议焦点多标签和裁判文档的语义表示,以便减少其在建模语义相关性阶段的偏差问题,可以提高争议焦点多标签的精度,可以更好地为法官以及律师等人员快速提取案件的争议焦点提供帮助。

[0087] 其中,争议焦点(191种)包括以下内容:

[0088] 1、诉讼费;2、鉴定费;3、保全费;4、营业损失;5、担保费;6、快递费;7、公告费;8、施救费;9、评估费;10、律师费;11、停车费;12、停运损失;13、免赔率及免赔额调整;14、责任比例调整;15、法律限制性行为;16、责任认定争议;17、他方责任;18、见义勇为;19、自身原发性疾病;20、虚假陈述;21、未及时报案;22、第三方责任;23、公车私用;24、身份转化;25、自甘风险;26、医疗事故;27、违法行为未劝阻;28、陈旧性损伤;29、一事不再理;30、不当得利;31、疏于管理;32、承揽;33、车辆所有方审查义务;34、非被保险人掌控的情形;35、雇佣;36、挂靠;37、保险期间;38、追偿;39、份额分摊;40、租赁;41、法律禁止行为禁止驾驶;42、职务行为;43、离开现场;44、法律禁止行为号牌问题;45、不利解释;46、受害人存在重大过失;47、拒赔争议;48、非交通事故;49、危险程度增加;50、法律禁止行为证件问题;51、出借;52、安全措施不当;53、故意行为;54、贬值损失;55、帮工;56、共同侵权;57、免责未告知;58、路权争议;59、逆行;60、法条适用争议;61、法律关系认定错误;62、诉讼超期;63、虚假证据材料;64、虚假鉴定;65、虚假事故;66、非法律规定的赔偿义务人;67、不具备民事行为能力;68、不具备保险利益;69、多因一果;70、乘客责任;71、连环事故;72、超标电动车;73、多方事故;74、好意同乘;75、紧急避险;76、责任无法认定;77、无接触事故;78、违规搭乘;79、违法改装车辆;80、实习期驾驶;81、违反特别约定;82、产品质量问题;83、非医保扣减;84、自费药扣减;85、外购药扣除;86、可替代药扣除;87、非事故治疗扣除;88、挂床治疗;89、自身疾病用药扣减;90、后续治疗费没有实际发生;91、伙食补助标准;92、伙食补助天数;93、是否满足护理条件;94、护理人数;95、护理人员年长不具备护理能力;96、护理人误工费标准;97、护工标准;98、护理天数;99、误工真实性-无业;100、误工费标准问题;101、误工期问题;102、营养费标准过高;103、营养期过长;104、残疾等级过高;105、申请重新鉴定;106、赔付整容费后是否赔付残疾赔偿金;107、赔付康复费后是否赔付残疾赔偿金;108、非事故造成死亡;109、死亡原因不明;110、精神抚慰金过高;111、处理丧葬人员误工费;112、处理丧葬人员住宿费;113、票据规范性;114、是否属于残疾辅助器具;115、鉴定标准使用错误;116、残疾赔偿金赔付后是否赔付康复费;117、是否需要赔付康复费;118、过度用药;119、护理标准问题;120、护理人数问题;121、护理依赖程度比例争议;122、护理依赖程度争议;123、残疾赔偿金赔付后是否赔付整容费;124、已获补偿;125、票据不规范;126、其它医疗费争议;127、无需后续治疗;128、无证据支持;129、二次手术费;130、复查费;131、标准问题;132、非事故相关治疗;133、其它后续治疗费争议;134、其它伙食补助费争议;135、是否需要院外护理;136、护理人不存在收入减少;137、护理人员与伤者关系不符合常识;138、重症期间家人护理;139、其它护理费争议;140、无劳动能力是否赔偿误工费;141、误工人员不存在收入减少;142、其它误工费争议;143、无鉴定结论或无医嘱;144、其它营养费争议;145、鉴定程序错误;146、鉴定时机不合理;147、经常居住地或户籍地认定不合理;148、赔偿标准适用争议;149、与伤情诊断明显不符;150、造作伤与诈病;151、伤残系数计算错误-残晋级规则错误;152、其它残疾赔偿金争议;153、赔偿标准适用争议;154、其它死亡赔偿金争议;155、低等级伤残是否需要给付被扶养人生活费;156、无劳动能力认定标准争议;157、夫妻间扶养义务;158、遗腹子的扶养认定;159、非婚生子的扶养关系确定;160、不存在抚养关系的继父母;161、过继子女的生父母;162、非法定扶养义务的其余亲属;163、扶养义务人认定争议;164、其它被扶养人生活费争议;165、致害人已追究刑事责任;166、事故责任无法认定;157、受害人存在重大过失;168、多方事故均存在事故责任;169、不构成伤残;170、交通意外事

故;171、其它精神损害赔偿金争议;172、其它交通费争议;173、是否需要残疾辅助器具;174、义肢;175、义齿;176、义眼;177、其它残疾辅助器具费争议;178、非功能康复;179、康复费标准过高;180、其它康复费争议;181、丧葬费计算标准;182、丧葬费重复计算;183、其它丧葬费争议;184、长期护理证据不足;185、长期护理费给付周期;186、其它长期护理费争议;187、可自行修复;188、高龄人士;189、存在自身基础性疾病;190、给付标准问题;191、其它整容费争议。

[0089] 本发明将司法领域争议焦点识别任务改为基于多头注意力的争议焦点多标签文本分类任务,提出了一种融合多头注意力机制,以多头注意力的形式同时对两种关系进行建模,避免了误差传递并使它们的信息可以同步交互。对于争议焦点多标签语义的表示提取中存在的问题,本文提出了通过使用BERT预训练语言模型作为编码器,使争议焦点多标签的语义表示和裁判文书的语义表示处于同一语义空间且两者互不扰,充分利用模型学习的语义信息。本发明的主要有益效果如下:(1)创新性地提出了一种提取司法领域裁判文书争议焦点多标签的技术方案。(2)提出了一种可以提升司法领域争议焦点多标签分类精度更高的技术方案。(3)从模型结构层面,提出了融合多头注意力机制同时建模争议焦点多标签-裁判文书关系和争议焦点多标签-争议焦点多标签关系,使模型学习到两者的交互信息并避免误差传递。(4)从模型结构层面,提出了一种共享语义方法提取争议焦点多标签和裁判文书的语义表示,使其在同一语义空间并避免了在编码阶段互相干扰。

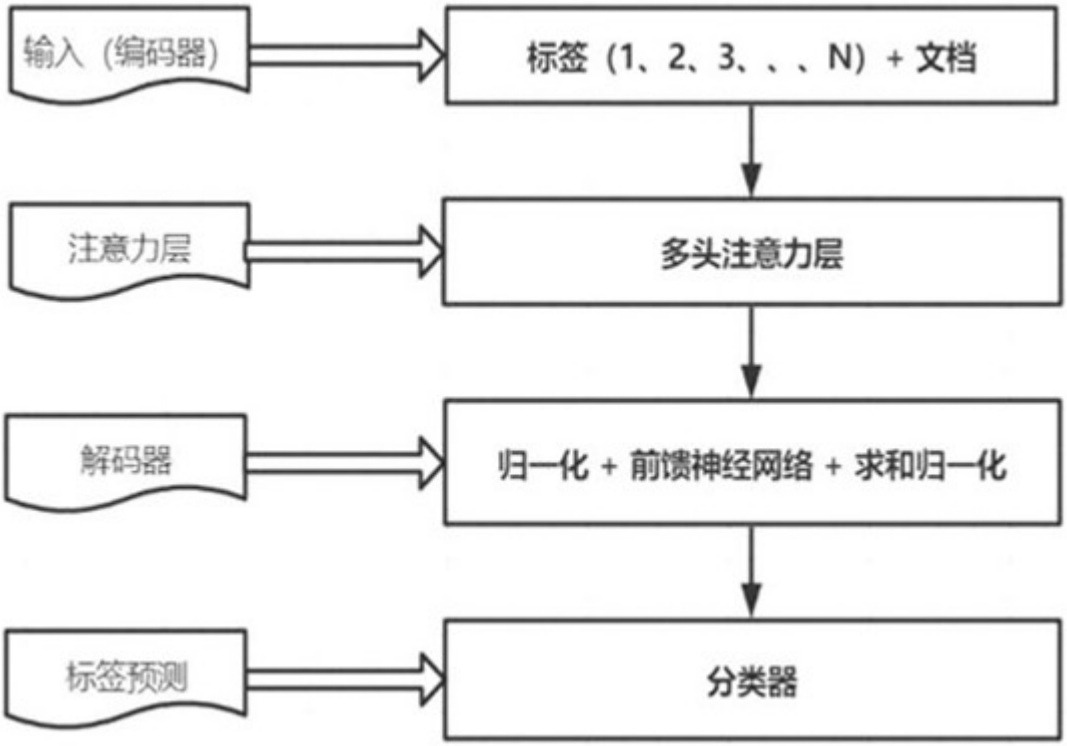



图1

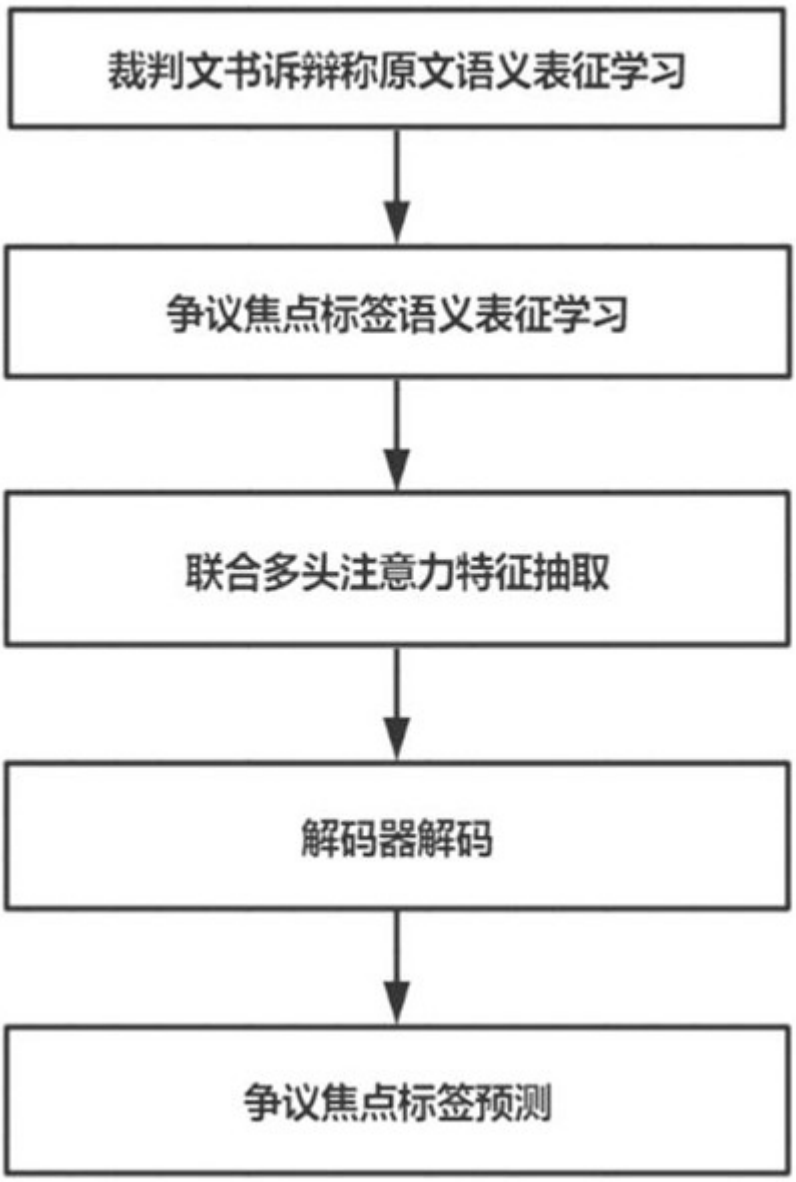

裁判文书诉辩称原文语义表征学习
争议焦点标签语义表征学习
联合多头注意力特征抽取
解码器解码
争议焦点标签预测

图2


裁判文书争议焦点
非费用项争议焦点标签数：82
费用项争议焦点标签数：109
平均每条数据标签数：2.19
平均每条数据标签数：2.71

图3