

文章编号: 2095-2163(2021)08-0042-05

中图分类号: TP183

文献标志码: A

基于混合特征提取的判决预测模型

刘璐瑶, 李实

(东北林业大学 信息与计算机工程学院, 哈尔滨 150006)

摘要: 法律判决预测指的是在给定案情描述的情况下,对案件的罪名和刑期进行预测。当前罪名预测主要使用深度神经网络模型,刑期预测主要使用混合深度神经网络模型。现有研究只关注局部特征或全局特征,没有考虑到将二者结合。因此本文使用双向门控循环单元提取上下文特征,并结合注意力机制学习文本中词的重要性,使用胶囊网络克服卷积神经网络丢失空间信息的缺点,学习文本局部与全局之间的关系特征。同时由于刑期分类时分类粒度较大,导致辅助决策结果不够理想。为实现更加理想的分类效果,将刑期按年进行更加细粒度的分类,分为28类。实验结果表明,该混合模型比基线模型效果更好,在罪名预测和28类刑期预测准确率分别为98.88%和74.32%。

关键词: 罪名预测; 刑期预测; 双向门控循环单元; 注意力机制; 胶囊网络

Rendering judgment predictions with the hybrid feature extraction learning model

LIU Luyao, LI Shi

(College of Information and Computer Engineering, Northeast Forestry University, Harbin 150006, China)

[Abstract] Legal judgment prediction refers to predicting accusations and prison terms of criminal cases on the basis of case descriptions. At present, accusation prediction mainly uses deep neural network models to extract features, and prison term prediction is grounded on hybrid deep neural network models. Existing works either focus on global or local features, not considering their combination. Therefore, this paper proposes Bidirectional Gated Recurrent Unit (BiGRU) to extract context features and merge attention mechanism (Attention) to learn the importance of words in text. Then it combines with CapsNet, which can overcome the disadvantage that Convolution Neural Network (CNN) will lose spatial information. The combination makes it possible to learn the relationship between local and global text features. At the same time, due to the large classification granularity of sentence classification, the result of auxiliary decision is not ideal. In order to achieve a more ideal classification effect, the sentence is classified into 28 categories by year. The experimental results show that the mixed model is better than the baseline model, and the accuracy of the prediction is 98.88% and 74.32% respectively.

[Key words] accusation prediction; prison term prediction; BiGRU; Attention; CapsNet;

0 引言

在传统的司法领域中,案件判决依赖于法官、律师等法律专业人士的专业解答和辩论流程。对于普通人而言,复杂的法律条文构成了专业壁垒,对于案件结果的预判也与专业人士存在较大差距^[1]。对于专业人士而言,大部分案件都属于常见案件,预测过程较为简单。如果能用技术手段学习到这些案件的共性,让量刑过程实现自动化或者半自动化,辅助法官的决策过程,一方面能将司法工作者从琐碎的事务中解放出来,另一方面也有利于消除人的主观因素的影响^[2],实现同案同判。

当今时代,随着法律不断完善,人民的法律意

识也不断提高,与此同时历史案件在不断累积,新案件也在不断的增加^[3],司法领域的各种公开案件量已足够满足深度学习需求,使得用深度学习技术进行司法领域相关问题的研究成为可能。

国外的研究者已经开展了大量的关于人工智能在司法领域各个方向的研究。Vlek等通过贝叶斯网络对案件的现有证据进行建模、分析与推理,在刑事案件的审判中,能够更好的通过模型从文本中提取出当前案件的场景描述,并能够直接的向法官或者陪审团提供对于现有证据的分析与结果展示^[4]; Ashley等人通过建立案件数据库,从已判决案件的文本描述中提取信息,并应用这些信息使用决策树算法完成自动预测新案件的结果^[5]。

基金项目: 中央高校基本科研业务费资助项目(2572019BH03)。

作者简介: 刘璐瑶(2000-),女,本科生,主要研究方向:自然语言处理;李实(1975-),女,博士,副教授,主要研究方向:自然语言处理、Web挖掘、商务智能。

通讯作者: 李实 Email:lishi@nefu.edu.cn

收稿日期: 2021-06-03

英美司法智能研究与中国在许多方面有所不同。首先, 英美量刑模式为普通法体系, 又称为判例法, 而中国的量刑模式是基于成文法; 其次, 英语与汉字的差异导致文本处理方法的不同; 另外, 汉语法律词汇与非法律词汇的差异很大, 一些流行的文本分类方法难以直接应用。基于此, 在关于机器学习的研究中, 高菲等在研究和借鉴英美法系的量刑模式后, 提出了改进中国量刑模式的新思路^[6]。通过对盗窃案件中年龄、主共犯、认罪态度等的量刑情节进行统计和数值化后, 使用支持向量机技术预测刑期, 使刑期结果精确到月份。本文是基于深度学习技术实现中文司法智能领域的罪名预测和刑期预测, 拟利用抽取式文本摘要, 对案情描述文本进行预处理, 同时提出基于 BiGRU-Attention-CapsNet 的文本分类模型进行罪名预测和 28 类刑期预测。

本研究的主要贡献如下:

(1) 提出了一种基于注意力机制的混合特征提取网络的文本分类模型, 解决全局和局部特征的不完全特征提取问题。

(2) 由于硬件和模型的限制, 每条案情描述文本长度有限, 通过对案情描述文本进行压缩, 在文本长度和文本信息量中找到平衡, 在指定长度中保留尽可能多的信息。

(3) 在刑期分类中, 实现了较小的分类粒度, 有更高的实际应用和参考价值。

1 相关工作

近年来判决预测的研究越来越受到人们的关注。目前在智能司法领域的研究方法主要分为 3 类: 数学统计方法、机器学习方法和深度学习方法。

早期的判决自动预测主张使用统计学方法, 试图分析大量历史案例找出共性规律并使用统计学模型模拟判决流程。典型工作如文献[7]中提出的量化分析法和如文献[8]提出的关联分析法。但此类方法仅在特定领域的数据上有效, 较难推广到一般性案件中。

随着机器学习的发展, 一些研究者开始从案件文书中提取特征, 利用机器学习模型来解决智能司法领域的问题。代表性工作如文献[9]中基于文书浅层文本特征的 K 近邻算法分类预测模型; 文献[10]中将 3 层神经网络与退火算法相结合, 通过定义和量化 28 种监禁情况的特征, 预测有期徒刑、死刑和无期徒刑; 文献[11]中利用多元伯努利模型进行分类的不均匀分布, 然后采用朴素贝叶斯算法进

行分类。该方法大大提高了分类的精度, 但提取出的浅层文本特征只能针对特定案例, 泛化能力不强。虽然机器学习模型极大地自动化了学习, 并且总是随着经验而改进, 但其需要手动对大量的特性进行标记, 这需要大量的时间和专业知识。

由于深度学习模型具有不需要标注大量特征的优点, 研究者们开始基于文本分类框架构建预测模型, 即以大量历史法律文书作为训练文本, 以罪名为类别标签, 训练深度学习分类模型^[12-15]。Ye 等人从事实描述中生成法院视图来解释判决预测^[16]。代表性工作如文献[17]和文献[18]中提出的基于深度神经网络罪名分类模型; 文献[19]在 2018 年使用“中国法研杯”司法人工智能挑战赛(CAIL-2018)的数据集, 提出了一种长文本分类的混合深度神经网络模型 HAC (hybrid attention and CNN model), 利用残差网络, 融合了改进的层次注意力网络和深度金字塔卷积神经网络, 使用分类方法将刑期分为 18 类对刑期进行预测; 文献[20]中, 在 CAIL-2018small 数据集上, 针对单人多罪名多法条的刑事案件对比了 3 种平均词向量模型, 并在多核 CNN 模型中加入不同层次的 Attention 机制, 融合 BERT 句向量特征, 提出了 BERT-ACNN 模型。

此类方法在预测效果上取得了一定的进步, 但对特征提取不够完整, 没有考虑局部特征和全局特征的融合。因此, 本文使用胶囊网络提取局部语义特征信息后, 再使用加入注意力机制的 BiGRU 提取全局语义特征信息, 最后将两个网络合并, 提取更加完整的信息来提升罪名预测和刑期预测的性能。

2 基于 BiGRU-Attention-CapsNet 的预测模型

2.1 数据预处理

数据预处理包括案情描述预处理和标签预处理。在案情描述预处理部分, 考虑到法律文书的关键句经常在句子结尾才出现, 而部分文本的长度超过了允许读入的最大长度, 本文先对数据进行抽取式摘要处理以确保关键句被读入。抽取式摘要采用 TextRank 算法抽取重要度最高的 10 个句子作为摘要, 之后主要是加入自定义词典对摘要进行分词去停、构建事实词典、将分词去停后的文本序列化、将序列处理为同一长度等预处理。在标签预处理部分: 对于罪名标签, 将 202 种罪名放到一个文本文件中, 再将其转化为数字编号; 而对于刑期标签, 考虑到以月为单位进行分类效果不佳, 所以以年为单位

将0~25年的刑期分为26类,无期徒刑和死刑各为一类,刑期一共分为28类。

2.2 BiGRU-Attention-CapsNet 模型研究

2.2.1 BiGRU-Attention-CapsNet 模型

本文搭建的基于 BiGRU-Attention-CapsNet 的文本分类模型结构如图1所示。文本分类模型主要包含输入层、Embedding 嵌入层、BiGRU-Attention-CapsNet 层、全连接层、输出层几个部分。

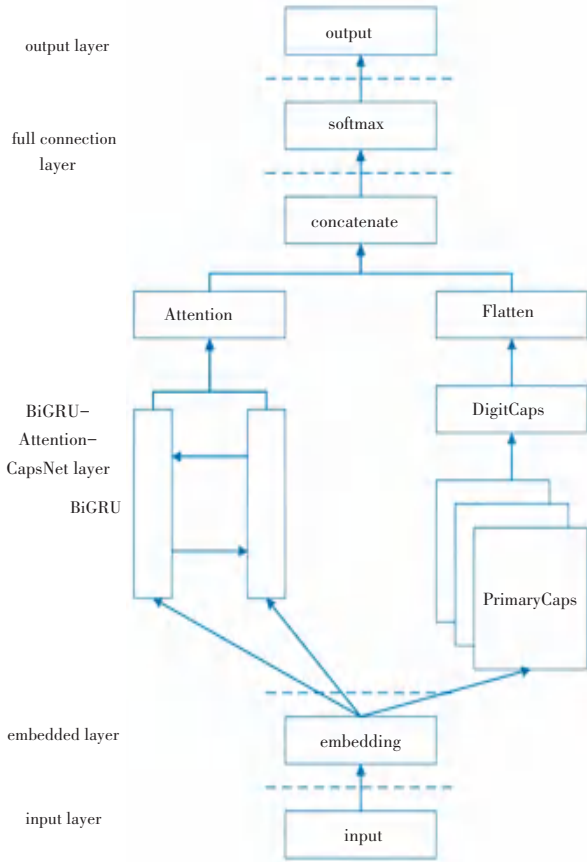


图1 BiGRU-Attention-CapsNet 模型

Fig. 1 BiGRU-Attention-CapsNet model

其中, BiGRU-Attention-CapsNet 层合并了 BiGRU-Attention 层提取的全局特征向量和 CapsNet 模块提取的局部特征向量。

2.2.2 BiGRU 层

门控循环单元 GRU 是对长短期记忆网络的一种改进,保留长期序列信息的同时通过门控机制优化了参数的规模^[21]。在 GRU 网络中信息只能单向传递,但词语可能与上下文的词语都有依赖关系,使用 BiGRU 融合上下文的语义信息,实现信息的双向传递,模型效果会更好。本文 BiGRU 层的目的是对输入文本词向量进行文本深层次特征的提取。式(1)、式(2)表示对输入词向量 \mathbf{x}_i 正向、反向编码。式(3)表示对 $\vec{\mathbf{h}}_i, \overleftarrow{\mathbf{h}}_i$ 进行向量拼接操作。

$$\vec{\mathbf{h}}_i = f_{GRU}(\mathbf{x}_i), i \in [1, n], \quad (1)$$

$$\overleftarrow{\mathbf{h}}_i = f_{GRU}(\mathbf{x}_i), i \in [n, 1], \quad (2)$$

$$\mathbf{h}_i = \text{concat}(\vec{\mathbf{h}}_i, \overleftarrow{\mathbf{h}}_i). \quad (3)$$

2.2.3 词级别注意力层

为捕获更准确的语义表达,本文在 BiGRU 层后引入注意力机制,对案情描述语句进行编码。不同的词对句子意思的表达所起的作用也有所不同,因此采用词级别 Attention 机制来提取对句子含义重要的词语。

词级别 Attention 机制可通过以下3个步骤实现:

(1)使用多层感知机获取 h_i 的隐藏表示 u_i , 如式(4)所示。

$$u_i = \tan h(w_i h_i + b_i), \quad (4)$$

式中, w_i 为模型权重, b_i 为偏置。

(2)计算 u_i 和上下文向量 \mathbf{u}_c 的相似性,并归一化得到度量词语重要性指标,如式(5)所示。

$$\alpha_i = \frac{\exp(u_i \mathbf{u}_c)}{\sum_t \exp(u_t \mathbf{u}_c) + \varepsilon}. \quad (5)$$

其中,上下文向量 \mathbf{u}_c 是对输入的一种语义表示,是在训练过程中随机初始化和共同学习的, ε 是一个很小的正数,是为了避免出现除零异常而添加的。

(3)计算词语的加权向量得到句子向量,如式(6)所示。

$$\mathbf{h}_i^{\text{att}} = \sum_i \alpha_i h_i. \quad (6)$$

2.2.4 胶囊网络模块

网络使用了 CapsNet 算法的最后一层 DigitCaps 层,将池化层用动态路由代替。该胶囊网络模块包括以下4层:

第一层:卷积层

通过不同的卷积核在句子的不同位置提取 N-gram 特征。其输入是文本词向量,卷积操作就是卷积核矩阵 \mathbf{M}^a 和对应输入层中一小块矩阵的点积相乘。卷积核通过权重共享的方式,按照步幅上下左右的在输入层滑动,提取特征,以此将输入层做特征映射 m_i^a 作为输出层。具体形式如下:

$$m_i^a = f(z_{i,i+k-1} \cdot \mathbf{M}_a + b_0), \quad (7)$$

其中, b_0 是偏置项, f 是非线性激活函数 $ReLU$ 。

第二层:主胶囊层(第一个胶囊层)

胶囊将卷积操作的标量输出替换为矢量输出,从而保留实例化参数。每一个胶囊 p_i 可由式(8)得到:

$$p_i = g(WM^i + b_1). \quad (8)$$

其中: $g()$ 表示非线性压缩函数; b_1 为胶囊的偏置项; W 是不同滑动窗口的共享滤波器; M^i 是 M^a 中第 i 行向量。

第三层: 卷积胶囊层

在这一层中, 每个胶囊仅与下面层中的一个局部区域相连。这些胶囊与转换矩阵相乘来计算子胶囊(低层胶囊)与父胶囊(高层胶囊)之间的关系, 然后根据协议路由计算出上层的父胶囊。

第四层: 全连接胶囊层

上一层的胶囊被展平成一个胶囊列表, 并送入全连接胶囊层。在全连接胶囊层中, 胶囊乘以变换矩阵, 然后按协议路由生成最终的胶囊及其对每个

类别的概率。

3 实验

3.1 数据集

本文使用的数据集为“中国法研杯”司法人工智能挑战赛(CAIL-2018)的数据集^[22], 数据集是来自中国裁判文书网公开的刑事法律文书。其中每份数据由法律文书中的案情描述和事实部分组成, 同时也包括每个案件所涉及的法条、被告人被判的罪名和刑期长短内容。数据集共包括 268 万条刑法法律文书, 共涉及 202 条罪名、183 条法条, 刑期长短包括 0~25 年, 无期徒刑和死刑。数据格式如图 2 所示。



图 2 数据示例

Fig. 2 Sample data

其中, fact 表示案情描述; meta 表示标签信息; punish_of_money 表示罚金(单位: 元); accusation 表示罪名; relevant_articles 为相关法条; term_of_imprisonment 为刑期。刑期分为: 是否死刑(death_penalty)、是否无期(life_imprisonment)、有期徒刑(imprisonment)等。

3.2 实验参数设置

参数设置上, 使用 word2vec 模型训练词向量, 维度为 100。由于本文训练样本的字符长度为 1 000 时, 样本覆盖全部语料集的 90% 以上, 故设定读取的序列长度为 1 000。对于长度不符的样例进行 padding 或 cut 处理。训练时部分参数见表 1。

表 1 部分参数设置

Tab. 1 Some parameters settings

参数/函数	实际意义	值
maxlen	句子最大长度	1 000
max_features	词汇表大小	20 000
embedding_dims	词向量维度	100
dropout	神经元失活概率	0.2
l2_penalty	学习率	0.000 1
activation	激活函数	ReLU
num_capsule	胶囊个数	10
dim_capsule	胶囊维度	16
batch_size	批处理大小	64
epochs	神经网络迭代次数	10

3.3 结果与分析

表 2 列出了 CAIL2018 数据集在 TextCNN, BiGRU-

Attention, CapsNet 和 BiGRU-Attention-CapsNet4 个模型上的罪名和刑期的预测结果。评估指标为测试集的准确率和损失值。

表 2 比较实验

Tab. 2 Model comparison experiments

模型	罪名		刑期(28 类)	
	accuracy	loss	accuracy	loss
TextCNN	0.953 5	0.182	0.734 1	0.729
BiGRU-Attention	0.982 1	0.088	0.736 6	0.731
CapsNet	0.980 2	0.081	0.741 6	0.695
BiGRU-Attention-CapsNet	0.988 8	0.057	0.743 2	0.701

TextCNN 是 2014 年由 Yoon Kim 提出的经典文本分类模型; BiGRU-Attention 融合上下文的语义信息, 实现信息双向传递, 注意力机制能对文本重要部分赋予更高的权重, 起到优化特征向量的目的; CapsNet 与 TextCNN 相比, 具有空间同变性, 将数个连续的神经元封装为一个胶囊输出, 保留了文本中词的本地顺序和词的语义表示。由表 2 可知, 在罪名预测和 28 类刑期预测上, 两模型的分类准确率均优于 TextCNN, 而融合了全局特征和局部特征的 BiGRU-Attention-CapsNet 是一种双向门循环单元注意机制混合胶囊神经网络模型, 其全局特征由 BiGRU-Attention 提取, 局部特征由 CapsNet 提取。由此表明, 由于合并特征提取, 提高了模型学习的语义信息, 该模型的分类精度高于两个独立模型分类精度。

本研究使用了与文献[1]相同的数据集和评价指标。由表3中数据表明,结合全局和局部特征的模型 BiGRU-Attention-CapsNet 具有最佳的精度。

表 3 不同模型实验结果

Tab. 3 Prediction results for different models

分类模型	罪名 accuracy	刑期(18类) score
CNN	96.73%	-
HAC	-	77.14
BiGRU-Attention-CapsNet	98.88%	82.76

对于刑期预测,文献[19]与本文数据集相同,把预测刑期与真实刑期的差异作为评估指标。假设第 i 起案件的真实刑期是 t_i ,而预测的结果是 \bar{t}_i 。其定义差异 d_i 如下:

$$d_i = |\log(t_i + 1) - \log(\bar{t}_i + 1)|, \quad (9)$$

然后,将得分函数 $f(v)$ 定义如下:

$$f(v) = \begin{cases} 1.0, & v \leq 0.2; \\ 0.8, & 0.2 < v \leq 0.4; \\ 0.6, & 0.4 < v \leq 0.6; \\ 0.4, & 0.6 < v \leq 0.8; \\ 0.2, & 0.8 < v \leq 1.0; \\ 0.0, & 1 < v. \end{cases} \quad (10)$$

最终得分如下:

$$Score = \sum_{i=1}^M \frac{f(d_i)}{M}. \quad (11)$$

文献[19]将刑期分为18类。为了便于比较,本研究也将刑期分为18类。表3显示,BiGRU-Attention-CapsNet 的得分为82.76分,比HAC的得分高5.62分。

4 结束语

本文在预处理部分对长文本采用 TextRank 算法抽取关键句作为模型输入,提出 BiGRU-Attention-CapsNet 模型,将全局特征和局部特征进行融合,在罪名预测和28类刑期预测的准确率上都有所提升。在未来的工作中,可以考虑引入外部法律知识库或融合更多知识模型。实际司法过程中,刑期还受到许多因素的影响,如被告和受害者的年龄,是否存在自首行为等。因此,在刑期预测中还可以添加命名实体识别和实体关系提取等技术,以提高预测的准确率。

参考文献

[1] 邓文超. 基于深度学习的司法智能研究[D]. 哈尔滨: 哈尔滨工业大学, 2017.
 [2] 闵丰锦. 同案不同判的理性思考[J]. 中国检察官, 2016(20): 64-68.
 [3] 靳高风, 朱双洋, 林晞楠. 中国犯罪形势分析与预测(2017-2018)[J]. 中国人民公安大学学报(社会科学版), 2018, 192(2): 34-43.

[4] VLEK C, PRAKKEN H, RENOUIJ S, et al. Constructing and understanding Bayesian networks for legal evidence with scenario schemes[C]//In Proceedings of the 15th International Conference on Artificial Intelligence and Law. 2015: 128-137.
 [5] ASHLEY K D, BRÜNINGHAUS S. Automatically classifying case texts and predicting outcomes. Artificial Intelligence and Law, 2009, 17(2): 125-165.
 [6] 高菲. 基于机器学习的计算机辅助量刑初探[D]. 上海: 华东政法学院, 2005.
 [7] KORT F. Predicting supreme court decisions mathematically: A quantitative analysis of the "right to counsel" Cases" [J]. American Political Science Review, 1957, 51(1): 1-12.
 [8] NAGEL S S. Applying correlation analysis to case pre-diction[J]. Texas Law Review, 1964, 42(7): 1006-1017.
 [9] LIU C L, CHANG C T, HO J H. Classification and clustering for case-based criminal summary judgments[C]//Proceedings of the 9th international conference on Artificial Intelligence and law, ACM, 2003: 252-261.
 [10] 邹晓玫, 修春波. 基于神经网络的刑事案件量刑决策系统[J]. 微计算机信息, 2008(3): 264-265.
 [11] 程春惠, 何钦铭. 面向不均衡类别朴素贝叶斯犯罪案件文本分类[J]. 计算机工程与应用, 2009, 45(35): 126-128, 131.
 [12] HU Zikun, LI Xiang, TU Cunchao, et al. Few-shot charge prediction with discriminative legal attributes[C]. Santa Fe: In Proceedings of COLING, 2018: 487-498.
 [13] JIANG Xin, YE Hai, LUO Zhunchen, et al. Interpretable rationale augmented charge prediction system[C]// In Proceedings of COLING: System Demonstrations, 2018: 146-151.
 [14] ZHONG Haoxi, GUO Zhipeng, TU Cunchao, et al. Legal judgment prediction via topological learning[C]. Brussels: In Proceedings of EMNLP, 2018: 3450-3459.
 [15] SHUBHASHRI G, UNNAMALAI N, KAMALIKA G. LAWBO: A Smart Lawyer Chatbot[C]. Goa, India: In Proceedings of The ACM India Joint International Conference on Data Science & Management of Data, 2018: 348-351.
 [16] YE Hai, JIANG Xin, LUO Zhunchen, et al. Interpretable charge predictions for criminal cases; Learning to generate court views from fact descriptions[C]. New Orleans: In Proceedings of NAACL, 2018: 1854-1864.
 [17] LUO B, FENG Y, XU J, et al., Learning to predict charges for criminal cases with legal basis[C]// Proceedings of EMNLP, 2017: 2727-2736.
 [18] HU Z, LI X, TU C, et al. Few-shot charge prediction with discriminative legal attributes[C]// Proceedings of the 27th International Conference on Computational Linguistics, 2018: 487-498.
 [19] 王文广, 陈运文, 蔡华, 等. 基于混合深度神经网络模型的司法文书智能化处理[J]. 清华大学学报(自然科学版), 2019, 59(7): 505-511.
 [20] 杨彬. 基于 BERT 词向量和 Attention-CNN 的智能司法研究[D]. 大连: 大连理工大学, 2019.
 [21] SCHUSTER M, PALIWAL K K. Bidirectional recurrent neural networks[J]. IEEE Transactions on Signal Processing, 1997, 45(11): 2673-2681.
 [22] XIAO C J, ZHONG H X, GUO Z P, et al. Cail2018: A large-scale legal dataset for judgment prediction[J/OL]. arXiv.https://arxiv.org/abs/1807.02478.