Applied Mathematics, Modeling and Computer Simulation
C.-H. Chen et al. (Eds.)

doi:10.3233/ATDE221093

Crime Prediction with Quantitative Information

Zhiyuan WANG, Chen ZHANG[1], Dingding HU
College of Cyberspace Security, Gansu University of political science and law, Lanzhou, Gansu, China

Abstract. Aiming at the poor prediction effect in the crime prediction task, a crime prediction model incorporating quantitative information is proposed. According to the numerical basis of judicial sentencing, quantify the values in the text, use the self-attention mechanism and maximum pooling to highlight the weight of quantitative information and keywords, then use the bidirectional long and short memory cycle network to extract the joint features between quantitative information and keywords, and use the capsule network to extract the depth features; Due to the imbalance of data, a priori weight is used in the loss function to balance the data. Experiments show that this method has better prediction effect than other models.

Keywords. Crime prediction, Quantitative characteristics, A priori weight, Maximum pooling

1. Introduction

The crime prediction task is to predict the crimes committed by the case according to the description of the facts in the legal case in the legal document [1]. The data includes two parts: facts and crimes. Facts are the description of the actual situation, and the high generalization of the case [2]. For example: robbery, intentional homicide, etc.

In the 1950s, early scholars used mathematical and statistical methods to study crime prediction. In recent years, with the rapid development of deep learning and other related technologies, great progress has been made in the fields of crime prediction research. Zhong et al. [3] used the method of neural network model to build a topological structure among multiple subtasks, and complete multiple tasks of legal prediction; Chen[4] takes the prediction of each crime as a separate task by using the method of graph; Yang [5] designed the crime prediction model of recurrent neural network and introduced the rule information into the model; Guo Junjun et al. [6] proposed a crime prediction model incorporating case auxiliary sentences by constructing case auxiliary sentences.

In the crime prediction task, the descriptions of some crimes are very similar, and the prediction effect is poor. However, there is a numerical basis in the sentencing process of legal cases. According to this characteristic, the numerical value in the fact description of legal documents is quantified. Because the data of legal documents is

1 Corresponding Author, Chen ZHANG, College of Cyberspace Security, Gansu University of political science and law, Lanzhou, Gansu, China; E-mail: zc6454@gsupl.edu.cn

unbalanced data, the prior weight is obtained by statistics of data labels, added to the cross-entropy loss function to deal with the problem of data imbalance.

2. Model building

The crime prediction model integrated with quantitative information includes five parts, the model structure of this paper is shown in figure 1.

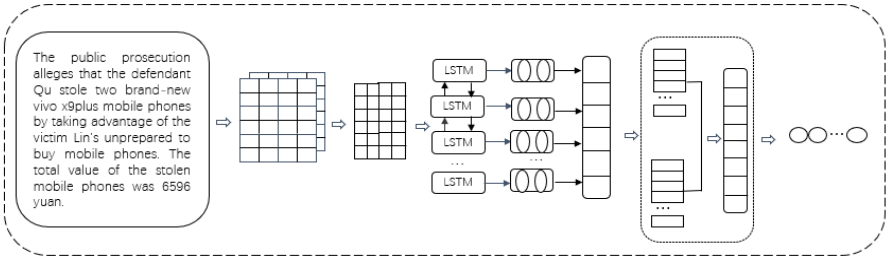

Figure 1. Structure diagram of crime prediction network model

2.1. Numerical Quantization

There is a certain numerical basis for the characterization of legal crimes, such as: drunk driving and drunk driving, robbery and robbery, and the description is shown in table 1. such as robbery and robbery, the facts are described in table 2.

Table 1. Description Conviction

Crime	Describe
Drunk driving	Driving behavior with alcohol content in blood greater than or equal to 80mg/100ml.
Drinking drive	Driving behavior with blood alcohol content greater than or equal to 20mg/100ml and less than 80mg/100ml.
Robbery	No matter the amount of robbery, criminal responsibility should be investigated.
Crime of robbery	The filing standard of the crime of robbery is 500 yuan to more than 2000 yuan (Interpretation on Several Issues concerning the application of law in the trial of criminal cases of robbery)

Table 2. Description of legal documents and facts of robbery and robbery

Crime	Legal document case fact description
Robbery	××× crime that Zhang pretended to ride the Zhejiang D driven by the victim Meng ××× Taxi No. 1 goes to a riverside park in Cao 1. Zhang robbed Meng with a fruit knife×××
Crime of robbery	×××, robbed two brand-new vivo x9plus mobile phones on the grounds of buying mobile phones while the victim, Mr. Lin, was unprepared. The value was 6596 yuan.

As shown in table 2, in the case of the crime of robbery, "6596 " meets the value basis standard of 500 to 2000 for the crime of robbery in table 1. If the value of the legal document data is not quantified, it may be unique for each value that appears, the model is not easy to distinguish the values. However, if quantization is carried out, the model is very easy to distinguish the values. The specific standards are shown in table 3.

Table 3. Quantitative criteria

numerical value	0-18	18-100	100-1K	1K-3K	3K-3W	3W-6W	6W-
Quantized value	1	18	100	1000	3000	30000	60000

2.2. Self- attention Mechanism

If only for extracting numerical features, it is impossible to convict the case, because, only numerical and other keyword features are combined together, can they be a very important basis for convicting the fact description. This paper uses self-attention mechanism to extract keywords in fact description.

Attention mechanism was first proposed in the field of image. When people observe an image, they selectively focus on a specific part of the image and ignore other parts. Researchers first applied it to the field of image, allocating different attention according to the importance of different areas in the image, and called this method attention mechanism. Self- attention is a kind of attention mechanism. It is an efficient analysis of data information to obtain internal weight information. The operation formula is shown in formula (1). Q, K and V represent query, key value.

$$Attntion(\mathrm{Q},\mathrm{K},\mathrm{V}) = \mathrm{softmax}(\frac{QK^{T}}{\sqrt{n}})\,\mathrm{V} \tag{1}$$

Suppose that the data input into the self-attention mechanism is E. Through the self-attention mechanism, the weight of some important words are increased, that is, the weight of keywords is increased. For example, in the robbery case in table 2, the weight information of each word is obtained through the self-attention mechanism, the weight of some more important words has increased, such as 'robbed' and '6596' quantitative value, because they play an important role in the case.

2.3. Bidirectional Long-Term and Short-Term Memory Recurrent Neural Network (Bi-LSTM)

After extracting keywords, if you analyse these keywords alone, you can't understand at all. This paper uses LSTM to extract the joint features between keywords.

In 1997, schmidhuber proposed an improved LSTM. The LSTM includes input gate $o(\mathrm{t})$,input gate $i(\mathrm{t})$ and forgetting gate, as shown in formula (2-7). The input gate controls the retention of those information at the current time, forgetting gate is used to control the abandonment of those information in the memory unit at the last moment [7]. The internal structure of long-term and short-term memory cycle neural networks is shown in figure 2.

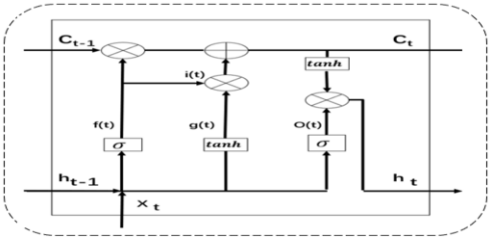

Figure 2. Internal structure diagram of long-term and short-term memory cycle neural network

$$\mathrm{i}_{\mathrm{t}} = \sigma(\mathrm{W}_{\mathrm{i}} \cdot [\mathrm{h}_{t-1}, \mathrm{x}_{t}] + \mathrm{b}_{\mathrm{i}}) \tag{2}$$

$$\mathrm{g}_{\mathrm{t}} = \sigma(\mathrm{W}_{\mathrm{g}} \cdot [\mathrm{h}_{t-1}, \mathrm{x}_{t}] + \mathrm{b}_{\mathrm{g}}) \tag{3}$$

$$f_{t} = \sigma(\mathrm{W}_{f} \cdot [\mathrm{h}_{t-1}, \mathrm{x}_{t}] + \mathrm{b}_{f}) \tag{4}$$

$$C_t = f_t \odot C_{t-1} \oplus i_t \odot g_t \tag{5}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t \odot \tanh(C_t) \tag{7}$$

For the text data, the information at the current time is not only related to the previous time but also to the next time. In order to improve the performance of the model, the Bi-LSTM is used to extract the timing characteristics of the data according to the input data. When the weight of a word is large, it is likely to be retained, so as to extract the joint information between keywords. As shown in table 2, the joint features between keywords are extracted, its structure is shown in figure 3.

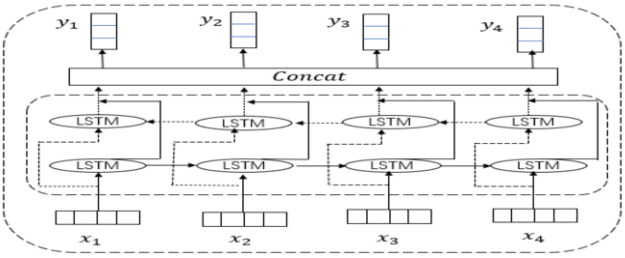

Figure 3. Structure diagram of cycle neural network

2.4. Capsule Layer

Through the Bi-LSTM, the joint features between keywords are extracted, but there are deep-seated features such as location between these joint features. This paper uses capsule network to extract deep-seated features between joint features.

2.5. Classification Layer

In the classification layer, Because the data is unbalanced, it may lead to insufficient training of low-frequency data. The prediction effect of crimes is not ideal. In order to solve the problem of uneven data, this paper adds a priori weight to each type of crime.

3. Experiment and Result Analysis

3.1 Date Set and Evaluation Index

Select Chinese criminal case data set. Select part of the data and construct C-Small and C-Big. C-Small contains 25667 pieces of legal document training data, 3076 pieces of test data, including 7 crimes; C-big contains 78193 pieces of legal document training data, 8561 pieces of test data, including 16 crimes. Statistical crimes, the rectangular chart is shown in figure 4. The statistical histogram of sentence length is drawn, as shown in figure 5. The length is set to 1200, the sample coverages are 0.9373 and 0.9689.

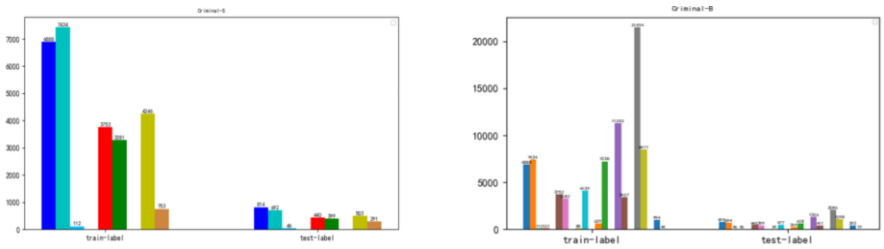

Figure 4. Count statistical histogram

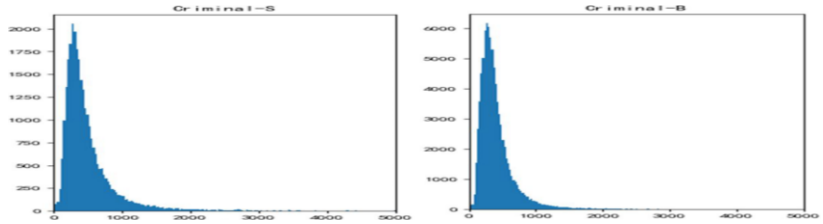

Figure 5. Sentence length statistical histogram

In order to evaluate the experimental results, the accuracy rate and micro average are selected as the evaluation indicators. The calculation formulas are shown in (8).

$$F1=\frac{2*\sum_{i=1}^{n} TP}{2*\sum_{i=1}^{n} TP+\sum_{i=1}^{n} FP+\sum_{i=1}^{n} FN} \tag{8}$$

3.2 Experimental Environment and Parameter Setting

Windows 10 operating system, pytorch1.6.0, python 3.8, Pycharm; GPU model tesla-p100-pcie-16gb.Batch size is 16, word vector dimension is 50.

3.3 Comparison Model

CNN: The CNN model with multiple filters is used to carry out the classification task. Self-Attn: Self-attention mechanism is used to carry out the classification task. Capsule: The capsule network model is used to carry out the classification task. Bi-LSTM: Using the bidirectional LSTM network to carry out the classification task

*3.4 E*xperimental Results and Analysis

The experimental results are shown in table 4.

Table 4. Statistical table of experimental results

Model	Criminal-B		Criminal-S	
	ACC	F1	ACC	F1
CNN	89.38	93.16	89.82	93.28
Self-Attn	90.12	93.80	88.85	93.25
Capsule	87.34	90.89	88.63	92.03
Bi-LSTM	90.87	93.57	90.15	93.63
CNN-LSTM	92.16	94.10	91.01	93.16
Self-Attn-CNN	91.00	93.60	90.05	93.07
CNN-Bi-LSTM-Self-Attn	93.10	95.32	92.53	94.13
Method in this paper	94.72	96.10	93.40	95.12

In order to verify the influence of numerical quantization and prior weight on the experiment, ablation experiments were carried out. The results are shown in table 5.

Table 5. Comparison of results of ablation experiment

Model	Criminal-B		Criminal-S	
	ACC	F1	ACC	F1
ABC	93.17	95.36	91.84	94.48
NL-CBA	93.50	95.67	93.37	94.59
Method in this paper	94.72	96.10	93.40	95.12

Further analyse the experimental results and make statistics on the number of prediction errors, as shown in table 6.

Table 6. Statistical table of false test

Model	Criminal-B		Criminal-S	
	Number	Mean value	Number	Mean value
ABC	[13,85,77,92,62,73,128,54,41,46,14,16,85,112,38,25]	60.06	[57,51,45,58,38,35,36]	45.71
NL-CBA	[6,136,20,35,74,48,303,41,24,30,5,12,153,158,4,19]	66.75	[49,42,58,50,37,41,39]	45.14
Method in this paper	[2,38,69,73,67,88,88,27,38,44,15,19,38,50,10,19]	42.88	[29,36,41,18,37,26,43]	32.85

4. Conclusion

Aiming at the fact that the description of accusations in the accusation prediction task is similar, which leads to the poor accusation prediction, this paper proposes an accusation prediction model integrated with numerical quantitative characteristics. It is easy to distinguish crimes and extract features by quantifying the fact text data of case documents. Due to the unevenness of the data, a priori equilibrium data is added to the loss function. The results show that the proposed method has a good effect.

Acknowledgments

Project: 2022 provincial industrial support project (2022CYZC-57); Gansu soft science project (20CX4ZA074); Scientific research project of colleges and universities of Gansu Provincial Department of Education (2015A-114); 2020 school level education reform project of Gansu University of political science and Law (GZJG2020-B06); Gansu Province 2020 provincial virtual simulation first-class course (GZYL2020-18).

References

[1] Cui bin, Zou Lei, Xu Mingyue Research on criminal case decision-making integrating CNN and bilstm [j] Computer application research, 2020,37 (S2): 181-183.

[2] Wang Jiawei, Zhang Hu, Tan Hongye, Wang Yuanlong, Zhao Hongyan, Li Ru Multi label crime prediction based on semantic differences of words [j] Chinese Journal of information technology, 2019,33 (10): 127-134.

[3] Zhong HX, Guo ZP, Tu CC, Xiao CJ, Liu ZY, Sun MS. Legal judgment prediction via topological learning. In: Proc. of the 2018 Conf. on Empirical Methods in Natural Language Processing. Brussels: ACL, 2018. 3540-3549.

[4] Chen S , Wang P ,Fang W , et al. Learning to Predict Crimes for Judgment with Legal Graph[M], 2019.

[5] Yang Z ,Wang P, Zhang L , et al. A Recurrent Attention Network for Judgment Prediction[J],2019.

[6] Guo Junjun, Liu Zhencheng, Yu Zhengtao, Huang Yuxin, Xiang Yan Prediction of low-frequency and confusing crimes integrated into the auxiliary sentence of the case [j] Journal of software, 2021,32 (10): 3139-3150. 006028.

[7] Zhao Dongmei, Wu Yaxing, Zhang Hongbin Network security situation prediction based on ipso bilstm [j/ol] computer science:1-10[2022-05-17].