

基于BERT与改进BP神经网络的盗窃案刑期预测方法研究

郭彬彬^{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)

✉gbb96@qq.com



摘要: 为了应对智慧法院项目中刑期预测任务的实际需求, 提出了基于BERT与改进BP神经网络的刑期预测模型。以盗窃案为切入点, 剖析相关案情要素, 介绍刑期预测的整体框架和具体过程。基于大量真实案件数据, 结合法官的审理流程, 首先使用BERT识别裁判文书中的案情要素, 然后基于规则抽取对应的涉案金额, 最后使用改进的BP神经网络预测刑期, 并与传统模型对比。实验证明, 提出的模型刑期预测的平均误差小于2.5个月, 优于进行对比的传统模型。

关键词: 神经网络; 刑期预测; 盗窃案件; BERT

中图分类号: TP39 **文献标识码:** A

Research on Prediction Model of Sentence for Theft based on BERT and Improved BP Neural Network

● GUO Binbin^{1,2}

(1. College of Computer Science and Technology, Guizhou University, Guiyang 550025, China;

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

✉gbb96@qq.com

Abstract: In order to meet the actual needs of sentence prediction task in the smart court project, this paper proposes a sentence prediction model based on BERT (Bidirectional Encoder Representation from Transformers) and improved BP neural network. Starting the theft cases, relevant case elements are analyzed, and the overall framework and specific process of sentence prediction are introduced. Based on a large amount of real case data and the judge's trial process, BERT is used to identify the case elements in the judgment documents. Then the amount of money involved based on the rules is extracted. Finally, the improved BP neural network is used to predict the sentence period and the proposed model is compared with the traditional one. Experiments show that the average error of the proposed sentence prediction model is less than 2.5 months, which is better than the traditional model used for comparison.

Keywords: neural network; sentence prediction; theft case; BERT

1 引言(Introduction)

智慧法院是我国2016年提出的一项任务, 主要目的是提高案件受理、审判、执行、监督等环节的信息化水平, 推动司法信息公开, 促进司法公平正义。在当前的“智慧法院”建设中, 刑期预测是其中一项重要任务, 其核心目标是

通过分析裁判文书中案情描述部分, 自动预测出案件的判决刑期。该任务的实现结果可以作为补充来提高法官的审判效率, 且促进同案同判。

早在20世纪, 自动法律判决预测就已经引起研究者的关注, 这个时期主要是利用数学模型和统计学原理对法律案

件进行分析^[1-3]。这种方法对专业性知识要求较高，且效率低下。随着人工智能的发展，研究人员逐渐将AI应用到法律方面。KATZ^[4]使用随机森林(Random Forest)从案情描述中提取有效特征对美国最高法院的判决结果进行预测。王文广等^[5]将层次注意力网络^[6](Hybrid Attention Network, HAN)应用到刑期预测模型中，提出混合注意力和卷积神经网络模型(Hybrid Attention and CNN model, HAC)。谭红叶等^[7]采用多模型投票方法结合量刑属性进行刑期预测。ZHONG等^[8]认为法律审判的多个子任务之间存在依赖关系，提出了多任务拓扑依赖学习模型TOPJUDGE。YANG等^[9]基于多个子任务之间的拓扑依赖关系，引入词与词之间的组合语义关系，提出了多视角双向反馈网络MPBFN。以上这些模型均是基于分类的方法进行刑期预测，并没有给出最终的预测刑期；且不同类型的案件存在一定的差异，缺乏对某一类型案件的针对性。

本文获取了近十年某省的裁判文书，分析发现其中盗窃案件占比最大，达到27%以上，且盗窃案件审判流程和案情要素较为清晰，因此以盗窃案为切入点进行刑期预测实验。

真实的盗窃案审判中，法官的审判流程如图1所示。首先从案情描述中识别对应的案情要素和涉案金额，然后根据相关法条和涉案金额确定基准刑期，再根据案情要素对基准刑期进行修改，最终在修改后的基准刑期的一定幅度内确定宣告刑期。由于法官自身的影响，本模型的任务是预测修改后的基准刑期。根据法官的要求，刑期预测的误差在3个月以内是可以被认可的。

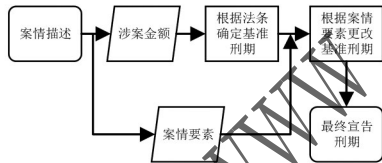


图1 法官审判流程

Fig.1 Trial process of the judge

本文借鉴了法官的真实审判流程，使用当下性能较强的预训练模型BERT作为要素识别模型，将要素识别作为多标签分类任务；然后使用基于规则的方法获取裁判文书中的涉案金额，将涉案金额和要素识别的结果作为BP神经网络的输入，最终输出预测刑期。经实验证明，本文所提出的盗窃案刑期预测方法平均误差达到2.5个月以内，已经具有实用价值。

2 数据集和案情要素(Data set and case elements)

2.1 数据集

本文数据集使用由贵州省高级人民法院提供的盗窃案件裁判文书，共包含8,000余篇带有标签的裁判文书。由于数据标注可能存在不规范或者标注错误的情况，因此对数据进行清洗。首先删除数据集中某些非初审案件的数据，原因在于非初审案件的裁判文书中一般不包含案情描述或者只有较少

的案情描述；然后去除某些标注数据明显错误的文本，通过规则识别案情描述中的涉案金额和刑期的比例，剔除某些明显错误的案例，如涉案金额1,000元、审判刑期10年；最终选取6,000篇裁判文书作为本次实验的数据集。

2.2 案情标签

案情标签是审理过程中的重要组成部分，主要是依据相关案情确定是否符合某一法条的某一部分，如未成年、累犯、扒窃等。对于某一案件，可能有多个案情标签同时存在，因此，案情要素的识别任务是一个多标签分类任务。

根据法官和相应法律文书的指导，本文选取了16个案情标签作为BP神经网络的输入，这些标签及标签相应的数量如表1所示。

表1 案情要素数量
Tab.1 Numbers of case elements

案情要素	解释	数量	百分比/%
hospital_loc	在医院盗窃	153	2.6
recover_goods	退还物品，赔偿	3,542	59
frank	坦白	3,699	61.7
in_home	入户	1,392	23.2
surrender	自首	688	11.5
young_to	未成年	858	14.3
criminal_record	具有前科	1,177	19.6
theft	扒窃	848	14.1
voluntary	当庭认罪	2,298	38.3
forgive	受到被害人原谅	605	10.1
offend	初犯，偶犯	454	7.6
again	累犯	2,083	34.7
principal	主犯	172	2.9
drug_mot	吸毒盗窃	152	2.5
accessory	从犯	165	2.8
disability	残疾人盗窃	69	1.2

3 模型构建(Model building)

模型构建分为两部分，分别是案情要素识别部分和刑期预测部分。案情要素识别的结果将会作为刑期预测模型的输入。

3.1 基于BERT的案情标签识别

2019年，谷歌提出了预训练模型BERT，在自然语言处理领域各个方面都具有非常好的效果。图2展示了基于BERT的案情要素识别模型图。

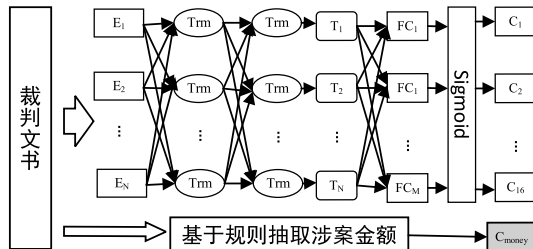


图2 基于BERT的案情要素识别模型

Fig.2 BERT-based case element identification model

本文所使用的BERT模型为谷歌官方发布的中文版，最大句子长度设为128。根据BERT模型的设置，句子长度超过该设置的部分将会被截断，同时该参数如果设置过大，将会影响训练效率及增加设备内存要求。因此，无法将整段裁判文书放入模型。

将裁判文书以句子为单位进行划分，使用句子中所包含的案情要素作为标签。在BERT模型的最后添加一个 768×16 的输出层，同时以Sigmoid作为激活函数，设置0.5为阈值，计算方法如公式(1)所示。多标签分类事实上可以看作多个二分类问题的集合，因此可以直接使用二分类的交叉熵损失函数，计算方法如公式(2)所示。

$$\sigma(Z_j) = \frac{1}{1 + e^{-Z_j}} \quad (1)$$

$$L = -\frac{1}{k} \sum_{i=1}^k y_i \log(\sigma(l_i)) + (1 - y_i) \log(1 - \sigma(l_i)) \quad (2)$$

3.2 基于自适应学习率的BP神经网络

BP神经网络是目前应用最为广泛的神经网络之一，其特点是可以很好地拟合一些非线性的函数。根据对盗窃案审判流程的分析，法官依据案情要素对基准刑期的修改过程就可以看作一个非线性函数的计算过程。因此，本文选取BP神经网络作为最终的刑期计算模型。

图3展示了具有三个隐藏层的简单BP神经网络。其中 H_1 即是隐藏层的计算过程，后面的两个隐藏层的计算方式相同，所使用的是Relu激活函数。最终的输出相当于是对输入的矩阵 X 的一个函数映射。

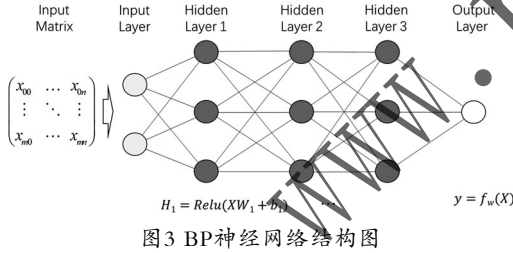


Fig.3 BP neural network structure diagram

在模型的输入中，将会增加一个维度，即通过规则提取出的金额，相当于一个特殊的要素。最终的输出节点只有一个，使其可以做回归分析。

本文所使用BP模型共有17个输入节点，1个输出节点，289个隐藏节点，以及三层隐藏层。通过输入层、隐藏层、输出层进行前向传播得到预测值，利用MSE损失函数来衡量真实值和预测值之间的误差。MSE损失函数的计算方法如公式(3)所示。利用梯度下降的方法来不断更新权重参数和偏置参数，不断缩小误差，最终获取和真实值最接近的预测结果。

$$MSE = \frac{1}{m} \sum_{i=1}^m (y_i - y_i^p)^2 \quad (3)$$

在梯度下降的过程中，学习率的设置会对结果和效率产

生一定的影响。学习率设置过大，有可能会错过极值点，且有一定可能无法收敛。学习率设置过小，将会消耗大量的时间来收敛，或者由于梯度的消失而无法收敛，即梯度趋近0，且固定学习率也有可能使函数陷入局部最小值。因此，本文使用自适应学习率方法，主要思路是使用前后两次迭代的误差来确定是增加还是减少学习率。同时根据差值的大小，确定学习率的变化幅度，最终使用差值的log函数值作为学习率变化的参考值。使用对数函数可以减缓变化的速度，同时减少网络发散的可能。

如图4所示，固定学习率的梯度下降曲线较为平滑，但有可能会陷入局部最小值，且随着loss越来越小，相对较大的学习率还会使曲线出现振荡。而可变学习率虽然下降曲线波动幅度较大，但最终也会趋向最小值，且可以探测到更多的极值点。

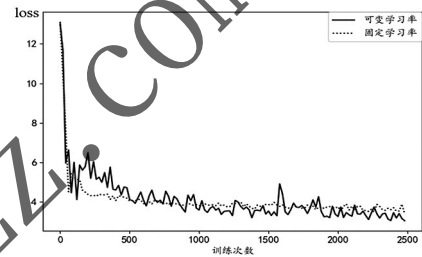


图4 不同学习率的loss曲线

Fig.4 Loss curve of different learning rates

4 实验与结果分析(Experiment and result analysis)

本文按照7:3的比例划分训练集和测试集，同时采用保留交叉验证，每次随机选择数据作为训练集和测试集。在进行10次训练后，选取模型在测试集上的平均值作为模型最终得分。模型在要素识别部分的最终结果如表2所示。可以看到，其他模型在要素识别部分的效果均低于BERT模型，BERT模型的F1值可以达到0.89。其中，全文输入的BERT模型效果较差，原因便是上文所说的最长句子截断问题，导致输入和输出不再匹配。

表2 不同模型的分类结果

Tab.2 Classification results of different models

编号	分类器	F1值
1	支持向量机(SVM)	0.66
2	CNN	0.78
3	RNN	0.82
4	BERT(全文输入)	0.74
5	BERT(分句)	0.89

使用效果最优的BERT模型的预测结果作为BP神经网络的输入，在衡量模型的优越性方面使用直观的MAE误差，即真实值和预测值的平均绝对值偏差。最终模型的效果如表3所示。同样可以看到，BP神经网络的MAE偏差小于其他模型，

达到了2.5个月以内。

表3 不同模型的MAE误差

Tab.3 MAE error of different models

编号	回归模型	MAE/月
1	Linear Regression线性回归	4.3
2	决策树(Decision Tree)	3.35
3	K近邻(KNN)	3.1
4	支持向量机(SVM)	2.78
5	BP神经网络	2.46

5 结论(Conclusion)

刑期预测作为当前智慧法院建设的重要组成部分,引起了众多学者的关注,无论是案情要素识别还是最终的刑期预测,都提出了很多新方法。本文所提出的刑期预测方法以盗窃案为切入点,将刑期预测的误差缩小到了2.5个月以内,具有实用价值,且本文对盗窃案裁判文书所预测的案情要素可以在法官判案时提供较多的参考。本文同样存在一定的不足:盗窃案件案情要素较为清晰,相对其他类型案件更易识别。如果拓展到其他类型的案件,需要对案情要素进一步分析,划分更具细粒度的案情要素,以保证刑期预测的准确率。此外,需要重新筛选案情要素,才能应用到其他类型的案件中。

参考文献(References)

- [1] 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.
- [2] RINGQUIST E J, EMMERT C E. Judicial policymaking in published and unpublished decisions: The case of environmental civil litigation[J]. Political Research Quarterly, 1999, 52(1):7-37.

(上接第13页)

- [16] 兰天翔,向子彧,刘名果,等.融合U-Net及MobileNet-V2的快速语义分割网[J].计算机工程与应用,2021,57(17):175-180.
- [17] 吕璐璐,陈树越,王利平,等.水体微纤维图像识别的改进MobileNetV2算法[J].郑州大学学报(工学版),2021,42(05):25-31.
- [18] CHENG J, WANG P, LI G, et al. Recent advances in efficient computation of deep convolutional neural networks[J]. Frontiers of Information Technology & Electronic Engineering, 2018, 19(01):64-77.
- [19] PAN H H, PANG Z J, WANG Y W, et al. A new image recognition and classification method combining transfer learning algorithm and mobilenet model for welding defects[J]. IEEE Access, 2020, 8:119951-119960.

- [3] LAUDERDALE B E, CLARK T S. The supreme court 's many median justices[J]. American Political Science Review, 2012, 106(4):847-866.
- [4] KATZ D M. Quantitative legal prediction—or—how I learned to stop worrying and start preparing for the data-driven future of the legal services industry[J]. Emory Law Journal, 2012, 62:909.
- [5] 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(07):505-511.
- [6] YANG Z, YANG D, DYER C, et al. Hierarchical attention networks for document classification[C]// KNIGHT K. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. San Diego, California: ACL, 2016:1480-1489.
- [7] 谭红叶,张博文,张虎,等.面向法律文书的量刑预测方法研究[J].中文信息学报,2020,34(03):107-114.
- [8] ZHONG H, GUO Z, TU C, et al. Legal judgment prediction via topological learning[C]// RILOFF E. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. Brussels, Belgium: EMNLP, 2018:3540-3549.
- [9] YANG W M, JIA W J, ZHOU X J, et al. Legal judgment prediction via multi-perspective bi-feedback network[J]/OL. (2019-05-16) [2021-10-04]. <https://arxiv.org/abs/1905.03969>.

作者简介:

郭彬彬(1996-),男,硕士生.研究领域:自然语言处理。

- [20] SO M H, HAN C S, KIM H Y. Defect classification algorithm of fruits using modified mobilenet[J]. The Journal of Korean Institute of Information Technology, 2020, 18(7):81-89.
- [21] 陈智超,焦海宁,杨杰,等.基于改进MobileNet v2的垃圾图像分类算法[J].浙江大学学报(工学版),2021,55(08):1490-1499.
- [22] 周博文.基于MobileNetV2的目标检测算法研究与应用[D].武汉:湖北工业大学,2020.
- [23] 丁常宏,王守宇,高鹏.基于SSD-MobileNet V1深度学习算法的药用植物叶片识别方法[J].江苏农业科学,2020,48(22):222-228.

作者简介:

李文远(1993-),男,硕士生.研究领域:图像识别。

韩俊英(1975-),女,硕士,教授.研究领域:软件工程理论与方法.本文通讯作者。