[1]Tianyuan Si

Problem Analysis and Application Research in Legal Judgment Prediction Based on Data Mining

Journal of Electrical Systems

Abstract: - The judgments of some hot cases that exceeded public expectations seem to confirm this. Although the legal certainty is questioned, the judgment remains predictable. The predictability of judgments is a way to enhance judicial authority and maintain judicial credibility. It is also a way to achieve overlapping consensus between the judiciary and the public, provide stable value guidance and behavioral expectations for the public, and promote the generation and development of public rational trust. To enhance public legal trust by improving the predictability of judgments, it is necessary to increase the burden of reasoning for judicial judgments, avoid randomness and contingency, and ensure the adequate provision and substantial disclosure of previous judgment information, so that judgment prediction in the era of big data can truly become possible. The research is to use the optimized particle swarm algorithm as the underlying model to carry out joint modeling and prediction research on the analysis and application of problems in the prediction of legal judgments. According to experimental calculations, the optimized particle swarm algorithm can significantly improve the accuracy and universality of legal judgment prediction. After optimization, the convergence speed is increased by about 5%, and the value of the acceleration factor is more significant.

Keywords: ItemCF; optimized particle swarm algorithm; data mining; legal judgment; legal prediction

I. INTRODUCTION

The judicial field is no exception. In today's society, the means of committing crimes are changing, resulting in more and more complex cases[1]. At the same time, criminal investigation methods are becoming more and more perfect, and the amount of case data is also increasing[2]. The increase in the complexity of cases and the massive growth of data also bring judicial work. With more challenges, judges also face more dilemmas in the process of adapting to the new changes brought about by informatization: First, the cases are becoming more and more complex, and the results of the judgment are difficult to grasp and unify; second, there are a lot of relevant materials in the case, which makes it difficult to sort out and check the facts of the case[3]. Legal judgment prediction technology alleviates the above problems to a certain extent, and is a key technology of the intelligent legal assistant system[4]. On the one hand, it provides a convenient reference for lawyers, judges and other professionals to improve work efficiency. On the other hand, it can avoid the deviation of judgments caused by different judgment scales among different judges, and effectively solve the problems of different judgments in the same case and similar cases being handled differently.[5]

Legal judgment prediction refers to the technology in which machines predict the results of judgments based on the description of the facts of the case[6-8]. Although the machine learning method is effective, its model does not have a clear reflection in the interpretation of prediction results. On the one hand, such as deep neural networks or complex models, this type of method usually has high accuracy, but the internal principles and mechanisms of these methods and models are difficult to understand, and the influence of features on the model prediction results cannot be obtained[9-11]. On the other hand, simple models like linear regression and decision trees usually have better interpretability, but their predictive power is usually limited and their accuracy is lower. This leads to the fact that

[1]School of Marxism, Leshan Vocational and Technical College, Leshan 614000, Sichuan，China

the existing legal judgment prediction research usually has high accuracy, but cannot give a reasonable explanation for the reason of the prediction result, and lack the supporting basis for decision-making.

As far as the judicial field is concerned, the unpredictability of judgments will increase the public's doubts about judicial credibility, cause the consensus of judicial elites and the general public to split on justice issues, thereby weakening social cohesion and affecting the establishment of legal and judicial authority[12-15]. Starting from the level of bridging the gap between the judiciary and the public, enhancing the predictability of judgments and striving for the public's sympathetic understanding of the judiciary has become another important way to boost the public's legal trust. What is especially worth looking forward to is that, after entering the era of judicial big data, powerful storage, extraction and analysis functions such as cloud computing and artificial intelligence have revolutionized the "small data" limitation of sample collection in the past, and provided a new basis for improving the predictability of judgments. "Wind Under the Wings". According to statistics, more than 20 million judgment documents have been collected and published by the China Judgment Document Network. The Supreme People's Procuratorate's "Procuratorial Big Data Resource Database" plan is in full progress, and the "Similar Case Guidance Project" established by the Anhui Provincial High Court has been On the basis of mining and analysis of big data, it is possible to quantitatively analyze a large number of judgment documents with the goal of the same judgment in similar cases. The application of judicial big data will make judgment prediction truly possible, thus The development process of particle swarm optimization analysis and legal judgment prediction[16-18]. The main development directions are as follows[19].

(1) The adjustment of typical molecular group parameters is an algorithm to balance global and local detection capabilities and mining. For example, Shi and Ebert made linear (or nonlinear) biological measurements of inertia according to repeated processes and particle flight conditions, showing the inertia of velocity, that is, the general molecular group close to global equilibrium and studying velocity. Chang et al. (2009) studied the influence of acceleration factors on the predicted location and difference according to the conventional molecular group standard (ECX) prediction location and stability analysis, and obtained a good data set.

For example, X-ray particle swarm optimization is combined with other improved algorithms (or strategies) to form a hybrid form of X-ray particle swarm optimization. For example, Tseng Yi and others are part of the X-ray particle swarm optimization algorithm model. The purpose is to introduce local research capabilities from the research algorithm model and provide global improvement functions that can improve the X-ray observation system. (4) Use surplus technology. Life technology, simulation environment balance, to get the best multimedia and multi problem target function.The termination conditions are shown in Figure 1.

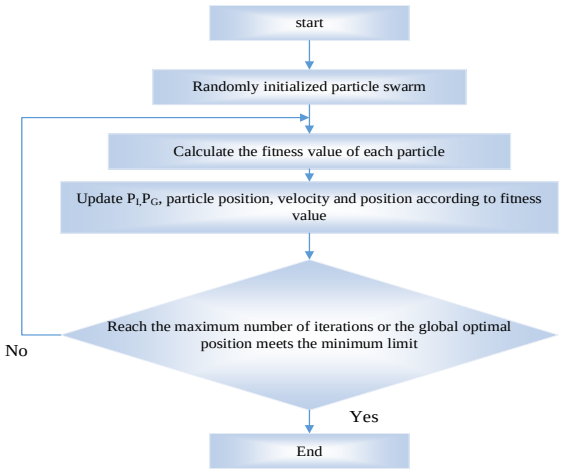

Figure 1. Termination condition of hybrid PSO algorithm

II. RELATED WORK

The importance of the application of model calculation and algorithm replacement technology in law has been deeply rooted in the hearts of the people. Automated legal judgment generation. Many scholars at home and abroad have conducted research from different perspectives and have drawn many valuable conclusions.

Some researchers mainly discussed the problems and countermeasures of the construction of judicial artificial intelligence in my country, and believed that my country needs to vigorously build a model calculation and algorithm replacement system in the legal field to assist judicial personnel to simplify judicial processes and improve work efficiency. For advanced concepts and methods at home and abroad, We should learn with an open mind, so that model calculation and algorithm replacement technology can be better and faster integrated into judicial modernization.

In recent years, neural networks have made great progress in dealing with natural language (neural language programming)[18], text classification and automatic translation[12]. From this point, researchers began to study the prediction of important legal provisions by combining neural network models and legal knowledge. [24], and most of the studies tried Text features are extracted from the case description part.The crime label information, Reference [25] proposed a legal graph network (LegalGraphNetWork) composed of 2 entity types and 4 relation types. It is the adjacency relationship between words and words in the definition of the same crime label, the co-group relationship between the crime label and the crime label with similar attributes, the co-occurrence relationship between the crime label and the crime label corresponding to the same case description, and the words in the definition of the crime label- The component relationship of the count label. Finally, the crime label representation obtained through the legal graph network is integrated into the attention mechanism model for crime prediction. As shown in Figure 2.

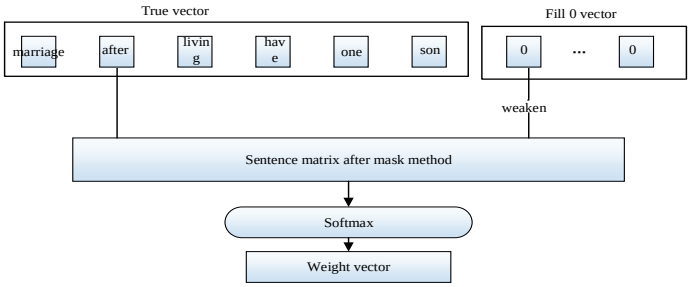

Figure 2 NLP method renderings

To sum up, legal judgment prediction is an application of NLP technology in the judicial field. It can not only provide professional guidance and advice to personnel engaged in legal-related work, but also to ordinary people who lack basic legal knowledge.Appropriate legal advice. Legal judgment prediction has attracted the attention and attention of a large number of scholars, and many new topics have emerged, which is the significance of this research.

III. RESEARCH OBJECTIVE

Matrix-based distributed representation constructs a "word-context" matrix by describing the semantic relationship between words and contexts. Reference [8] proposes a Latent Semantic Analysis (LSA) model, in which singular value decomposition technology is used to project vocabulary into a latent low-dimensional semantic space, thereby reducing the dimension of feature representation and eliminating some noise. The left and right singular matrices in LSA can capture the correlation between word-topic and topic-document respectively, which basically solves the problem of polysemous words, but cannot distinguish polysemy problems well, and it is difficult to explain singular value decomposition. The meaning of negative values appearing in . In this regard, the literature [21] proposes

Probabilistic Latent Semantic Analysis (PLSA), which treats both documents and vocabulary as random variables and introduces them into the latent topic layer. Specifically, the PLSA model gives PLSA has the ability to distinguish polysemy problems, but due to the lack of assumptions about the prior distribution of topics, the parameters increase linearly with the number of training samples, and even overfitting occurs. In response to the above problems, the literature proposes the Latent Dirichlet Allocation (LDA) model, which introduces the Dirichlet distribution of topics and vocabulary on the basis of PLSA, so that PLSAThe medium topic distribution parameters are determined, thus solving the overfitting problem of the PLSA model. However, the matrix factorization technique has a general effect on the linear analogy task of words, which means that its vector space structure is not optimal.

3.1 Theoretical Premise

The data set used in this paper is the data set challenging the "China legal research" judicial artificial intelligence cup (cail2018) [14]. Cail 2018 includes more than 2.6 million criminal cases announced by the Supreme People's Court of China. Compared with other datasets used in legal judgment prediction work, CAIL2018 is larger in scale and has several times the amount of data. In addition, CAIL2018 has more detailed and richer annotations on the verdict. Each case is composed of two parts: a description of the facts and the corresponding verdict. The verdict includes the applicable legal terms, charges and sentences, which are based on the factual description of the case. inferred. CAIL2018 can be used as a reference for professionals to improve work efficiency, and has a certain reference value for the study of legal intelligent systems. As shown in Table 1.

Table 1 Examples of CAIL2018 datasets

Factual description	On the morning of November 5, 2015, the defendant Hu had a quarrel with the victim sun in the workshop of Jiaxing duoling Jinniu Clothing Co., Ltd. in Zhapu Town, Pinghu City, over trivial work. Later, the defendant Hu injured the victim sun's left abdomen with a wooden cushion. The Pinghu public security Judicial Expertise Center identified that sun's left abdominal injury had reached the second level of serious injury.	
Judgment result	defendant	Hu Mou
	charge	Intentional injury
	Relevant provisions	Article 234 of the criminal law
	Term of imprisonment	12 months

3.2 Model Construction

How to judge the relevance of pending cases to past cases to be predicted? The essence of judging the similarity of cases is a kind of analogical reasoning, that is, people deduce the similarities that two things may have in other aspects according to the similarities in some aspects. The conclusion of this logical reasoning is highly probabilistic: both objects A and B have a, b, and c, and object d at the same time, and it is uncertain whether object B must have attribute d. There are no two identical leaves in the world, and the phenomenon of "different judgments for the same case" is not uncommon in practice. However, the implementation of the case guidance system clearly room for increasing the predictability of judgments.

The supervised keyword extraction method, its advantages of easy calculation and simple operation make it useful in fields such as search and document classification.The domain is very widely used, mainly used to judge the effect of a single word on a text in a dataset or corpus Whether it is critical and the critical size is quantified. Its value is used to measure the weight of words in the field of information retrieval.Importance is a measure of the amount of

information a term can provide in a query. Due to the algorithmic assumption: in the given textThe information value of words with more occurrences and less occurrences in the corpus for distinguishing the text istallest. So the main idea of the algorithm is: the criticality of words and their occurrence in a given textIt is proportional to the number of occurrences.

Frequency (task summary) indicates the frequency of words in the text:

$$TF_{w,D_i} = \frac{\text{count}(w)}{|D_i|} \tag{1}$$

Inverse Document Frequency (IDF) is just the opposite. The IDF value is inversely proportional to texts in which the word. The calculation formulas (2) and (3) are as follows:

$$IDF_w = \ln \frac{n}{\sum_{i=1}^{n} I\left(w, D_i\right)} \tag{2}$$

$$I\left(w, D_i\right) = \begin{cases} 1 & w \in D_i \\ 0 & w \overline{\in} D_i \end{cases} \tag{3}$$

When the word w does not appear in any document in the dataset, then formula (2) is meaningless, so the calculation of the IDF value isThe calculation is slightly modified, as shown in formula (4): $IDF_w = \ln \frac{n}{1+\sum_{i=1}^{n} I\left(w, D_i\right)}$

(4)

Thus, the TF-IDF value of the text can be calculated according to formula (5).

$$\text{TFIDF}_{w,D_i} = TF_{w,D_i} \times IDF_w \tag{5}$$

IV. METHODS

The basic idea optimization: firstly, the text is multiple units (words, phrases, etc.) and a word graph model is established with this as a node, and then the text units are sorted according to their criticality. The sorting rule is a voting mechanism, that is, a The criticality of a cell is determined by the sum of the criticalities of the cells that point to it. Because it does not require corpus learning and training, the keyword extraction process can be completed in the text itself, which is concise and efficient. Its underlying logic is shown in Figure 3:

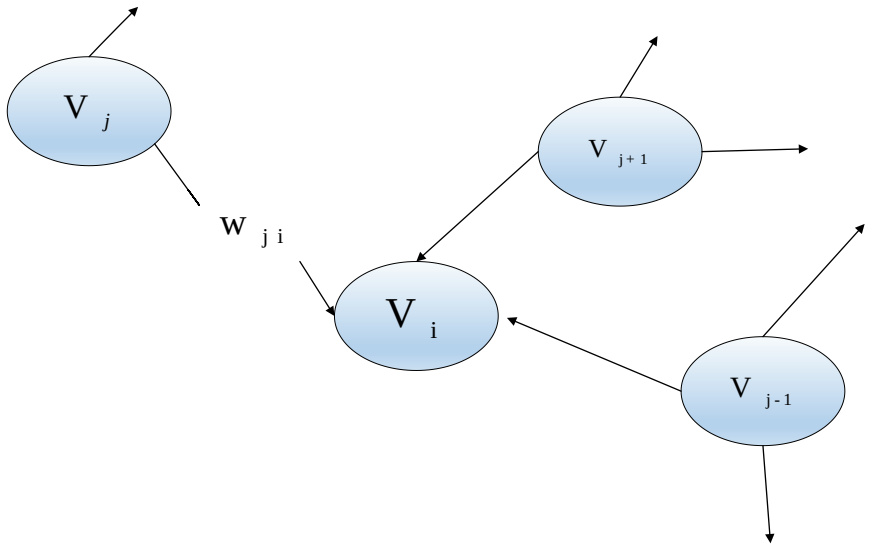

Figure 3 The underlying logic diagram of the algorithm

3.1 Text Classification

Text rendering is a method to convert text into computer understanding, and it is the most important part to determine the quality of text. The second chapter provides search content. Most text classification functions (i.e. text groups) show complex multi-level repetition functions, which will have a negative impact on the performance of text classification. Therefore, it is necessary to take advantage of the advantages of choosing more different and representative text expression methods. The selection method can not only reduce disasters to a certain extent, but also reduce the difficulty of text classification while eliminating independent functions. According to the form of topic selection, the selection method can be divided into three categories: filtering, packaging and packaging. This is part of the selection function.As shown in Figure 4.

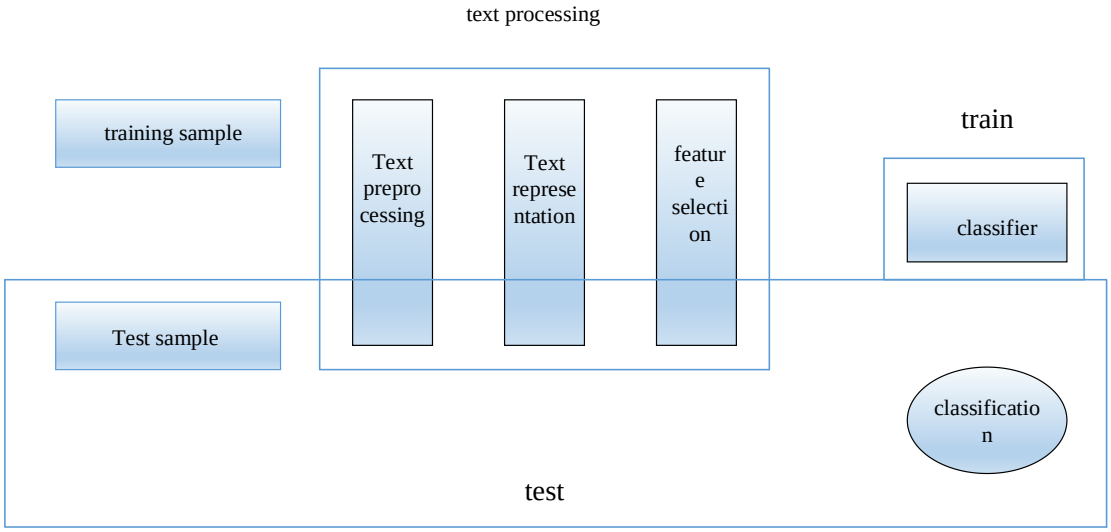

Figure 4 The underlying logic diagram of text classification

3.2 Extraction of decision elements

In order to facilitate model processing, this chapter converts fact description paragraphs in the dataset into corresponding multiple fact description sentences, combines multiple labels into a single label, and adds “0” labels

to correspond to non-element sentences. The processed experimental data are shown in Table 2. Show.

Table 2 Number of samples of each type of dataset

data	Training set	Validation set	Test set	Number of tags
Divorce	30525	3815	1500	36
Labour	26394	3300	1500	38
To loan	17825	2228	1500	39

In order to improve the quality of education data and reduce the imbalance of the data group, this chapter excludes the "monthly pension" of divorce data and the "job introduction arbitration stage" of work data. About 0.1% of names and data shows the number of samples and tags of the processed dataset. According to the name of each data and the ratio of total data, 1500 data are divided into three groups to evaluate the test results, and other data are subdivided into training and group verification at the ratio of 8:1.Its model structure is shown in Figure 5.

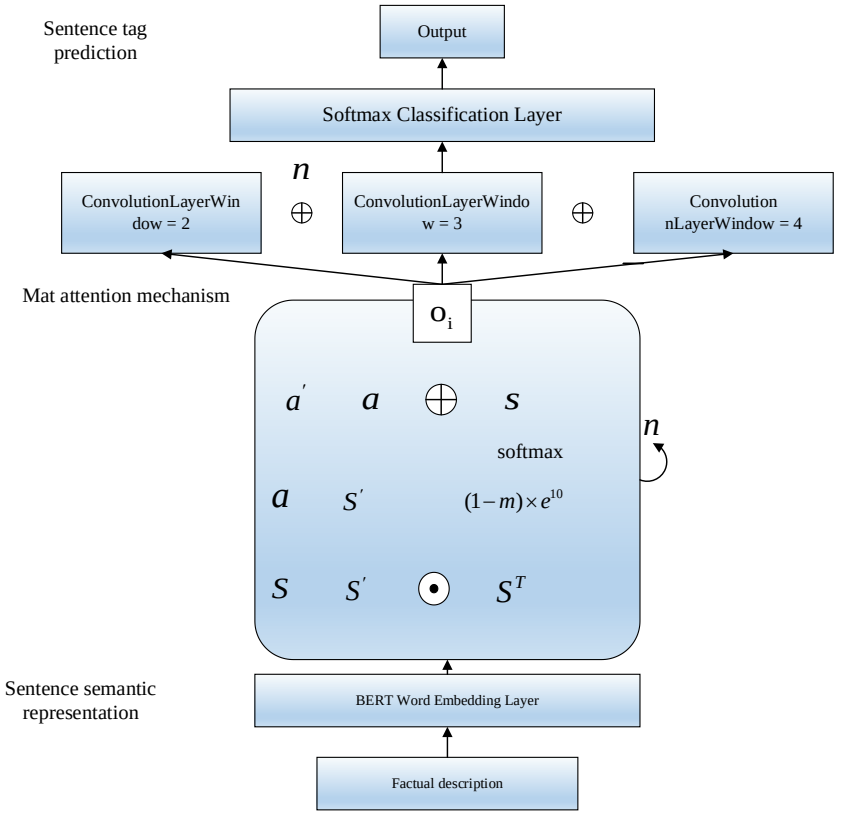

Figure 5. Structure diagram of decision element extraction model

V. CASE STUDY

This chapter uses a convolutional neural network with multiple convolution kernels. Synaptic connection uses nuclei to extract local information, which has more accurate kernel fusion function.The size of the information range. The matrix is generated using three filters to generate local features, as shown in Equation (6):

$$C_{i,j} = f\left(W \cdot H^{att}_{(i,i+h-1)} + b\right) \tag{6}$$

Based on this formula, the optimization frequency upgrade of its particle algorithm can be realized, as shown in Figure 6.

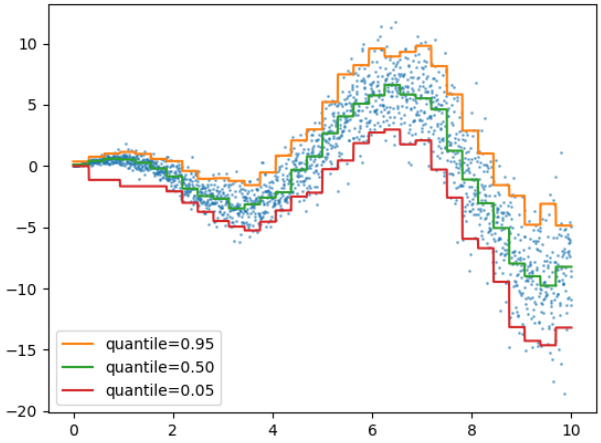

Figure 6 Optimization frequency upgrade of optimized particle swarm optimization

5.1 Optimization of Scaled Dot Product Attention

Through X, K and various linear transformations, product input is the focus of maximization, product output is the focus of maximization, the focus of maximization is mixing, followed by the final linear transformation. This value is the result of Sue's interest. 10. After k-linear transformation, each ADHD calculation method is different, so the model can only calculate once, not once, so the model can focus on the information reflecting different trends in different micro spaces. Leaders' attention is equal to the average representation of all micro spaces, which limits their ability to learn in different micro spaces.Since self-attention is calculated for each word and all words in the sequence, the distance between any positions in the sequence is reduced to a constant, which can well capture long-distance features. At the same time, because it does not rely on the calculation of the previous moment, it also has a strong parallel ability, which solves the shortcomings of RNN and CNN well. The optimization process is shown in Figure 7.

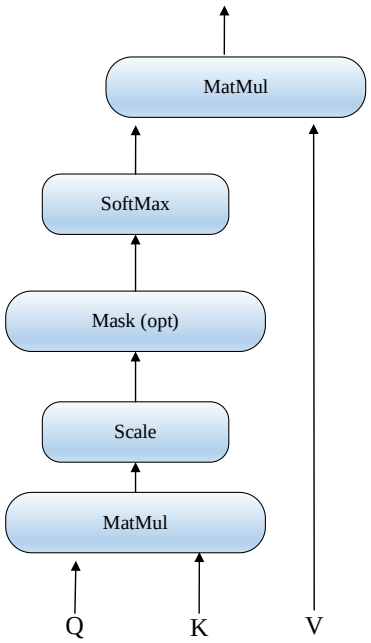

Figure 7 Schematic diagram of the optimization of zoom point and attention

5.2 The efficiency of coding and extraction of legal articles is improved

Through the extraction stage of relevant laws, k relevant laws are obtained for each given case description, and the text information of these relevant laws is encoded by the DPCNN model to obtain their relevant laws. As can be seen from the introduction in Chapter 2, the architecture of the DPCNN model is mainly divided into three parts: The first layer is a region embedding layer, which aims to convert a generic word vector into a text region vector covering one or more words; followed by a stack of convolutional units and pooling layers, where the convolutional unit includes two convolutional layers and a residual connection, and the pooling layer is used for downsampling; the final pooling layer aggregates the internal data of the text into a vector for classification. The network structure of DPCNN is shown in Figure 8.

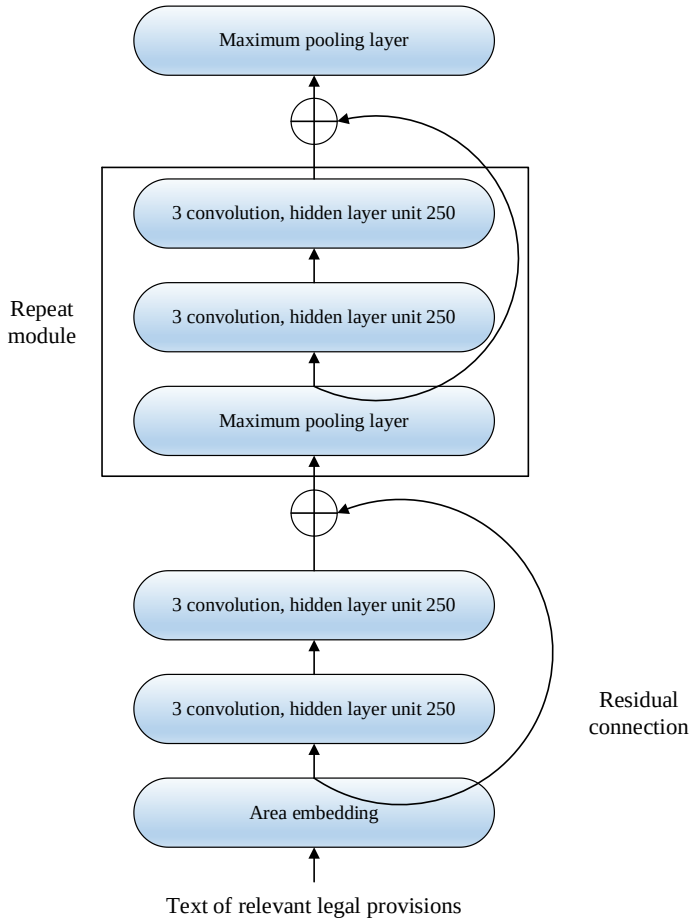

Figure 8 DPCNN network structure diagram

5.3 Entry confusion efficiency improvement

Table 3 lists the correspondence between some typical crimes and specific keywords. Therefore, this paper adopts the method of incorporating specific keyword information related to crimes into the model to solve the problem of confusing crimes.

Table 3 Confused terms for some typical crimes

Charge	Key word
Robbery	Robbery, robbery, interception, fruit knife, dagger, robbery, resistance
To rob	Snatch, handbag, satchel, snatch, snatch, seize, follow
Intentional injury	Intentional injury, dispute, trivia, quarrel, fight,

	cutting nasal bones and ribs
Intentional homicide	Neck, stabbing, chopping, killing, stabbing, abdomen, sharp tools, counting knives, fierce stabbing, continuous stabbing
Damage to vehicles	Tires, engines, nails, spikes, leech nails, brakes
Damage to traffic facilities	Railway, expressway, signboard, rail, guardrail, pole adjuster
Traffic accident	Car accident, accident scene, traffic accident, overturning, falling, touching, hanging and knocking down
Dangerous driving	Blood, ethanol, drunkenness, alcohol, concentration, exhalation, blood sample, threshold, detector

VI. CONCLUSION

For legal professionals, it can improve their work efficiency, give more professional legal advice, avoid subjective interference factors, and achieve a certain degree of judicial fairness and openness; for the public, they can have a broad understanding of legal knowledge, understand the situation of the case, and obtain a certain degree of justice. legal guidance. This paper uses deep learning technology to study the task of legal judgment prediction. By analyzing the relationship between relevant laws and judgment elements, a judgment element extraction method incorporating legal information is proposed. By integrating the semantic weight of the law into the fact description sentence, the decision element extraction task is guided by the law information, performance of the decision element extraction task is improved.

Data Availability

The experimental data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declared that they have no conflicts of interest regarding this work.

Funding Statement

There is no specific funding to support this research.

REFERENCES:

[1] Wang, J. (2021). Research on data mining and prediction of large-scale competitions based on online data migration model. *Journal of Intelligent and Fuzzy Systems*(2), 1-12.

[2] Gong, B. (2022). Convergence analysis of two finite element methods for the modified maxwell's steklov eigenvalue problem. *ESAIM: Mathematical Modelling and Numerical Analysis, 56*(1), 287-301.

[3] Zhang, X. Shen, J. Wu, P. & Sun, D. . (2021). Research on the application of big data mining in the construction of smart campus. *Open Access Library Journal, 8*(11), 10.

[4] Zhang, Y. K. Gong, G. F. Yang, H. Y. , Chen, Y. X. , & Chen, G. L. . (2022). Towards autonomous and optimal excavation of shield machine: a deep reinforcement learning-based approach. *Journal of Zhejiang University-SCIENCE A, 23*(6), 458-478.

[5] Ye, S. Nedzved, A. Chen, C. Chen, H. Leunikau, A. , & Belotserkovsky, A. . (2022). Shadow detection on urban satellite images based on building texture. *Pattern Recognition and Image Analysis, 32*(2), 332-339.

[6] Yanping, C. & Linlin, C. (2022). Research on face-voiceprint correspondence learning with aux-output supervised convolution neural network. *Pattern Recognition and Image Analysis, 32*(2), 448-455.

[7] Egurnov, D. A. & Ignatov, D. I. (2022). Triclusters of close values for the analysis of 3d data. *Automation and Remote Control, 83*(6), 894-902.

[8] Afanasieva, T. V. (2022). Granulation of multivariate time series in the problem of descriptive analysis of the state and behavior of complex objects. *Automation and Remote Control, 83*(6), 884-893.

[9] Chang, J. Xu, M. & Xue, R. . (2022). Public auditing protocol with dynamic update and privacy-preserving properties in fog-to-cloud-based iot applications. *Peer-to-Peer Networking and Applications, 15*(4), 2021-2036.

[10] Ellouze, M. & Hadrich Belguith, L. (2022). A hybrid approach for the detection and monitoring of people having personality disorders on social networks. *Social Network Analysis and Mining, 12*(1), 1-17.

[11] Yemm, L. (2022). Design and analysis of the extended hybrid high-order method for the poisson problem. *Advances in Computational Mathematics, 48*(4), 1-25.

[12] Heid, P. & Wihler, T. P. (2022). A modified kaanov iteration scheme with application to quasilinear diffusion models. *ESAIM: Mathematical Modelling and Numerical Analysis, 56*(2), 433-450.

[13] Jingchun Zhou, Boshen Li, Dehuan Zhang, Jieyu Yuan, Weishi Zhang, Zhanchuan Cai. "UGIF-Net: An Efficient Fully Guided Information Flow Network for Underwater Image Enhancement," IEEE Transactions on Geoscience and Remote Sensing, vol. 61, pp. 1-17, 2023, Art no. 4206117, doi: 10.1109/TGRS.2023.3293912.

[14] Ali, J., Shan, G., Gul, N., & Roh, B. H. (2023). An Intelligent Blockchain-based Secure Link Failure Recovery Framework for Software-defined Internet-of-Things. Journal of Grid Computing, 21(4), 57.

[15] Ren, X., Ahmed, I., & Liu, R. Study of Topological Behavior of Some Computer Related Graphs. Journal of Combinatorial Mathematics and Combinatorial Computing, 117, 3-14.

[16] Nazeer, S., Sultana, N., & Bonyah, E. Cycles and Paths Related Vertex-Equitable Graphs. Journal of Combinatorial Mathematics and Combinatorial Computing, 117, 15-24.

[17] Zeng, Y., & Chu, B. The Appropriate Scale of Competition Between Online Taxis and Taxis Based on the Lotka-Volterra Evolutionary Model. Journal of Combinatorial Mathematics and Combinatorial Computing, 117, 25-36.

[18] Luo, X., Zhang, C., & Bai, L. (2023). A fixed clustering protocol based on random relay strategy for EHWSN. Digital Communications and Networks, 9(1), 90-100.

[19] Chunjiong Zhang, Byeong-hee Roh, and Gaoyang Shan. 2023. Poster: Dynamic Clustered Federated Framework for Multi-domain Network Anomaly Detection. In Companion of the 19th International Conference on emerging Networking EXperiments and Technologies (CoNEXT 2023). Association for Computing Machinery, New York, NY, USA, 71–72. https://doi.org/10.1145/3624354.3630086