

Research Article

Artificial Intelligence-Assisted Decision-Making Method for Legal Judgment Based on Deep Neural Network

Wenqing Ma 

Gansu University of Political Science and Law, Gansu, Lanzhou 730070, China

Correspondence should be addressed to Wenqing Ma; mwq6824@163.com

Received 19 July 2022; Revised 9 September 2022; Accepted 26 September 2022; Published 11 October 2022

Academic Editor: Xuefeng Zhang

Copyright © 2022 Wenqing Ma. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the arrival of the third revolution of artificial intelligence, the applications of artificial intelligence in the fields of automatic driving, image recognition, smart home, machine translation, medical services, e-sports, and so on can be seen everywhere, and topics about artificial intelligence are constantly emerging. Since 2017, the discussion on artificial intelligence in the field of law has become more and more active. In this context, the application of artificial intelligence in the field of legal judgment and the hypothetical system based on this technology in court judgment has also become the object of discussion from time to time. In this paper, based on the artificial intelligence decision-making method of the deep neural network, aiming at the three subtasks of legal judgment prediction, namely, crime prediction, law recommendation, and sentence prediction, a multi-task judgment prediction model BERT12multi and a sentence interval prediction model BERT-Text CNN are proposed, which improve the prediction accuracy and adopt the knowledge distillation strategy to compress the model parameters and improve the reasoning speed of the judgment model. Experiments on the CAIL2018 data set show that the performance of the deep neural network model in crime prediction and law recommendation tasks can be significantly improved by adopting the pre training model adaptive training, grouping focus loss, and gradient confrontation training strategies. Using a step-by-step sentence prediction strategy can realize the weight sharing of pre training model and make use of the prediction results of charges and laws in sentence prediction. The recall training-prediction strategy can avoid error accumulation and improve the accuracy of sentence prediction. By integrating the artificial intelligence decision-making method, the case reasoning speed can be greatly improved, the highest compressible model volume can be about 11% of the original one, and the reasoning speed can be increased by about 8 times. At the same time, performance close to that of the deep neural model can be obtained, which is superior to other legal decision prediction models based on word embedding.

1. Introduction

With the advent of the Internet and the big data era, the development of all walks of life is more and more combined with computer science. As the most cutting-edge technology of computer science, artificial intelligence is also frequently combined with the industry in the era of network development. In recent years, the combination of the judicial field and artificial intelligence technology is generally favored by the outside world. The combination of the two is an innovative exploration of the application of artificial intelligence technology, which is conducive to improving the level of judicial reform. The reason why it is generally optimistic is that the application of artificial

intelligence technology improves the efficiency of judicial adjudication, and further saves manpower and time [1]. However, due to the limitation of the current level of information network technology, there are still many problems in the application of artificial intelligence technology in the field of legal judgment. Legal judgment prediction is the most typical application of artificial intelligence technology, especially natural language processing methods in the judicial field. The legal judgment prediction task generally includes sub tasks such as crime prediction, relevant law prediction, criminal sentence prediction, and so on. Through the study of legal materials such as adjudicated historical cases and published judgment documents, the prediction model is constructed by

using machine learning or a deep learning algorithm to infer the judgment results.

Aiming at crime prediction, legal recommendation, and judicial decision prediction, this paper proposes a hybrid deep neural network model HAC (hybrid attention and CNN model) for long text classification. And Deep Pyramid Convolutional Neural Networks. The model's prediction of crime and the recommended F1-Score (mean of Micro-F1 and Macro-F1) for the relevant statutes are 85% and 87%, respectively. For the prediction of the sentence, due to the differences in regions, ages, courts, judges, and defendants' attitudes, it will become more difficult to predict judicial decisions [2]. The model has excellent predictive performance and generalization ability and can adapt well to these differences. At the same time, the output results of the model's crime prediction and legal recommendation are added to the input of the judicial decision prediction task, and the classification method is used to predict the judicial decision, which further improves the effect of the model. Finally, in the sentence prediction task F1-Score was more than 77% and obtained excellent results predicted by the CAIL2018 judicial ruling.

It is undeniable that the application of artificial intelligence technology has further improved legal efficiency and saved legal resources, but there are still many drawbacks in terms of the application status of artificial intelligence. On the one hand, we have high expectations for the promotion of artificial intelligence technology; on the other hand, we are troubled by the threat theory of artificial intelligence. Whether the bottleneck of artificial intelligence technology in the application of legal judgment will hinder further integration and development in the legal field poses a severe challenge to both the theoretical circle and the expert group [3]. Therefore, scholars and specialized institutions are committed to judicial reform, aiming at realizing a networked and intelligent legal adjudication system by studying the combination of judicial and artificial intelligence technology. The innovation of this paper is the following: First, by analyzing the specific practice and theoretical research of artificial intelligence in legal judgments at home and abroad, we will gain insight into the loopholes existing in artificial intelligence in legal judgments in our country, and put forward relevant suggestions for the problems that arise. Second, the status and role of artificial intelligence in the application of legal judgments are demonstrated from three aspects: legislation, judiciary, and court system; evidence. By demonstrating the rationality of the application of artificial intelligence in the judicial field of our country, we will discuss the development direction of artificial intelligence technology in the field of law in China in the future, and gradually improve the new development model of the combination of law and computer science in my country.

This article is organized into seven chapters. The first chapter is the introduction part. This part analyzes the application of artificial intelligence in the field of legal judgment and the current situation of the hypothetical system based on the practice of this technology in court judgments summarizes the reasons for the problems and compresses new models to improve the judgment model.

Inference speed. The second chapter mainly summarizes the relevant literature, summarizes the advantages and disadvantages, and puts forward the research ideas of this paper. The third chapter introduces the application characteristics of artificial intelligence in legal adjudication in detail. Chapter 4 discusses pre trained models based on deep neural networks. The fifth chapter expounds on the application of artificial intelligence to assist decision-making in legal judgments, which is enough to prove the positive significance of the application of artificial intelligence technology to judicial construction. The sixth chapter is the part of the experimental results, which analyzes the research results in detail and the outlook based on the results. The seventh chapter is the conclusion, summarizing the significance of the research.

2. Related Work

The most far-reaching impact of training word vector models with neural networks is the Neural Network Language Model (NNLM) proposed by Chen et al. The NNLM model is based on the n-gram language model and uses a three-layer linear neural network. 1 word is used as input to predict the nth word [4]. This method of word vector training lays the groundwork for the technical direction of word vector training, and later generations of research on word vector training methods are mostly inspired by this. In their published papers, Hong et al. proposed the idea of using word vectors to solve various tasks such as part-of-speech tagging and named entity recognition in natural language processing. The open-source system SENNA [5]. According to the idea of transfer learning and the use of Transformer structure, researchers proposed a large-scale pre training model represented by Schemmer et al. [6]. The pre training model generally adopts a multi-layer Transformer structure to perform unsupervised pre training on a large corpus. In practical applications, the pre training model has the defects of low computational efficiency and large resource consumption. The researchers mainly use strategies such as knowledge distillation to compress the volume of the pre training model. Rastogi et al. proposed a Text CNN text classification model based on a one-dimensional convolution structure, which first maps the text into vectors and then uses one-dimensional convolutions of different sizes to capture the local semantic information of the text and capture the local semantic information of the text through pooling. Important feature information is input to the classifier to obtain the probability distribution of the labels [7]. Gerards and Borgesius proposed a text classifier based on a deep convolutional structure. By increasing the network depth, the vector representation of the text is refined for classification [8]. The patient knowledge distillation strategy (Patient Knowledge Distillation, PKD) proposed by Wu and School for pre training models such as BERT, extracts rich information through the deep structure of the teacher network and adds supervision from the middle layer of the teacher model in the distillation process, which improves the depth. The performance of the network model [9]. Qian et al. designed a dual feedback mechanism with

multi-view forward prediction and backward verification to match the numerical collocations in the text, aiming at the numerical unit keywords in the legal text, such as the content of alcohol, the weight of drugs, and the amount of theft. It improves the ability of the model to capture the matching information of numbers and keywords [10]. Solum used a keyword extraction algorithm to mine judgment documents to extract the keywords of crimes, integrated the deep learning model, and proposed the MTL-Fusion model. It effectively improves the model's ability to distinguish easily confusing charges [11]. Walker et al. applied the deep learning text classification model HAN and DPCNN to legal judgment prediction, integrated and improved the model for judicial intelligence applications, and proposed a legal judgment prediction model HAC (Hybrid attention and CNN model) [12]. Applying the pre training model to the legal judgment prediction task can further improve the performance of the judgment prediction model and improve the reliability and accuracy of the prediction. Brennan-Marquez and Henderson proposed a BERT-based HIER-BERT model to handle very long case texts [13].

3. Application Characteristics of Artificial Intelligence in Legal Adjudication

3.1. Electronicization of Legal Information. In recent years, the electronic application of data and information has brought technological innovation to the court. The rise of the "smart court" is the technical product of the combination of artificial intelligence and the court system. The combination of artificial intelligence and judicial technology makes judicial judgment have the characteristics of electronic information. Through the data collection and analysis of laws and cases, artificial intelligence can process judicial data more scientifically and accurately [14]. The exertion of this feature allows the judge to find the necessary theoretical basis at any time according to the circumstances and needs of the case. Through intelligent analysis and statistical data, it not only saves a lot of time for investigation and data search but also prevents information omission. The application of this technology also provides convenience for the smooth development of judicial adjudication, such as intelligent identification of the parties, and retrieval of all relevant information. Finally, the application of artificial intelligence makes the office of judicial adjudication intelligent and further liberates manpower [15].

3.2. Intelligence to Prevent Judgment Risks. Artificial intelligence can prevent the judgment trap of "similar cases with the same judgment" to the greatest extent. Through the judge's ability to analyze massive case files using artificial intelligence technology, it can help judges identify the differences between massive cases and check the differences between similar cases, preventing the judgment trap of "similar cases with the same judgment." Secondly, artificial intelligence can prevent data loopholes through the use of artificial intelligence technology to supervise judicial personnel, intelligently monitor the process of legal judgment,

monitor the illegal behavior of judges in real-time, and prevent judicial personnel from operating under the shadows to affect judicial justice and cause unnecessary legal judgments, anthropogenic risk [16]. In this way, judicial corruption can be effectively suppressed and the risk of legal judgment can be reduced, as shown in Figure 1.

4. Pre Training Model Based on Deep Neural Network

4.1. Pretrained Model Encoder. In this paper, the BERT pre training model is used as the encoder of the model to enhance the performance of the model, and further adaptive training is carried out on the CAIL2018 data set to enhance the performance of the pre training model. The subtasks of crime prediction and law recommendation are classified by multi-labels, two subtasks are jointly trained in one model, and each category of crime and law is classified by binary [17]. The categories of crimes and laws are grouped according to the number of samples, and the focus loss of grouping is used to improve the classification ability of the model for unbalanced samples. In the fine-tuning of the pre training model, gradient confrontation training is adopted to further improve the performance of the pre training model. The algorithm flow is shown in Figure 2.

For the legal judgment prediction task, due to the complex content of the judgment document, the defendant's behavior differences, the items involved and other information will directly affect the judgment results, and there are high requirements for the information extraction ability of the model [18]. And in the task of accusation prediction and related law prediction, the number of samples of some categories is scarce. If the model of traditional word vector combined with neural network structure is used, it is difficult to accurately understand its semantic information only through a small number of samples in the training set. Using the pre training model can effectively enhance the performance of the model, and the pre training model has strong generalization, which is conducive to the model's understanding of some rare categories of samples [19].

In the current research, the encoder or decoder in the Transformer structure is often applied independently to extract the features in text or sequence data. For example, the Bert and GPT prep models adopt the transformer multilayer encoder and decoder structure as a hidden layer, as shown in Figure 3.

The Transformer structure abandons the convolutional neural network structure and the recurrent neural network structure and adopts the self-attention mechanism to extract the information in the sequence data [20]. First, map the sequence into three sets of vectors of key (Key), value (Value), and query (Query) through three sets of neural networks with different weights, and then map the query vector of each unit in the sequence to the key vector of all units in the sequence. The scaling dot product is calculated, and the attention score is calculated through the softmax function mapping. The final encoded hidden representation at a certain position is jointly determined by the attention score and the value vector, namely,

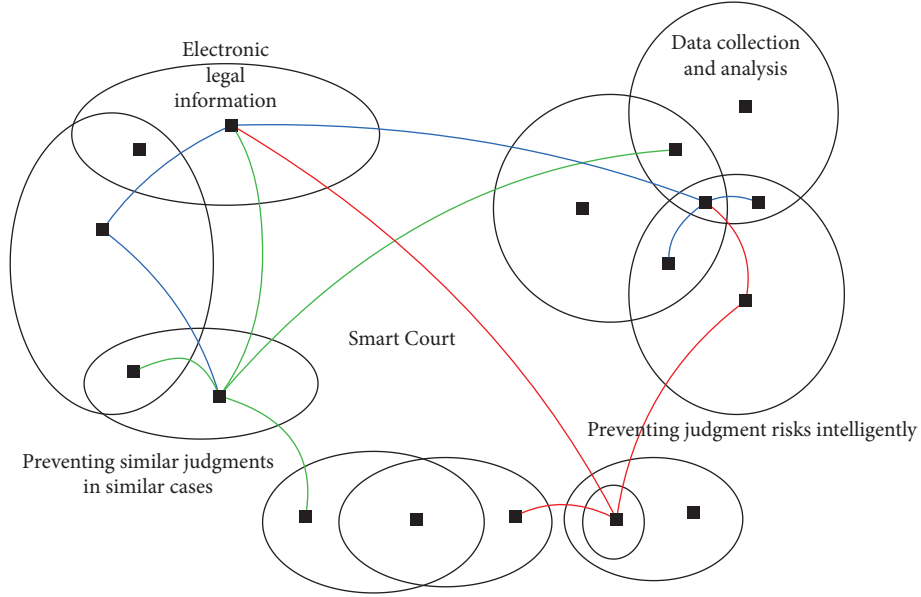


FIGURE 1: Application features.

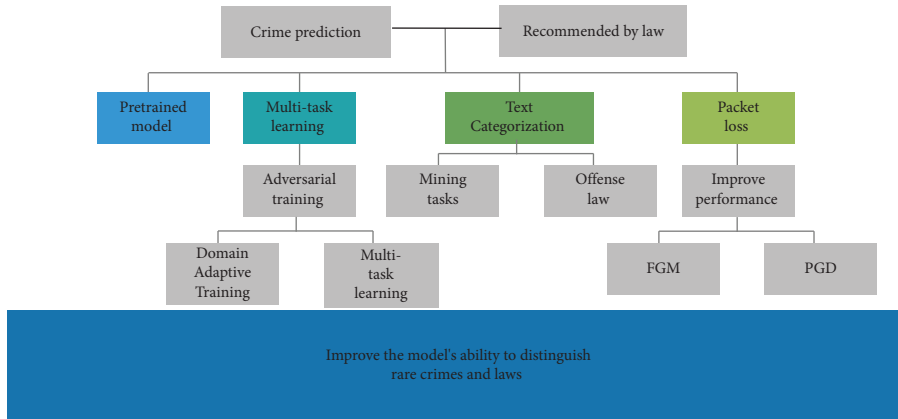


FIGURE 2: Algorithm flow chart.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK}{\sqrt{d_k}}\right)V. \quad (1)$$

In practical use, the Transformer structure adopts a multi head attention mechanism similar to the multi convolution kernel structure in a convolution neural network, which further improves the nonlinearity of the model and enables the model to pay attention to multi-level semantic information, namely,

$$\begin{aligned} \text{head}_i &= \text{Attention}(QW_i, KW_i, VW_i), \\ \text{MultiHead}(Q, K, V) &= \text{Concat}(\text{head}_1, \text{head}_2, \dots, \text{head}_h). \end{aligned} \quad (2)$$

Compared with the mainstream recurrent neural network structures such as (RNN, GRU, and LSTM) used for text or sequence data modeling, the Transformer structure does not have the risk of gradient disappearance, has a stronger ability to capture long-distance dependencies, and

can be calculated in parallel, which can effectively improve the efficiency of training and reasoning, but the calculation of the self-attention mechanism in the Transformer has a high space complexity and requires high memory resources in the calculation.

4.2. Multi-Task and Multi-Label Text Classification Model.

For the task of crime prediction and law recommendation, in studies, only one crime or law regulation was often considered, and it was regarded as a multi-category text classification task. In practice, the trial of a case often involves several related laws, and the defendant has the situation of combined punishment for several crimes. Therefore, the task of crime prediction and law recommendation is treated as a multi-label text classification task, and all crimes and law categories are classified in binary, so as to calculate the probability that each judgment document belongs to each category.

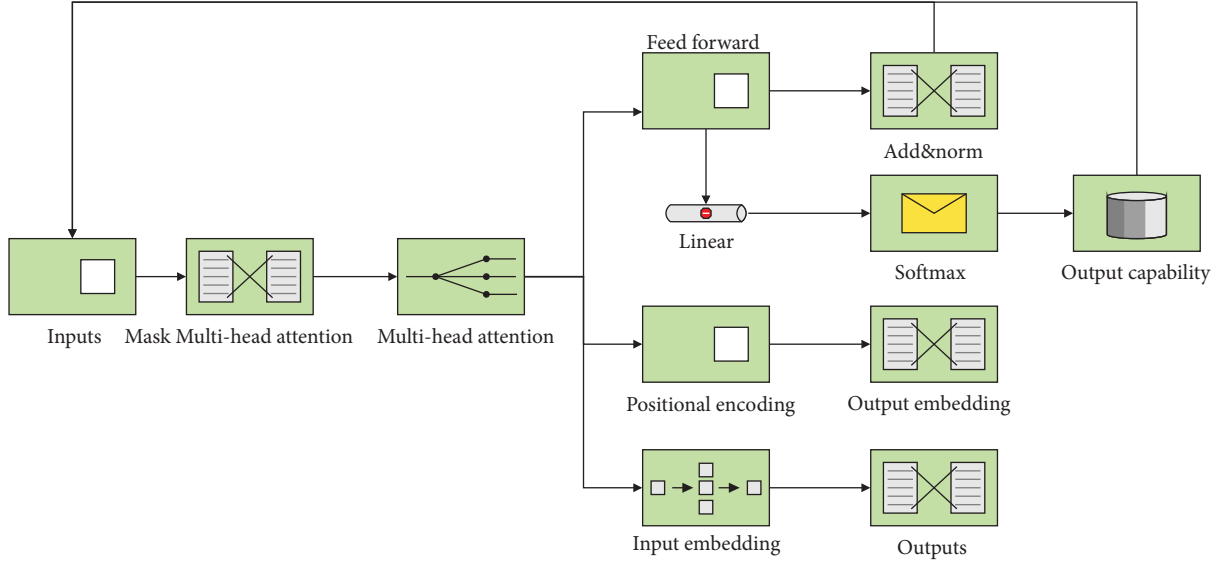


FIGURE 3: Schematic diagram of the Transformer structure.

Completes or truncates the preprocessed case text into a uniform length, converts it into an index sequence matrix according to BERT's pre training vocabulary, inputs BERT models in batches, and uses the output vector corresponding to the special symbol position of (CLS) after coding by BERT model as the sentence vector of the input text to input into the classifier. After dropping, the text sentence vector is mapped into the corresponding category by the classifier, namely,

$$\begin{aligned} h_i &= \text{BERT}(x_i, \theta), \\ y_i &= \delta(\text{Dropout}(W \cdot h_i)). \end{aligned} \quad (3)$$

Here, h_i represents the output vector of the sample x_i input BERT model at the corresponding position of the (CLS) mark; W is the weight matrix of the classifier; $\delta(x)$ represents the sigmoid activation function, and y_i represents the class probability predicted by the model for sample x_i . This multi-task learning model with 12 hidden layers is named BERT_{12multi} to distinguish it from the student model that uses fewer layers of Transformer encoding layers in knowledge distillation.

In the training of the neural network, for binary or multi-label classification problems, the binary cross-entropy (Cross Entropy Loss) is generally used as the loss function, namely,

$$\begin{aligned} L_{\text{CE}}(x_i) &= -\log y_i, y_i = 1, \\ L_{\text{CE}}(x_i) &= -\log((1 - y_i)). \end{aligned} \quad (4)$$

The two-category cross-entropy loss function has no tendency to positive and negative samples. When the negative samples in the data set are far more than the positive samples, the ordinary cross-entropy loss function will make the classification results tend to be negative samples. Moreover, for multilabel or multiclassification, the problem is that the uneven distribution of samples in different categories will cause the classification results to tend to the

categories with more samples so that the recall rate of the categories with fewer samples can be lower than that of the categories with more samples, which affects the final classification results.

4.3. Adversarial Training. In the field of computer vision, a certain amount of noise interference is often added to the picture samples of the training set to carry out a certain degree of anti disturbance training, enhance the ability of the model to resist disturbance, reduce the dependence of the model on the original training set, make the model more robust, have higher anti disturbance ability, and also play the role of expanding data to a certain extent. Image data itself can be regarded as a matrix, and each pixel can be regarded as a group of vectors. However, for text data, because the character representation is discrete, it is impossible to directly add disturbance in the field of computer vision. In natural language processing tasks, it is often necessary to vectorize the discrete character representation through the embedding layer, so the gradient confrontation training can be realized by disturbing the embedded layer according to the gradient direction of the embedded layer. Confrontation training can generally be expressed in the following format:

$$\min_{(x,y) \sim D} [\max_{\Delta x \in \Omega} L(x + \Delta x, y, \theta)]. \quad (5)$$

Among them, D represents the training set, x represents the input of a single sample, Δx represents the applied perturbation, L represents the loss function of the sample x , and Ω represents the perturbation space. Common adversarial training methods in natural language processing are FGM (Fast Gradient Method) and PDG (Projected Gradient Descent).

The FGM adversarial training strategy is as follows:

The sample x mapped by the embedding layer, through forwarding propagation, calculates the loss function, and obtains the gradient g of x :

$$g = \Delta_x L(\theta, x, y). \quad (6)$$

Calculate the perturbation radius r_{adv} according to the perturbation range ε and the gradient g of the embedding layer:

$$r_{adv} = \varepsilon \cdot \frac{g}{|g|_2}. \quad (7)$$

Compared with the FGM confrontation method, which directly increases the disturbance to the boundary of the disturbance radius, the PGD confrontation method realizes the confrontation process step by step and gradually adjusts the disturbance. When the disturbance range exceeds the disturbance space, it reduces the disturbance momentum and looks for the best disturbance range.

5. Legal Decision-Making AI-Assisted Decision-Making Applications

Now, in relatively developed cities, intelligent trial systems have been specially developed to assist in the trial of judicial cases. For example, the “trial-centered litigation service software” system developed by a city’s high court, the system uses intelligent scoring and induction to conduct case trials, shortening the trial time. This system converts litigation materials into digital materials by scanning and systematically entering them, which is convenient for digital extraction in judicial practice. It provides basic information such as intelligent comparison and verification of evidence for judicial adjudication by automatically pushing cases of the same nature, which provides convenience for judicial case handling. There is also the “intelligent trial-assisted sentencing adjudication system” of a provincial intermediate court. Through further analysis of previous cases, the court’s rate of re sentencing and retrial initiation has shown a downward trend, as shown in Figure 4. It is enough to illustrate the positive significance of the application of artificial intelligence technology to judicial construction.

The specific parameters and settings of the whole artificial intelligence design decision RESNET algorithm include the input layer, residual module, batch normalization layer, pooling layer, and activation function. The input layer takes the preprocessed construction machinery modeling image metadata as the data set, including the preprocessed image, semantic label, and scoring data. The final data of the convolution layer is converted into a 13-layer fully connected network output. After the operation of the first layer of the convolution layer, it is divided into three main residual modules. The main diameter is extracted through three convolution layers. The purpose of the two-path fusion is to fuse the deep and shallow features of the design image data and semantic data. In this way, we can get a better feature classification effect. As shown in Figure 5, the design decision accuracy curve can be used to judge the quality of the model. It can be seen that with the increase in the number of iterations, the design decision accuracy of the RESNET artificial intelligence decision model gradually increases. When the number of iterations is about 160, the decision accuracy tends to be stable.

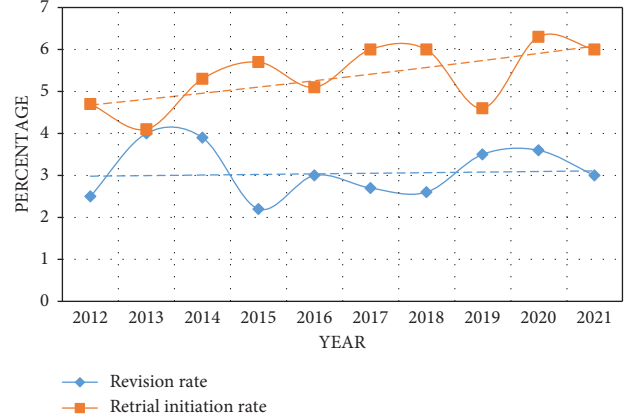


FIGURE 4: The retrial initiation rate and the rate of revision of the judgment in the provincial intermediate court.

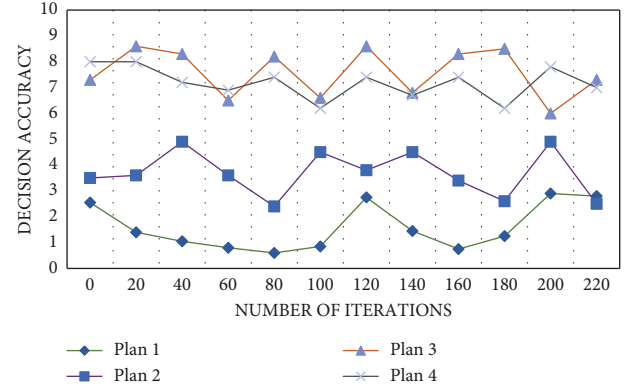


FIGURE 5: Accuracy comparison.

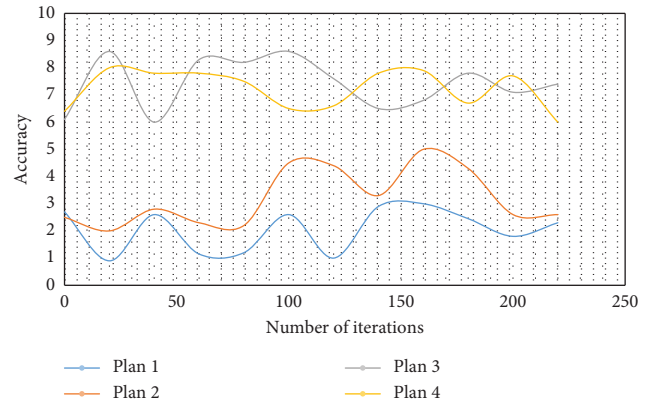


FIGURE 6: Comparison of decision satisfaction.

Figure 6 shows that the three modeling image words of science and technology, sensibility, and integrity are used as the semantic vocabulary labels of product design, and the three design scheme renderings are used as the decision accuracy curve comparison chart after image input. After the corresponding data training process, the design decision accuracy of the RESNET artificial intelligence decision model is further improved and shows a stable trend.

The input of the deep neural network generally uses the embedding method to convert the sparse one-hot encoding into a dense vector representation. At the same time, studies have shown that the combined use of multiple different types of embeddings can effectively improve the ability of the model. In our model, three embedding connections (concatenates) are used: a 256-dimensional word embedding followed by training, a pre trained 100-dimensional word vector, and a pre trained 200-dimensional word vector. The pre training uses the C language version of Word2Vec released by Google and performs unsupervised training on the case description and the original text of the facts of the data set provided by the sponsor. The training results are shown in Figure 7.

6. Experimental Results

Referring to the existing research of text classification technology, the tf-idf + svm machine learning model and the mainstream deep learning text classification model based on word embedding text CNN model, Han model, and DPCNN model are used as comparison models. After word segmentation of the case text, the fixed sentence length is 300 words, word2vec and glove word embedding algorithms are used, and the fixed word embedding dimension is 200. Word vectors are trained on all data sets respectively. It is compared with the open source word vector with a dimension of 200 obtained by directional skip Gram (DSG) algorithm in a large corpus. According to the performance of each model in different word vectors, word2vec word vectors are finally used in the text CNN model and Han model, and open source word vectors are used in the DPCNN model. For the two tasks of crime prediction and legal recommendation, each comparison model is modeled independently and jointly, and the best results of the model are recorded.

First, the comparison models such as TF-IDF + SVM and TextCNN are compared with the BERT12multi which uses the BERT pre training model as the encoder. At the same time, ablation experiments are performed on the effects of joint training and adaptive training, as shown in Table 1.

Experiments are carried out on the effect of grouping focus loss and the effect of gradient adversarial training. Considering the computational resource consumption of establishing independent models for the two tasks, the comparison models in the experiment all use the same joint training method as BERT12multi for multi-task learning. Because BERT12multi using adaptive training has better results in the experiment, BERT12multi in the following experiments adopts adaptive training. The experimental results are shown in Table 2 and Figure 8:

As shown in Table 2 and Figure 8, each model using grouped focus loss has improved on the Macro-F1 index, indicating that grouped focus loss can effectively improve the performance of each high model in the unbalanced data set of category samples and reduce the judgment error for small sample categories without significantly increasing the amount of super parameters. The bert12multi model perturbed by FGM and PGD improves the F1 score of the two tasks by about 0.001 and 0.003 respectively, indicating that

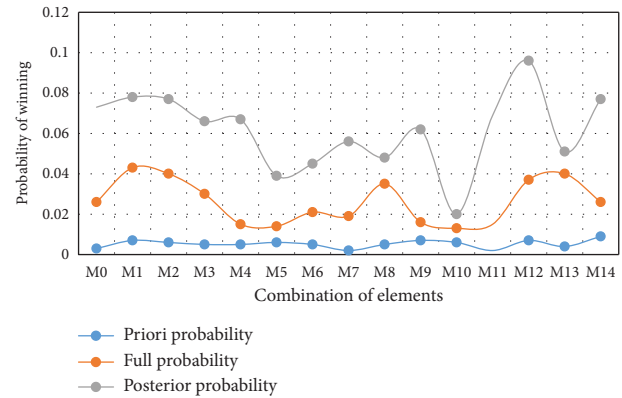


FIGURE 7: Prediction results of winning probability of experimental data points and similar data points.

the gradient confrontation training strategy can improve the generalization of the text classification model, and then improve the accuracy of decision prediction.

With the rapid development of artificial intelligence technology, the main development trends for assistant decision-making systems in the future are as follows:

- (1) Assistant decision-making under man-machine fusion. Due to the physiological limitations of human perception and decision-making ability, human beings cannot control future wars. The traditional method of controlling the whole battlefield factors by human beings has been unable to cope with the rapidly changing high-intensity future battlefield environment. Human soldiers will have to gradually jump out of the battle chain and cooperate with machine intelligence as planners, administrators, and commanders in most cases to control the process of the whole war, so as to promote the war from the war between humans to a new form of man-machine collaborative warfare, including precise perception, natural interaction in the environment, man-machine collaborative perception, and man-machine fusion computing.
- (2) Evolution of human-machine intelligence for decision-making. Machine intelligence may overturn the traditional human war decision logic. The new game confrontation strategy based on deep reinforcement learning, big data, and supercomputing shows a new data-driven optimal game strategy generation method and shows a strong game confrontation ability and self-learning evolution ability. In the future, the learning of evolutionary war decision logic driven by the mixture of knowledge and data will comprehensively replace the traditional military game decision logic based solely on operations research. The evolving new artificial intelligence technology has become a disruptive force in future military game decision-making. Mechanized war thinking will be replaced by new intelligent war thinking. The machine intelligence of non human traditional war methods has greatly extended human

TABLE 1: Comparison of evaluation indicators of each model

Mission name Model name	Crime prediction task			Statutory recommendations		
	Micro-F1	Macro-F1	F1-score	Micro-F1	Macro-F1	F1-score
SVM for task1	0.815	0.648	0.792	—	0.592	0.683
SVM for task2	—	—	—	0.773	—	—
Text CNN for task1	0.841	0.751	0.813	0.848	0.721	0.785
Text CNN for task2	0.875	0.753	0.814	—	0.721	0.753
Text CNN multi	—	—	—	0.85	—	—
HAN for task1	0.878	0.733	0.822	0.841	0.737	0.783
HAN for task2	0.987	0.764	0.819	—	0.739	0.791
HAN multi	—	—	—	0.842	—	—

TABLE 2: Comparison of evaluation indicators of each model 2.

Mission name Model name	Loss function	Crime prediction task			Statutory recommendations	
		Micro-F1	Macro-F1	F1-Score	Micro-F1	Macro-F1
Text CNN	Cross entropy	0.875	0.753	0.814	0.85	0.721
HAN	Cross entropy	0.879	0.764	0.822	0.842	0.739
DPCNN	Focal loss	0.882	0.766	0.824	0.858	0.739
BERT12multi	Focal loss	0.884	0.782	0.822	0.847	0.758

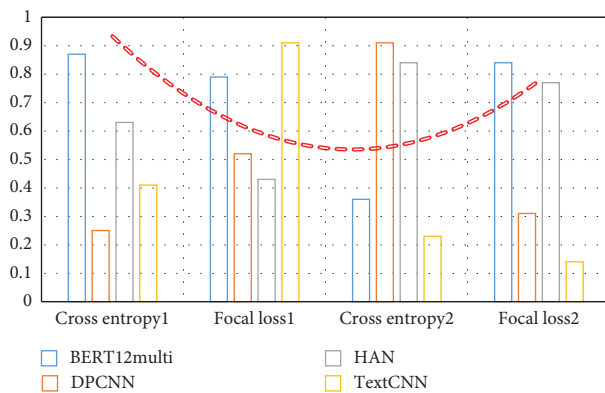


FIGURE 8: F1-score comparison of each model under different loss functions.

intelligence, which is the latest cognition of human beings to explore the laws of war and enhance intelligence by human-computer mixing.

7. Conclusions

With the progress of the times, the series connection of information and intelligent network pulse is the requirement of the times of legal development. The application of artificial intelligence in legal judgment is a progressive exploration of the combination of computer science and law. The application of artificial intelligence technology further improves the efficiency of legal case handling, promotes the standardization of the legal process, and strengthens the construction of legal data. The application of artificial intelligence in legal judgment also provides a reference for solving problems and improves the ability of dispute resolution. The practice has proved that the combination of the legal field and artificial intelligence technology is the choice

of the times, and the better the development, the more prosperous the legal civilization can be created. However, from the perspective of the long-term development of the Internet and big data, the exploration of artificial intelligence in legal judgment still needs to overcome many difficulties, for example, strengthening the research of artificial intelligence in cognitive reasoning and solving technical logic obstacles. By incorporating artificial intelligence decision-making methods, this paper can greatly improve the speed of case reasoning, with the highest compressible model volume reaching about 11% of the original, and the reasoning speed being increased by about 8 times. At the same time, performance close to that of the deep neural model is obtained, which is superior to other prediction models of legal judgment based on word embedding. It is far from enough to have such research results. It takes a long time to explore and accumulate practical experience in order to establish a complete artificial intelligent legal judgment theory.

Data Availability

All data used in this study are presented in the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] H. Karimi and J. Tang, "Decision boundary of deep neural networks: challenges and opportunities," *WSDM '20: The Thirteenth ACM International Conference on Web Search and Data Mining*, vol. 29, no. 11, p. 389, 2020.
- [2] Y. Guan, J. Zhao, and Y. Qiu, "Melodic phrase segmentation by deep neural networks," *Social Science*, vol. 20, no. 1, p. 20, 2018.

- [3] M. Pirhooshyaran and L. V. Snyder, "Simultaneous decision making for stochastic multi-echelon inventory optimization with deep neural networks as decision makers," 2020, <https://arxiv.org/abs/2006.05608>.
- [4] L. Chen, J. W. Guo, and W. U. Jian-Gong, "Research on artificial intelligence assisted decision-making algorithms for lawyers based on legal-computing theory," *Journal of East China Normal University*, vol. 28, no. 23, p. 928, 2019.
- [5] H. Hong, X. Lin, and K. Tang, "Artificial-intelligence assisted decision making: a statistical framework," *Social Science Electronic Publishing*, vol. 39, no. 2, p. 488, 2020.
- [6] M. Schemmer, P. Hemmer, and M. Nitsche, "A meta-analysis on the utility of explainable artificial intelligence in human-AI decision-making," *Journal of Artificial Intelligence*, vol. 30, no. 4, p. 29, 2022.
- [7] C. Rastogi, Y. Zhang, and D. Wei, "Deciding fast and slow: the role of cognitive biases in AI-assisted decision-making," *Computer deep learning research*, vol. 29, no. 23, p. 98, 2020.
- [8] J. Gerards and F. Z. Borgesius, "Protected grounds and the system of non-discrimination law in the context of algorithmic decision-making and artificial intelligence," *Social Science Electronic Publishing*, vol. 249, no. 23, p. 48, 2019.
- [9] X. Wu and L. School, "The possibility and limits of modeling judicial decision-making in artificial intelligence," *Zhejiang Social Sciences*, vol. 3, no. 12, p. 48, 2017.
- [10] D. J. Qian, S. U. Hang, and S. O. Law, "The change of China's legislative decision-making mode under the perspective of artificial intelligence," *Journal of Hunan University of Science & Technology (Social Science Edition)*, vol. 2e, no. 2, pp. 18–28, 2018.
- [11] L. B. Solum, "Legal personhood for artificial intelligences," *Legal Information Retrieval and Applications*, vol. 38, no. 4, p. 98, 2019.
- [12] V. R. Walker, N. Carie, and C. C. Dewitt, "A framework for the extraction and modeling of fact-finding reasoning from legal decisions," *Artificial Intelligence and Law*, vol. 19, no. 2, p. 49, 2021.
- [13] K. Brennan-Marquez and S. Henderson, "Artificial intelligence and role-reversible judgment," *Journal of Criminal Law and Criminology*, vol. 20, no. 2, p. 423, 2019.
- [14] D. H. Berman and C. D. Hafner, "The potential of artificial intelligence to help solve the crisis in our legal system," *Communications of the ACM*, vol. 32, no. 8, pp. 928–938, 1989.
- [15] H. P. Olsen, J. L. Slosser, and T. T. Hildebrandt, "What's in the box? The legal requirement of explainability in computationally aided decision-making in public administration," *Social Science Electronic Publishing*, vol. 309, no. 33, pp. 128–210, 2019.
- [16] K. Hawkins, "On legal decision-making," *Washington and Lee Law Review*, vol. 23, no. 39, p. 477, 2017.
- [17] D. Merkl and E. Schweighofer, "The exploration of legal text corpora with hierarchical neural networks," *Guide tour in public international law*, vol. 34e, no. 23, p. 384, 2017.
- [18] A. Preece, R. Ashelford, and H. Armstrong, "Hows and whys of artificial intelligence for public sector decisions: explanation and evaluation," *Public affairs management*, vol. 43, no. 34, p. 280, 2018.
- [19] N. E. Sondak and V. K. Sondak, "Neural networks and artificial intelligence," *ACM*, vol. 2, no. 7, pp. 241–245, 2017.
- [20] J. Zeleznikow and A. Stranieri, "The split-up system: integrating neural networks and rule-based reasoning in the legal domain," *The fifth international conference of Springer-Verlag*, vol. 32, no. 89, p. 990, 2020.