# Applying Data Discretization to DPCNN for Law Article Prediction

Hu Zhang[1], XinWang[1], Hongye Tan[1], Ru Li[1,2]

[1]School of Computer and Information Technology, Shanxi University, Taiyuan, China
zhanghu@sxu.edu.cn
[2]Key Laboratory of Computing Intelligence and Chinese Information Processing, Ministry of Education, Shanxi University, Taiyuan, China

**Abstract.** Law article prediction is a crucial subtask in the research of legal judgments, aiming at finding out the adaptable article for cases based on criminal case facts and relevant legal provisions. Criminal case facts usually contain a lot of numerical data, which have an essential impact on law article predicting. However, existing charge prediction models are insensitive to the size of numbers such as money and age, and lack of special analysis and processing for these data. Moreover, the models currently applied to legal judgment still cannot effectively acquire long-distance dependencies of legal texts. In response to this, we propose an automatic law article prediction model based on Deep Pyramid Convolutional Neural Networks (DPCNN) with data preprocessing. Experimental results on three different datasets show that our proposed method achieves significant improvements than other state-of-the-art baselines. Specifically, ablation test demonstrate the validity of data preprocessing in law article prediction.

**Keywords:** Law article prediction, Legal judgments, Data discretization, DPCNN.

## 1 Introduction

In recent years, China's judicial institutions at all levels have entered the construction period of intelligent courts. In 2018, the Chinese AI and Law challenge (CAIL2018) further promoted the leap-forward development of judicial informatization to intellectualization. At present, intelligent judicial services are roughly divided into three levels: (1) Assist in some simple, mechanical and repetitive tasks, such as optical character recognition and legal text generation. (2) Learn decision-making rules to assist the legal judgment, such as recommendation of similar cases and legal document verification. (3) Carry out judicial-related services for the people’s convenience, such as legal consultation, and intelligent legal judgment. In these legal services, auxiliary and intelligent legal judgments have been widely concentrated by many research institutions.

As a promising application in intelligent judicial services, automatic legal judgment prediction has been studied for decades. Initially, most of the relevant researchers used mathematical and statistical methods to conduct the task. Under the impact of machine learning, afterward most scholars tried to extract textual features from legal texts and predict legal judgment decisions. With the development of deep learning technology, most of the mainstream methods of legal judgment prediction focused on using a variety of neural network models, and the corresponding experimental results have greatly improved.

Law article prediction is a crucial subtask in the study of intelligent legal judgments. It aims to use case facts and related legal provisions to predict the applicable law article for cases, the main challenges in current research include:(1) A lot of numerical data involving money and age appear in criminal case facts, and the existing prediction models cannot effectively acquire their true meaning. (2) Long-distance dependencies between the features exist in criminal judgments, and the existing law article prediction models cannot catch the dependency relations well.

To help address these issues, we preprocess the numerical data (including money, age, etc.) in case facts of the criminal judgments, and introduce the processed data into DPCNN model that can effectively acquire text long-distance dependencies [1]. The general process is shown in Figure 1. Among them, the input of the model is case facts, the output is law article number, and the detailed structure of the DPCNN model is partly omitted.

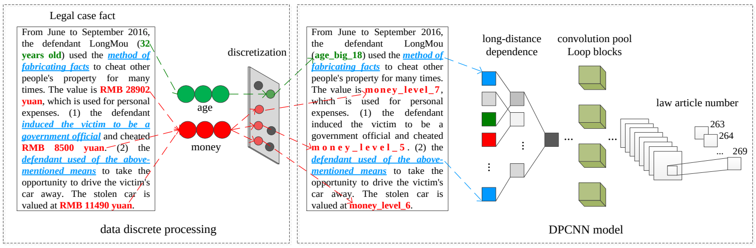


**Fig. 1.** Law article prediction process.

**Contributions**. Our contributions are the following:

(1) Combining with relevant law article, legal interpretation documents and criminal judgments, we construct the data discretization pattern to preprocess the numerical data in the case facts.

(2) According to the particularity of the law article prediction and the long-distance dependencies of the legal texts, we apply the data discretization to DPCNN for law article prediction.

(3)We conduct several experiments on three different datasets, and our proposed method achieves significant improvements than other state-of-the-art baselines.

## 2 Related Work

Using data analysis in legal judgments has attracted the attention of legal researchers in the 1950s. Early work focused on the use of mathematical and statistical algorithms to analyze legal cases in specific scenarios, such as Kort predicted the Supreme

Court's decision mathematically and made a quantitative analysis of “lawyer's rights” cases [2]. Ulmer used rule-based method to analyze legal fact data, and assist judges to tease case evidences [3]. Nagel counted a number of legal variables to serve judges, and helped the public to seek legal aid [4]. Keown carried on the legal forecast research based on the mathematical model [5]. Ringquist and Emmert studied judicial decisions by taking environmental civil action as an example [6]. Lauderdale and Clark applied the substantive similarity information between cases to estimate different substantive legal issues and long-term judicial preferences [7].

With the development of machine learning and text mining technology, more and more researchers have explored legal judgment tasks based on text classification framework. Most of these researches extract features from legal text [8-11] or case profiles [12]. Obviously, using the shallow text features and human design factors, it not only costs numerous labors but also has the poor generalization ability in cross-scenario applications.

In recent years, the neural network model has achieved excellent results in text classification tasks. Collobert used convolution filters to process text sequences in sliding windows, and utilized max-pooling to capture effective local features [13]. Kalchbrenner proposed a dynamic convolution neural network, which uses dynamic k-max pool operation to model sentences semantically [14]. Lei proposed a new feature mapping operator to generate discontinuous n-gram features for processing text data better [15]. Wang used a large number of classification knowledge base to enhance the model performance [16]. Johnson directly applied CNN to high-dimensional text data and proposed a variable of bag-of-words conversion in convolution layer to improve the accuracy of text classification [17]. Zhang conducted an empirical study on text classification using character level convolution network, providing a reference for scholars who later used character level convolution neural network [18]. Xiao proposed a neural network architecture, which uses convolution and cyclic layer to encode input character effectively, and can achieve better performance through fewer parameters compared with the above convolution model [19].

Inspired by the successful application of neural networks in natural language processing tasks, Kim tried to combine the neural network model with legal knowledge to conduct legal judgments prediction [20]. Luo proposed a neural network based on the attention mechanism, which incorporated law articles to the charge prediction task [21]. Hu attempted to use ten legal discriminant attributes to predict confusing charges [22]. The above studies all use criminal law cases as experimental datasets. Ye used the seq2seq model to generate interpretable court opinions based on the case facts and charge prediction in civil legal documents [23]. For the task of law article prediction, Liu designed a text mining based method, which allows the general public to use everyday vocabulary to describe their problems and find pertinent law articles for their cases [24]. Liu employed techniques of instance-based classification and introspective learning for the law article classification task [25].

At present, most of the studies on legal judgments focus on charge prediction, but few on the law article prediction. In addition, the existing researches mainly concerns on the shallow textual features and classification framework, lack of in-depth data analysis and application of law article content. Based on this, we focused on improving the method from two aspects: the influence of numerical data on the law article prediction and the acquisition of long-distance dependencies in legal texts.

## 3 Data Discretization

In this section, we propose a data discretization method which jointly applies case facts and criminal law articles. The used experimental dataset include the criminal case facts and the law articles. Criminal cases mostly contain numerical data, such as the money of theft, the weight of drug smuggling, the age of the plaintiff, and so on, there are obvious differences of the number in the case facts corresponding to the different law article. Therefore, we construct the data discretization pattern, and replace the original numerical content with the corresponding interval labels, which enable the model to recognize the specific meaning of numerical data of the different sizes. In the relevant legal interpretations, the amount of money is usually divided into more specific intervals, as shown in Figure 2.

> In accordance with the provisions of Article 264 of the Criminal Law and the current level of economic development and social security, the Supreme Law, the Supreme Procuratorate and the Ministry of Public Security stipulate the following criteria for determining the amount of theft:
> 1. The amount of personal theft of public and private property is relatively large, starting from 500 to 2000 yuan.
> 2. The amount of personal theft of public and private property is huge, starting from 5,000 to 20,000 yuan.
> 3. The amount of personal theft of public and private property is particularly huge, starting from 30,000 to 100,000 yuan.

**Fig. 2.** Examples of money interval in legal interpretation.

Among them, there are money number interval labels such as “relatively large”, “huge” and “particularly huge”. Judgment results of different money intervals are quite different, and the machine cannot directly acquire its specific meaning in the process of learning, such as money, age, etc. Hence, we combine the judgments characteristics and experimental requirement to preprocess the data, as shown below.

**Money interval division**. After analyzing of the case facts, we divided the amount of money in judgments into 24 sections, such as money_level_1: “0-1000 yuan”... Money_level_24: “More than 5000000 yuan”. The partition process and results are shown in Figure 3, in which the legal provisions in the text box are related to money regulations, some of which are omitted.

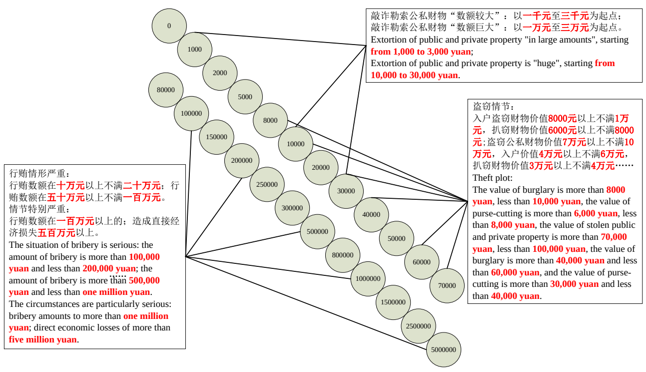


**Fig. 3.** The result of money interval division.

**Age interval division and name removal**. According to legal provisions on the offenders' age, the age-interval is divided into adults and minors with the labels "age_big_18" and "age_little_18" respectively. At the same time we remove names from legal documents, such as "Li Mou", "Qian XX" and so on. The specific processing flow is shown in Figure 4.

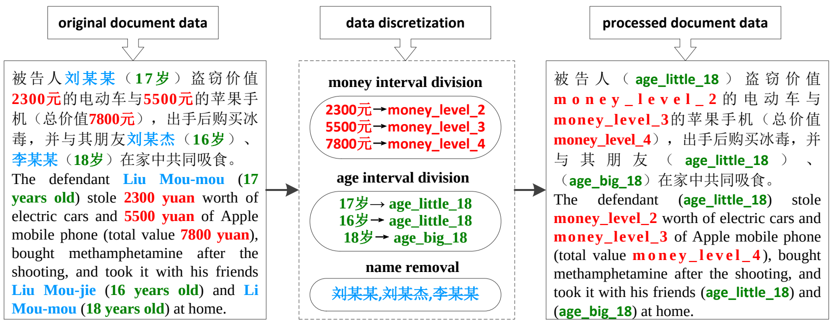


**Fig. 4.** Data discretization flowchart.

## 4 DPCNN Model for Law Article Prediction

Existing researches usually fuses LSTM model to acquire long-distance dependencies, such as CNN+LSTM [26]. However, the computational complexity of LSTM model is more than four times that of RNN, so the time complexity of LSTM fusion model increases dramatically. DPCNN model follows the bottom structure of CNN, thus it keeps low time complexity while acquiring long-distance dependencies. Therefore, we use DPCNN model to predict law article on 1.7 million legal dataset in this paper.

### 4.1 Bottom Structure

DPCNN model adopts the method of text region embedding. Similar to the bottom structure of CNN model, we first vectorize every word in text at the input level, and concatenate word vectors according to the corresponding location in legal text sequence, finally get the word vector matrix $\boldsymbol{X}$ for text sequence, as shown in Formula (1):

$$\boldsymbol{X}_{1:n} = \boldsymbol{x}_1 \oplus \boldsymbol{x}_2 \oplus \cdots \oplus \boldsymbol{x}_n \tag{1}$$

$\oplus$ is the word vector connection operator. $\boldsymbol{x}_i$ is the word vector of the $i_{th}$ word in sentence. $\boldsymbol{X}_{i:i+j}$ means $\boldsymbol{x}_i, \boldsymbol{x}_{i+1}, \ldots, \boldsymbol{x}_{i+j}$ has a total of $j+1$ word vectors. Convolution operations involve filter $\boldsymbol{W}$, which is applied to $h$ word windows to generate new features. For example, a window on the word vector $\boldsymbol{X}_{i:i+h-1}$ generates feature $\boldsymbol{C}_i$, as shown in formula (2):

$$\boldsymbol{C}_i = f(\boldsymbol{W} \cdot \boldsymbol{X}_{i:i+h-1} + b) \tag{2}$$

$b$ is a bias term and $f$ is a non-linear function. Apply max-pooling operation to select maximum features $\boldsymbol{C_{max}} = max\{\boldsymbol{C_i}\}$ , Dropout is used to prevent over-fitting. Give $\boldsymbol{Z} = [\boldsymbol{C_1}, \boldsymbol{C_2}, \dots, \boldsymbol{C_m}]$ with assuming that there are $m$ filters, the formula for calculating final feature vector $\boldsymbol{y}$ is shown in formula (3). Among them, $\boldsymbol{Z}$ denotes the feature set of $m$ filters, $\circ$ denotes the multiplication operation by elements, and $\boldsymbol{r}$ denotes the mask vector.

$$\boldsymbol{y} = \boldsymbol{W} \cdot (\boldsymbol{Z} \circ \boldsymbol{r}) + b \tag{3}$$

### 4.2 Long Distance Dependence

DPCNN model use two-level equal-length convolution and maximum pooling, and perform maximum pooling after each convolution, where $size = 3$ and $strid = 2$. In this model the length of the output sequence is half as long as before, hence, the legal text fragments that the model can perceive are twice as large as before, as illustrated in Figure 5. Before pooling, the model can perceive the information of position length is 3. After 1/2 pooling layer, it can perceive information about 6 position length. Therefore, repeated execution of the convolution pooling cycle block can capture the long-distance dependencies for legal texts.

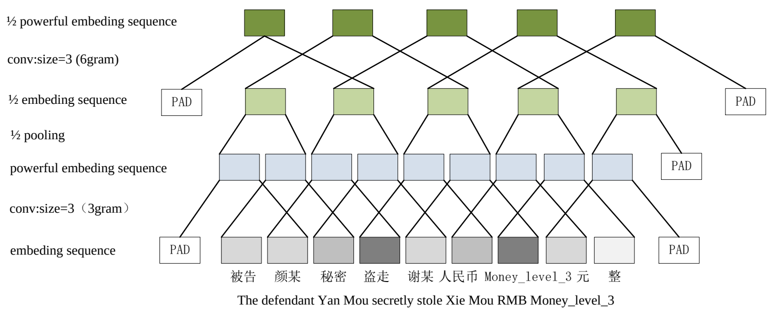


**Fig. 5.** Example of acquiring long-distance dependency.

### 4.3 Model Structure

As illustrated in Figure 6, the input and output of this model are the legal texts and the adaptable law article numbers respectively. Firstly, we preprocess legal texts with numerical data discretization and name removal and conduct text region embedding. Next, after two convolution layers are processed, block is recycled four times for down sampling, which includes the convolution and maximum pooling operations of size 3 and step 2. Then, we use the maximum pooling operation to aggregate the representation of each document into a vector, and output the prediction number of the law article through the full connection layer. Here, the illustration within the shaded box is an implementation process of one convolution pool block.

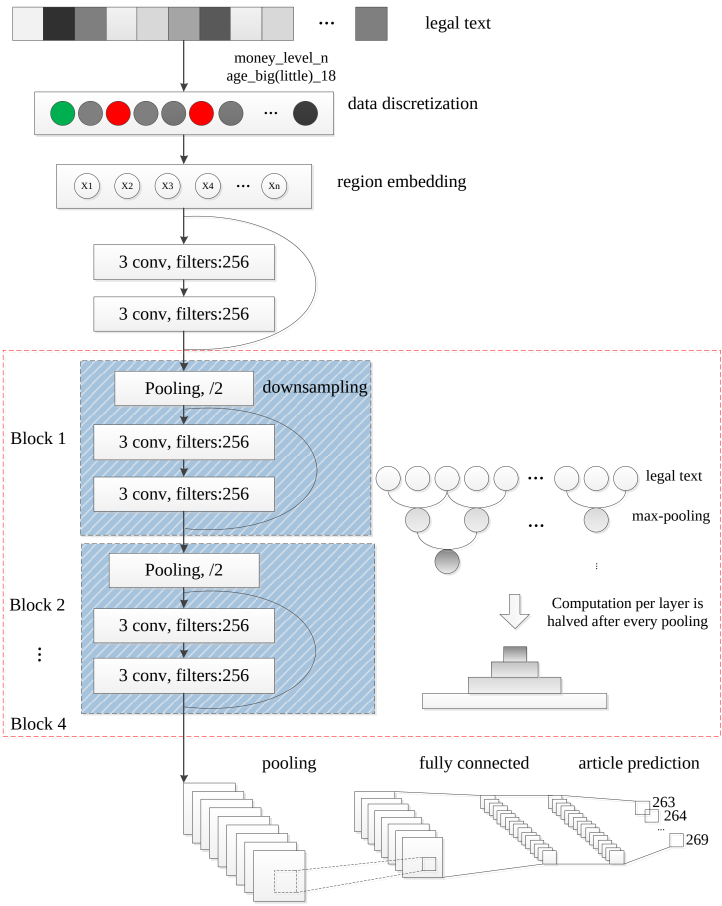


**Fig. 6.** The architecture of DPCNN.

# 5 Experiments

## 5.1 Dataset Collection

Since there is still no open available datasets for law article prediction at present, we collect three experimental datasets on the basis of “CAIL2018” from different perspectives, including CAIL2018-L, CAIL2018-H and CAIL2018-S. CAIL2018-L dataset consists of all charges and law articles cases, which is a typical category imbalance dataset including some fewer charges, such as “smuggle nuclear materials” and “unknown sources of huge property”. In addition, we removed some low-frequency law articles cases and constructed CAIL2018-H dataset, which can verify the prediction model on the category relative balance dataset. Furthermore, CAIL2018-S dataset including 196,231 cases that randomly selected from CAIL2018-L dataset is built to test the learning effect of the model on small-scale datasets, as shown in Table 1.

**Table 1.** Statistics of datasets.

| Datasets | CAIL2018-L | CAIL2018-H | CAIL2018-S |
|---|---|---|---|
| Number of cases | 1710856 | 1477184 | 196231 |
| Classification of articles | 183 | 62 | 183 |
| Training set | 1645840 | 1421921 | 146592 |
| Dev set | 32508 | 27632 | 24821 |
| Test set | 32508 | 27631 | 24818 |

## 5.2 Experimental Results

In this experiment we use common evaluation indexes in text classification field: accuracy (P), recall rate (R) and $F1_{macro}$. The final effect was evaluated by scoring, which fused $F1_{micro}$ and $F1_{macro}$. The exact process is shown in formula (4).

$$S = \frac{F1_{macro}+F1_{micro}}{2} \times 100 \tag{4}$$

**Table 2.** Comparison on the experimental results of models.

| Datasets | Methods | P | R | $F1_{macro}$ | S |
|---|---|---|---|---|---|
| CAIL2018-L | SVM | 0.751 | 0.763 | 0.695 | 72.598 |
| | FastText | 0.833 | 0.837 | 0.792 | 81.350 |
| | HAN | 0.864 | 0.869 | 0.821 | 84.375 |
| | TextCNN | 0.885 | 0.872 | 0.837 | 85.773 |
| | TextRNN | 0.846 | 0.836 | 0.808 | 82.449 |
| | DPCNN | 0.891 | 0.897 | 0.842 | 86.799 |
| | CNN fusion | 0.902 | 0.897 | 0.854 | 87.675 |
| | *DPCNN fusion | **0.913** | **0.906** | **0.866** | **88.526** |
| CAIL2018-H | SVM | 0.773 | 0.762 | 0.705 | 73.623 |
| | FastText | 0.872 | 0.876 | 0.823 | 84.850 |
| | HAN | 0.893 | 0.882 | 0.838 | 86.273 |
| | TextCNN | 0.896 | 0.901 | 0.857 | 87.775 |
| | TextRNN | 0.883 | 0.862 | 0.825 | 84.869 |
| | DPCNN | 0.917 | 0.904 | 0.865 | 88.773 |
| | TextCNN fusion | 0.924 | 0.908 | 0.873 | 89.447 |
| | *DPCNN fusion | **0.931** | **0.922** | **0.894** | **91.325** |
| CAIL2018-S | SVM | 0.713 | 0.706 | 0.651 | 68.024 |
| | FastText | 0.795 | 0.792 | 0.769 | 78.125 |
| | HAN | 0.836 | 0.823 | 0.777 | 80.322 |
| | TextCNN | 0.852 | 0.847 | 0.798 | 82.375 |
| | TextRNN | 0.801 | 0.794 | 0.765 | 78.124 |
| | DPCNN | 0.873 | 0.878 | 0.804 | 83.975 |
| | TextCNN fusion | 0.879 | 0.884 | 0.819 | 85.025 |
| | *DPCNN fusion | **0.903** | **0.894** | **0.821** | **85.974** |

As few of the existing studies involve the task of law article prediction, it is impossible to compare with the recent popular models. For experimental comparison, we use six common models of text classification: SVM, FastText, HAN, TextCNN, TextRNN and DPCNN. In order to fuse the slight difference of law article categories in this task, we add model fusion and threshold filtering to TextCNN and DPCNN. The experimental results are shown in Table 2. "*" denotes the best model, the roughened numbers represent the best results.

From Table 2, it can be seen that DPCNN model achieves the best results compared to all single models in three datasets. The operation of model fusion and threshold filtering further improves the results of law article prediction. The experiments in CAIL2018-H dataset significantly outperform other experiments, which show the imbalance of dataset has an important effect on the proposed model.

### 5.3 Ablation Test

In order to further illustrate the importance of our works to law article prediction, we design ablation test to investigate the effectiveness of these processing modules. DPCNN fusion model and TextCNN model were used to test on CAIL2018-H dataset respectively, and the following processes were eliminated one by one about "remove name", "age interval division", "money interval division", "model fusion" and "threshold filtering". The compared results of the experiments are shown in Table 3.

**Table 3.** Ablation test results.

| **Models** | **DPCNN Fusion** | | | **TextCNN** | | |
|---|---|---|---|---|---|---|
| **Evaluation Metrics** | P | R | $F1_{macro}$ | P | R | $F1_{macro}$ |
| all | 0.931 | 0.922 | 0.894 | 0.896 | 0.901 | 0.857 |
| w/o "remove name" | 0.929 | 0.917 | 0.889 | 0.894 | 0.897 | 0.849 |
| w/o "age interval division" | 0.927 | 0.915 | 0.885 | 0.891 | 0.889 | 0.837 |
| w/o "money interval division" | 0.921 | 0.906 | 0.877 | 0.876 | 0.864 | 0.823 |
| w/o "model fusion" | 0.904 | 0.873 | 0.856 | — | — | — |
| w/o "threshold filtering" | 0.895 | 0.861 | 0.842 | — | — | — |

Among them, "w/o" represents the removal process, and "-" means exclusion, "all" denotes all included operations. From Table 3, it can be seen that when DPCNN fusion model removes the "money interval division", the three evaluation indexes decrease significantly, but the removal processes of "age interval division" and "remove name" have little influence on the experimental results. The operations of "model fusion" and "threshold filtering" have a stable effect on improving the experimental results. Compared to the ablation test results of DPCNN fusion model, the change range of the index of "remove name" and "age interval division" on TextCNN model are a little increased, and the process of "money interval division" is more obvious, which shows the fusion of the different models can effectively make up the deficiencies of acquiring knowledge of one model. Therefore, for law article prediction task, the proposed data discrete processing, model fusion and threshold filtering operations play irreplaceable roles on improving task performance.

### 5.4 Ablation Analysis

The ablation test dataset of CAIL2018-H include 62 categories of law articles, which only contains numbers in part of the case facts and leads to little changes in ablation test results. To this end, we further extract the cases with article 264 (theft) and article 384 (embezzlement of public funds) to form a comparative dataset, and verify the role of the "money interval division" processing in law article prediction.

Example case analysis: it can be showed in figure 7, the corresponding cases of articles 264 and articles 384 are confused, the reason for that is the both cases facts contain the similar keywords such as "defendant", "deceived", "repay", "yuan".

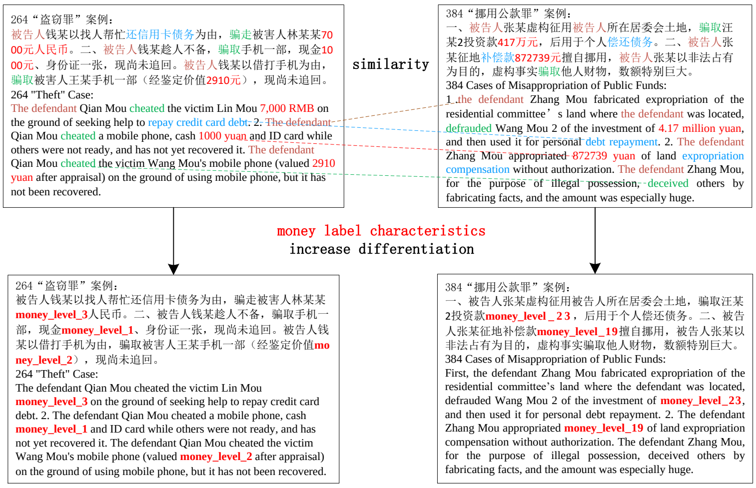


**Fig. 7.** Confusion case comparison.

For the confusing cases in figure 7, there is no better way to deal with the key features such as "defendant", or "repay", and direct deletion or substitution will lead to confusion with the facts of other similar law article. In view of the model cannot directly identify the numerical meanings of money, we use the operation of "money interval division", and replace money numbers with money labels. This preprocessing increases the distinctions between the different law articles, and effectively improves the performance of law article prediction for the confusing cases. The experiments fully verify the effect of numerical data discretization on law article prediction.

## 6 Summary

According to the requirement of law article prediction, we start from the characteristics of legal judgments and the challenges summarized in relevant research, and propose law article prediction method of applying data discretization to DPCNN. By applying numerical data discretization, model fusion, threshold filtering and other operations, the difficulties of law article prediction is solved to a certain extent, and the overall performance of law article prediction model is improved.

The experimental results show that this method can address some problems in the law article prediction, but the research still needs further improvement. In future, we will explore the following directions:

(1) In this work, we didn't introduce interpretability into the process of law article prediction, while it is usually necessary in judicial services. Thus, it is challenging to handle this specific need of legal judgment prediction.

(2) Our proposed prediction model is not well integrated with the process of manual decision, and lacks the reasoning ability in the legal judgment. Therefore, how to better solve the above problems is the focus of our next study.

## Acknowledgements

We want to express gratitude to the anonymous reviewers for their hard work and kind comments, which will further improve our work in the future. Thank the teachers and classmates of the project team for their guidance and help.

## References

1. Johnson, R., Zhang, T.: Deep pyramid convolutional neural networks for text categorization. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). pp. 562-570 (2017)
2. Kort, F.: Predicting Supreme Court decisions mathematically: A quantitative analysis of the "right to counsel" cases. American Political Science Review. 51(1), pp. 1-12 (1957)
3. Ulmer, S. S.: Quantitative analysis of judicial processes: Some practical and theoretical applications. Law & Contemp. Probs. 28: 164 (1963)
4. Nagel, S. S.: Applying correlation analysis to case prediction. Tex. L. Rev. 42: 1006 (1963)
5. Keown, R.: Mathematical models for legal prediction. Computer/lj, 2: 829 (1980)
6. Ringquist, E. J., Emmert, C. E.: Judicial Policymaking in Published and Unpublished Decisions: The Case of Environmental Civil Ligaton. Political Research Quarterly. 52(1), pp.7-37 (1999)
7. Lauderdale, B. E., Clark, T. S.: The Supreme Court's many median justices. American Political Science Review. 106(4), pp. 847-866 (2012)
8. Liu, C. L., Hsieh, C. D.: Exploring phrase-based classification of judicial documents for criminal charges in Chinese. International Symposium on Methodologies for Intelligent Systems. Springer, Berlin, Heidelberg. pp. 681-690 (2006)
9. Lin, W. C., Kuo, T. T., Chang, T. J., Yen, C. A., Chen, C. J., Lin, S. D.: Exploiting machine learning models for chinese legal documents labeling, case classification, and sentencing prediction. Processdings of ROCLING, 140 (2012)
10. Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., Lampos, V.: Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. PeerJ Computer Science, 2: e93 (2016)
11. Sulea, O. M., Zampieri, M., Malmasi, S., Vela, M., Dinu, L. P., van Genabith, J.: Exploring the use of text classification in the legal domain. arXiv preprint arXiv:1710.09306 (2017)

12. Katz, D. M., Bommarito, I. I., Michael, J., Blackman, J.: Predicting the behavior of the supreme court of the united states: A general approach. arXiv preprint arXiv:1407.6333 (2014)
13. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. Journal of machine learning research, 2493-2537 (2011)
14. Kalchbrenner, N., Grefenstette, E., Blunsom, P.: A convolutional neural network for modelling sentences. arXiv preprint arXiv:1404.2188 (2014)
15. Lei, T., Barzilay, R., Jaakkola, T.: Molding cnns for text: non-linear, non-consecutive convolutions. arXiv preprint arXiv:1508.04112 (2015)
16. Wang, J., Wang, Z., Zhang, D., Yan, J.: Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification. IJCAI. pp. 2915-2921 (2017)
17. Johnson, R., Zhang, T.: Effective use of word order for text categorization with convolutional neural networks. arXiv preprint arXiv:1412.1058 (2014)
18. Zhang, X., Zhao, J., LeCun Y.: Character-level convolutional networks for text classification. Advances in neural information processing systems. pp. 649-657 (2015)
19. Xiao, Y., Cho, K.: Efficient character-level document classification by combining convolution and recurrent layers. arXiv preprint arXiv:1602.00367 (2016)
20. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
21. Luo, B., Feng, Y., Xu, J., Zhang, X., & Zhao, D.: Learning to predict charges for criminal cases with legal basis. arXiv preprint arXiv:1707.09168 (2017)
22. Hu, Z., Li, X., Tu, C., Liu, Z., Sun, M.: Few-shot charge prediction with discriminative legal attributes. Proceedings of the 27th International Conference on Computational Linguistics. pp. 487-498(2018)
23. Ye, H., Jiang, X., Luo, Z., Chao, W.: Interpretable charge predictions for criminal cases: Learning to generate court views from fact descriptions. arXiv preprint arXiv:1802.08504 (2018)
24. Liu, Y. H., Chen, Y. L., Ho, W. L.: Predicting associated statutes for legal problems. Information Processing & Management. 51(1), pp. 194-211 (2015)
25. Liu, C. L., Liao, T. M.: Classifying criminal charges in chinese for web-based legal services. Asia-Pacific Web Conference. Springer, Berlin, Heidelberg, pp. 64-75 (2005)
26. Tan, M., Santos, C. D., Xiang, B., Zhou, B.: Lstm-based deep learning models for non-factoid answer selection. arXiv preprint arXiv:1511.04108 (2015)