Research Article

The Application of Artificial Intelligence Decision-Making Algorithm in Crisis Analysis and Optimization of the International Court System

Yuan Zhang,[1] Yuepeng Zhao,[2] and Yueqin Zhao (iD)[1]

[1]*Hebei Finance University, Baoding 071051, China*
[2]*Hebei College of Science and Technology, Baoding 071000, China*

Correspondence should be addressed to Yueqin Zhao; zhaoyueqin@hbfu.edu.cn

Regardless of theoretical research or practical experience, the construction of international wisdom courts has produced a large number of wisdom achievements; however, the scientific convergence between the institutionalized logic of sustainable development and the existing practical achievements has not yet been realized, and there are still problems in the construction of smart courts, such as insufficient practical application, poor system coherence, and insufficient online litigation. Therefore, this study proposes a fusion CNN-GRU network result model that uses a text classification neural network structure convolutional neural network with a good effect to combine the GRU neural network model structure to carry out the task of recommending the law. Promote the development of the connotation of smart courts through the collaborative integration of artificial intelligence technology and system to promote the deep integration of judicial theory and intelligent technology, the deep integration of litigation needs and intelligent technology, and the deep integration of general technology and proprietary development to provide technical mechanisms and legal institutions for the construction of smart courts.

1. Introduction

As the basic value of judicial practice, fairness and efficiency have always attracted much attention, and the creation and improvement of the new format of court modernization and smart courts precisely reflects the technical way of realizing the value of intelligent justice. As the basic plan for the modernization of courts in recent years, the construction of smart courts has received strong attention from academic circles and judicial practice departments [1]. Regardless of theoretical research or practical experience, although the current construction of smart courts has provided us with a large number of wisdom achievements, in terms of the institutionalized logic of sustainable development, the existing practical results have not yet achieved scientific convergence, and the practice of smart court construction is in some fields, such as technology exploration, knowledge structure, algorithm development, and other fields. There are fault phenomena, and it is necessary to further explore and

improve the theory of system engineering construction on the basics and practical solutions [2]. Therefore, starting from the practice of smart court construction, systematically summarize the basic issues of smart court construction at this stage, and on this basis, deeply analyze the origin of the problem and its development trend. Problem-oriented, explore the direction of the construction of smart courts and gather theoretical consensus and practical experience for the sustainability of international smart court construction [3]. In terms of development logic, the construction of smart courts belongs to the basic category of court modernization, and it is the basic result of court informatization construction under the pattern of court modernization, especially "artificial intelligence + justice." However, the question is also objective, which is related to the initial stage of the development of smart courts [4].

With the advent of the era of intelligence, although intelligent construction has touched all aspects of court informatization construction, the problem of insufficient

application in the core business of adjudication is also significant [5]. The reasons for the lack of practice are manifold, with the subjective understanding of judicial personnel, the maturity of technology, and the narrow scope of applicability as the main reasons. Because of the above factors, the application of the intelligent system at this stage has not yet achieved the expected effect, and it needs to be further developed and improved in terms of functional perfection, universality, and coverage [6]. Taking the electronic file as an example, the intelligent auxiliary function provided by the in-depth application of the electronic file cannot yet cover the whole process of trial enforcement and the full coverage of the application of courts across the country. It is not intelligent and humanized enough, i.e., the push of similar cases is not accurate enough, and the automatic backfill of case information is not perfect enough [7]. In addition, in litigation practice, the application of the smart court system should be based on the judge as the core, with the judge's business facilitation, specialization, and efficiency as the mainstay. The technical maturity not only needs to refer to the judge's mechanical duplication of labor reduction value but also has a direct relationship with the judge's personal understanding, cognition, and recognition of the system. The main promoters of the construction of smart courts are the court leadership, and the applicators are the judges of the trial-line [8, 9]. Comprehensively promoting the construction of smart courts should take high-intelligence judicial services as the direction and provide application suggestions for the improvement of convenient, efficient, and professional intelligent services through the seamless connection between technological innovation and the practical application of trial-line judges to realize the integrated development of smart court construction and the application of smart justice, e.g., litigation services, smart trials, and smart management [10, 11].

Insufficient system coherence is one of the main problems facing the construction of smart courts worldwide at this stage. As a systematic and technical complex project, there are many problems in the system collection of the wisdom court system at this stage. The compatibility between the subsystems needs to be improved, and the subsystems lack efficient connection, data sharing, coordination, and interaction between the operating procedures. The lag of the data sharing system based on judicial big data has become a basic problem restricting the efficient operation of the overall system of smart courts [4]. Only by accurately explaining the inherent functional mechanism of judicial big data through massive fragmented data aggregation and structured processing, systematically revealing the correlation between various modules, and realizing statistically significant value judgments and predictions, in addition to conducting diachronic and changeable research, can the purpose of case prediction and development trend analysis be achieved [12]. Driving the construction of an interconnected and seamless intelligent system requires all-round, all-encompassing, and all-field consensus among various judicial practice departments to promote the application of business data and the ability to share resources, which rely on the coordination and unification of intelligent

technology and data resources, which is the solution of this basic problem. In fact, it affects the repair problem of insufficient integration between the subordinate subsystems of the entire intelligent court system [13, 14]. However, the scope and objects involved in this restoration process are extensive, requiring the coordination of the Supreme People's Court and the grassroots courts, as well as advanced technical support from technology companies. The problems and solutions are often not completely consistent in form, and it takes time to correspond to them one by one. It is inseparable from the precise design of technical public relations personnel. An important foothold of the smart court is intelligent technology, and opening up the integration between intelligent technology and judiciary is also an inevitable requirement for the integration of the system [15]. To this end, it is necessary to smooth the supply channels of "artificial intelligence + justice" composite talents, which will inevitably involve the creation and development of new interdisciplinary disciplines. As a result, the problems involved in solving the lack of coherence of the smart court system are also extensive [16].

The basic embodiment of smart courts is online litigation. However, because of various reasons, the current dispute resolution mechanism of the International Court of Justice is still dominated by offline litigation, and the scope and application of online litigation are still insufficient. Usually, judicial activities have a strong on-site expression significance, reflecting a kind of on-site judicial prosecution activity, with significant physical space, subject participation, and activity openness. In contrast, online litigation in the context of smart courts has obvious virtual spaces, off-site participation, and the relative secrecy of activities. This significant difference between the two makes it difficult for citizens to adapt based on the inertia of traditional litigation thinking [17]. Under normal circumstances, the transfer of traditional offline case disputes to online will inevitably bring many discomforts to the parties, among which the degree of technical achievement, emotional acceptance, and acceptability of litigation results are the main bottlenecks restricting the development of online litigation. Although the number of e-commerce disputes continues to grow with the diversified development of new formats of the Internet economy, the current dispute resolution mechanism in China is still mainly offline [18]. It is true that the difference between Internet dispute resolution and traditional judicial disputes in the era of intelligence will promote the sustainable development of online litigation forms, and because of the limited online litigation method presented by the Internet court at this stage, ordinary e-commerce disputes are still fond of offline litigation. In addition to e-commerce litigation disputes that are suitable for resolution through the online litigation model, the practical conditions for other ordinary judicial disputes to be transferred from offline to online are not yet available. It is directly related to the online litigation cognition and acceptability of litigation subjects. In practice, the problems faced by the modernization of courts are diverse, for example, in the adjudication system, there are problems, such as the lack of clarity in the positioning of adjudication functions, and the standards and procedures

for applying for civil administrative retrials need to be optimized [19].

Smart courts pay more attention to the judicial application of artificial intelligence technology and the facilitation and intelligent services it brings to the trial business. At present, the construction of smart courts mainly revolves around the popularization and application of judicial big data and artificial intelligence technology in the process of court informatization, and the real sense of smart courts has not yet reached a sufficient degree of understanding in the scientific and technological dimensions. In this regard, the integration between the degree of technological achievement and the degree of integration of judicial spirit is obviously more complex than the simple application of artificial intelligence. From the perspective of conceptual analysis, the definition of smart courts by both academic circles and judicial practice departments gives more expectations for judicial fairness, convenient litigation, and technical neutrality. At this stage, the focus of the construction of smart courts is mainly on improving litigation efficiency, promoting the same case and the same sentence, criminal assistance, and smart enforcement. The theoretical discussion in the academic circles is also mostly focused on it. However, as an intelligent judicial system, in addition to the above functions, the smart court is also the main indicator for examining its scientific and sustainable nature, especially the realization of judicial justice. The link between smart technology and justice often lies not in court referees but in technology developers. Technological neutrality under the position of technological instrumentalism does not give more convenience to referees, nor does it restrict the exercise of their discretion. For referees, the neutrality of technology is mainly reflected in their adherence to legal provisions, which is contrary to judicial justice in the substantive sense. The riskiness of technology can usually be avoided through institutionalized design. The double-edged effect of smart court construction should be implemented under the framework of institutional rationality, and ultimately, one should realize its own auxiliary value through institutional control. Based on it, the construction of smart courts in the intelligent era urgently needs to take the creation of an ecological system that conforms to the logic of judicial justice as the basic development direction. In terms of practical rationality, it is imperative to explore the solution path of various problems in the initial stage of the construction of smart courts.

1.1. State of the Art.

The research of legal intelligence has a long history. With the progress of science and technology and the rapid development of Internet technology, the informatization in the judicial field is also basically completed; due to the use of judicial informatization has also produced a significant amount of data, big data are changing and will continue to profoundly change our social governance operation mode and organizational structure. On the one hand, in this era of large-scale production, sharing, and application of big data, it provides a huge amount of information for procuratorial organs to handle cases. On the other hand, it has also formed a trend of forcing procuratorial organs to introduce big data. In the field of public prosecution business, the main function of big data application and prediction can be well-combined with the task of sentencing recommendation to form a new business model: big data intelligent sentencing, which is also based on deep learning technology to process big data legal intelligence.

The task of crime prediction is to determine appropriate crimes in the course of the case, such as robbery, theft, and violence, and to predict the crime by analyzing the text and facts of the case description by the machine. The recommendation of legal provisions is to predict the relevant legal provisions and laws involved in the case according to the description of the case and the facts in the legal documents in the criminal case. Crime prediction and law recommendation are tasks in the sentencing process, and they are important subtasks related to the law, in the case of sentencing crimes and sentencing law provisions that play an important role in the judicial process, crime prediction, and law: article recommendation can be used as a sentencing auxiliary system, and it can benefit many people involved in the task procedure. For example, it can provide a convenient reference for legal experts (such as lawyers and judges) to improve their work efficiency. At the same time, it can also provide help and legal advice for ordinary people who are not familiar with legal knowledge. According to the above crime prediction and the description of the recommended tasks of the law, it is actually classified according to the text data of the case description and the facts. Now, there has been a lot of scientific research on the text classification task in the field of natural language. Text data classification problem is a major task in the field of natural language processing problems. The relevant scientific method research can be traced back to the 1950s when the scientific research method is to use expert rules to solve the classification problem of text data, and in the 1980s, the problem of text data classification has advanced to the use of knowledge engineering technology to form an expert system; however, the scope of classification and classification accuracy used is very limited. With the development of statistical learning-related method theories, especially the growth of text data on the Internet and the rise of machine learning disciplines after the 1990s, a class of classical algorithms to solve a large number of text classification tasks has been slowly established, and the main method technology in this period is artificial feature engineering combined with shallow classification models. The text classification task is divided into two components: feature engineering and classifier. The main methods are the naive Bayes classification algorithm, KNN, and SVM. The above method is a traditional solution to the problem of text data classification methods, the text representation of these methods, the main problem is high latitude and high sparseness, and the ability to express characteristics is very limited.

Now, the mainstream technology and methods are mainly used, and deep learning technology implements the text classification method. Bengio et al. proposed the neural probabilistic language model (NPLM) that uses a text

distributed representation. Each word representation is a dense real vector. The biggest advantage of text distributed representation is that it has a very strong ability to express features. The disadvantage is that the context semantic expression is lacking. Mikolov et al. proposed a distributed representation of the use of words and released the Word2Vec toolkit. The advantage is to achieve good results in rich semantic expression and promote the development of text analysis, through the word vector to represent text data, the text representation from high latitude and high sparse, and difficult to process the way, similar to the image and audio of the continuous dense characteristics of the data. The disadvantage is Word2Vec word vector text representation and text.

There is still a fixed gap in the original semantics, and more is to learn words with similar contexts. Word vector representation solves the problem of text data representation. It provides a method for deep learning networks to extract features using word vector representation, which can solve the text classification problem. Joulin et al. proposed that all word vectors in the sentence are averaged and then entered into the Softmax function to classify the text. This model is simple, and one can obtain a lot of classification information without too many nonlinear transformations and feature combinations. The disadvantage is that the structure does not fully consider the coherence of the word sequence information in the sentence. Kim proposed the use of convolutional neural networks for sentence classification problems in 2014. The advantage of the method is that it is possible to extract the correlation of local features of text data, and the use of CNN models can be used to extract key information features similar to n-grams in sentences. The disadvantage is that CNN has the size of a fixed filter, which cannot be molded for longer sequence information, and the hyperparameter adjustment of the filter is also very cumbersome. In natural language processing problems, to better express contextual information, the recurrent neural network (RNN) is more commonly used. It uses the RNN multitask learning framework in the literature to learn multiple related tasks together, thus providing a text classification design method for RNN learning in multitask; however, it has the disadvantage of being impact-scoring. The key information in the class process is not intuitive enough to portray it. In 2016, Yang et al. used the principle of attention mechanism to increase the weight of features that have a greater impact on the classification results. Li Mengtong proposed an intelligent flood evacuation model based on the deep learning of various flood scenarios to make flood control intelligently [20].

2. Methodology

2.1. Deep Learning. Deep learning is an important direction of scientific research in recent years, and it has achieved very good performance in many fields of artificial intelligence. Metrology is still the most important method support for evaluation science; however, the emergence of deep learning technology can well complement the data methods and measurement indicators in the existing evaluation system of

various types. Hence, it has become a research topic that evaluates science needs to focus on at present and for a period of time in the future. At its roots, deep learning is a branch of machine learning that refers to a class of problems and ways to solve them. First of all, the deep learning problem is a machine learning problem, which refers to the general law summarized by the algorithm from a limited sample. It can be applied to the new unknown data, for example, we can summarize the law between the case description and the sentencing from some historical court judgments, so that when there is a new case description, we can use the summarized law to judge what crime the case description has committed. Secondly, unlike traditional machine learning, deep learning uses models that are generally complex, meaning that the data flow between the original input and output targets of a sample passes through multiple linear or nonlinear components. Because each component processes the information, it affects subsequent components. It is not clear how much each component contributed in the period. This problem is called the credit assignment problem (CAP). In deep learning, the contribution allocation problem has always been a very critical issue, which is related to how to learn the parameters in each component. At present, a better model to solve the contribution allocation problem is the artificial neural network (ANN). Artificial neural networks are also called neural networks, and neural networks are mathematical models of construction that are inspired by the way neural networks in the human brain work. The structure of artificial neural networks is the composition of the connections between neurons, among which there are two special types of neurons.

One is to be able to receive data information from outside, and the second is to output data information to the outside. Artificial neural networks can be thought of as data processing networks that change input to output. A neural network unit can be seen as a complex function controlled by a set of parameters, and the parameters of a neural network can be learned from the data in a machine learning manner because neural network models are generally complex, from input to output.

The information transmission path is relatively long. Hence, the learning of complex neural networks can be regarded as a kind of deep machine learning, i.e., deep learning.

Deep learning (DL) can be seen as a problem of how to learn a "deep model" from data, and deep learning is a subproblem in machine learning. By creating a model structure with a certain "depth," it is advisable to let the model automatically learn feature representation, and the process of learning features is from the underlying features to the middle features to the high-level features. Hence, the accuracy of model prediction or recognition can be improved. Figure 1 shows the data processing process for deep learning.

Representation learning: to improve the accuracy of machine learning systems, we need to convert the input information into a valid feature, also known as representation. If there is an algorithm that automatically learns valid features and improves the performance of the final classifier,
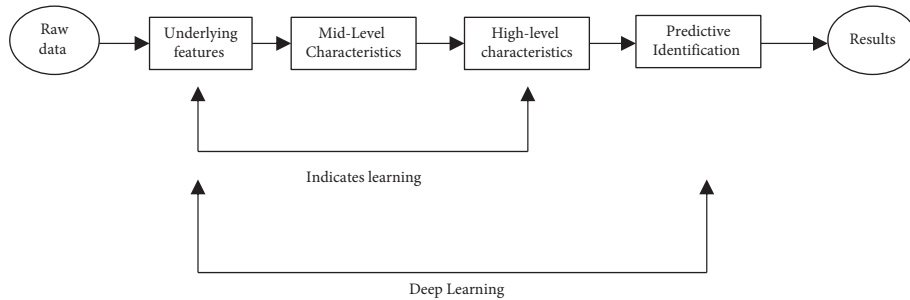
FIGURE 1: Data processing process of deep learning.

then this learning is called representation learning. In the field of deep learning, representation refers to the input observation sample X of the model in what form and in what way through the parameters of the model. Representation learning refers to learning a representation that is valid for observation sample X. There are many forms of representation learning, for example, the supervised training of CNN parameters is a supervised form of representation learning, and the unsupervised pretraining of autoencoders and restricted Boltzmann machine parameters is an unsupervised form of representation learning. With the rapid development of deep learning, the depth of the model continues to increase, and the ability to represent its features is also getting stronger, making subsequent predictions easier.

The main purpose of deep learning is to automatically learn valid features from data, i.e., representation learning. The technique of deep learning can be seen to some extent as a representation learning technique that turns raw data into a higher-level, more abstract representation through multilayered nonlinear transformation. These learned representations can replace artificially designed features. With the development of neuroscience and cognitive science, people have gradually realized that human intelligent behavior and brain activity are related. Inspired by the neural network system of the human brain, early scientists constructed a kind that mimics the human brain.

The mathematical model of the nervous system, called the artificial neural network, is referred to as the neural network. In the field of machine learning, the neural network model refers to a network structure model composed of many artificial neuron structures, and the strength of the previous connections of these artificial neurons is a parameter that can be learned.

Neuron models in computers: in 1943, psychologist McCulloch and mathematician Pitts drew on the network structure of biological gods and elements to propose an abstract neuronal model MP. The neuronal model consists of three parts, namely, the input, output, and computational function modules. The input part is equivalent to the dendrite of the neuron, the output is equivalent to the axon of the neuron, and the calculation module can be understood as the nucleus. An example of a neuronal network model in a computer is shown in Figure 2.

2.2. Neural Network Model Optimization Method. Neural network model optimization is to optimize the parameters in

the neural network model. In machine learning, the simplest and most commonly used optimization algorithms are the gradient descent method, batch gradient descent method (BGD), stochastic gradient descent method (SGD), and small batch gradient descent method (MBGD). This paper uses the stochastic gradient descent method to optimize the minimization loss function. The loss function is a nonnegative real function that is used to quantify the difference between the prediction of the deep learning model and the real classification label; for classification problems, now generally use the cross-entropy loss function. This study also uses this function to calculate the previous difference between the prediction and the real label. The cross-entropy loss function can be understood specifically to assume that the label $y \in \{1...C\}$ of the sample is a discrete category label, and the output of model $f(x, \theta) \in [0,1]^C$ is the conditional probability distribution of the class label, i.e., as shown in the following equation.

$$P\left(y = \frac{c}{x}\right) = f_c(x, \theta). \tag{1}$$

Satisfy the following equation.

$$f_c(x, \theta) \in [0, 1], \sum_{c=1}^{C} f_c(x, \theta) = 1. \tag{2}$$

A c-dimensional one-hot vector y can be used to represent the label of the sample. Assuming that the label of the sample is k and then the label vector y, only the value of the k dimension is 1, and the rest of the elements are 0.

The sample label y can be seen as the true probability distribution of the sample label, i.e., the cth dimension (denoted $_{yc}$, $1 \le c \le C$) is the true probability of the category c. Assuming that the class of the sample is k, then the probability that it belongs to the kth class is 1, and the probability of the other classes is 0.

For two probability distributions, the difference between them can be measured by cross-entropy. The true distribution of labels is represented as y. The cross-entropy before the predicted distribution of the modulus yc(x, 0) is shown in the following equation.

$$L(y, f(x, \theta)) = -\sum_{c=1}^{C} y_c \log f_c(x, \theta). \tag{3}$$

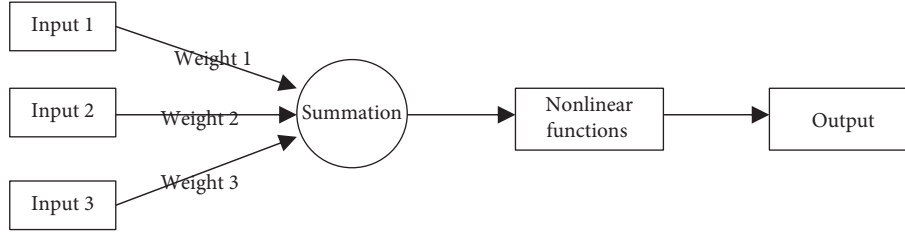In machine learning, the simplest and most commonly used optimization method is gradient descent, i.e., the

FIGURE 2: Neuron network model in a computer.

minimum value of the risk function on the training set D is calculated by an iterative method as shown in the following equation.

$$\theta_{t+1} = \theta_t - \alpha \frac{\partial R_D(\theta)}{\partial \theta}$$

$$= \theta_t - \alpha \frac{1}{N} \sum_{n=1}^{N} \frac{\partial L(y^{(n)}, f(x^{(n)}, \theta))}{\partial \theta}, \quad (4)$$

where 0 is the parameter value at the tth iteration and a is the search step. In machine learning, ∞ is generally called the learning rate, and in the gradient descent of formulae (1)–(4), the objective function is the risk function on the entire training set, which is called the batch gradient descent (BGD) method. The batch gradient descent method needs to calculate the gradient of the loss function on each sample and sum iterate. When the sample size N in the training set is very large, the spatial complexity is relatively high, and the computational overhead of each iteration is also very large; in machine learning, assuming that each sample is independent and distributed randomly, extracted from the real data distribution, the real optimization goal is to expect the least risk.

Batch gradient descent is equivalent to taking N samples from a real data distribution and calculating the gradient of empirical risk from them to approximate the gradient of desired risk. To reduce the computational complexity of each iteration, we can also take only one sample at each iteration, calculate the gradient of this sample loss function, and update the parameters, namely, the stochastic gradient descent method. When a sufficient number of iterations have passed, the stochastic gradient descent method, also known as incremental gradient descent, can converge to the local optimal solution.

2.3. Convolutional Neural Networks. Convolutional neural network (CNN) is a multistage globally trainable artificial neural network model and a deep neural network structure containing convolutional computation in the convolutional kernel. A convolutional neural network can pass fewer preprocessing processes and even learn from the original data to obtain an abstract. Essential and high-order feature representation, one of the classic algorithms in the field of deep learning, is a neural network model designed to process data with mesh-like structures. For example, on time series data, it can be considered that there is regular sampling on the timeline to form a one-dimensional grid in the image

data. Images can be seen as two-dimensional pixel meshes, and convolutional neural networks have performed extraordinary results in many fields. They are now widely used in computer vision, speech recognition, and natural language processing. Convolutional neural networks generally contain data input, convolutional layers, pooling layers, fully connected layers, and classifiers, and the specific extraction data characteristics and classification process of convolutional networks are shown in Figures 3 and 4.

(1) Input layer: assuming a two-dimensional tensor, the process of convolution using convolutional kernels is shown in Figure 4.

(2) Pooling layer: it is the operation after the convolutional layer in the convolutional neural network model structure. The pooling layer mainly uses the pooling function. The characteristic results of the convolutional layer output are pooled, and the principle of pooling function is to use the local features in the feature map instead of the overall features to represent the output of the position, for example, the max pooling function calculates the maximum value in the adjacent matrix. Like convolutional layers, the pooling layer has a fixed-shaped window (also known as pooling window) for the input data at a time in the computed output. Unlike the convolutional layer that calculates the intercorrelation between input and kernel, the pooling layer directly calculates the maximum or average value of the elements within the pooling window. This operation is also called maximum pooling or average pooling, respectively.

The average pooling function is to calculate the average value in the matrix, and the role of the pooling layer is to continuously reduce the spatial size of the data. The number of parameters and calculations in the neural network model will also decrease. Hence, the pooling layer also prevents the over-interaction of the model to a certain extent. The maximum pooling function specific pooling process is shown in Figure 5.

A schematic diagram of the specific process of the average pooling function is shown in Figure 6.

Convolutional neural networks also perform well in text classification tasks in the field of natural processing, and there are several parts of the convolutional neural network structure model in the process of text classification, namely data text preprocessing, data word embedding after processing, convolutional layer extraction of text features,
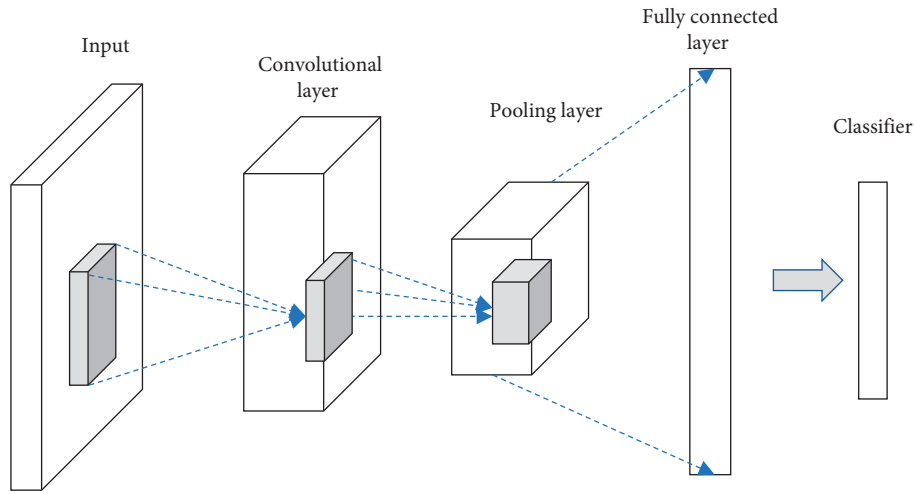
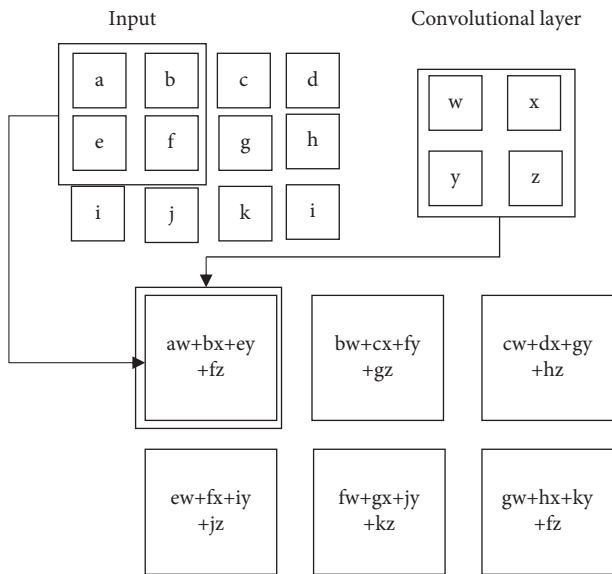FIGURE 3: Classification structure model of the convolutional neural network.

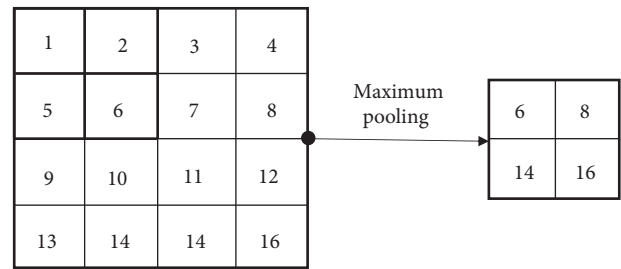FIGURE 4: Convolution of a convolutional kernel on a two-dimensional tensor.

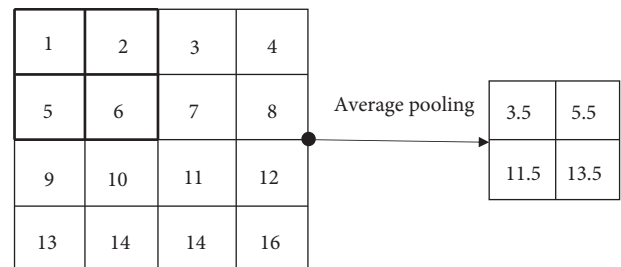FIGURE 5: Maximum pooling calculation chart.

FIGURE 6: Maximum pooling calculation chart.

pooling layer compression of data, and classification function classification. After comparative analysis of the structure of convolutional neural networks and general neural networks, there are three characteristics. The first is a local connection, and the second is weight sharing. The third is subsampling, which makes convolutional neural networks have better model generalization capabilities and powerful data feature extraction capabilities. The core components of convolutional neural networks mainly contain three parts, which are the convolutional layer, pooling layer, and fully connected layer. Usually, the convolutional layer and pooling layer are the main components of deep learning. The convolutional neural network can automatically select the feature representation of the data, and the convolutional neural network can obtain the classification target after extracting the feature representation and then going through the classification function.

3. Result Analysis and Discussion

3.1. Legal Recommendations Based on Converged CNN-GRU Networks. The recommendation of the law is to predict the relevant laws involved in this case based on the description of the case and the facts in the criminal legal documents. The law is the main basis for judicial sentencing. Hence, as the research object of this chapter, the main content of the research in this section is to propose the use of deep learning technology in the convolutional neural network and GRU neural network fusion method and to promote the legal intelligence law. Good results were achieved in the recommendation task, which resolved the relevant laws involved in the case based on the description of the case and the factual parts of the criminal legal documents. It provides a legal basis for rapid sentencing and provides legal support for personnel involved in the case, thus improving the

efficiency of judiciary and achieving rapid and intelligent sentencing assistance. The experimental data are based on the dataset of the China Law Research Cup competition, which is used as data for criminal cases provided by the Supreme People's Court of China.

3.2. Model of the Fusion CNN-GRU Network in Legal Recommendation.

The main work of this section is to propose the structure of the fusion CNN-GRU neural network model according to the inherent relationship between the case description of the facts and the provisions of the law, and the fusion CNN-GRU neural network is analyzed through the experimental results of other text classification models to achieve better results of the model. The proposed fusion CNN-GRU model is to recommend the overall accuracy of the criminal law based on convolutional neural networks. As a result of poor accuracy in multiple classification parts, the models of this structure usually encode the input sequence as a fixed-length vector representation, and for shorter-length input sequences, the mold can learn a reasonable vector representation. However, the problem with this model is that when the input sequence is not long, the model has difficulty learning a reasonable vector representation. Therefore, the introduction of the GRU neural network model to overcome the overall classification accuracy is high, and the problem of single classification accuracy bottom, through CNN and GRU to solve the problem of legal recommendation, is due to the recommended essence of the law and can be seen as a multi-classification problem, and the law recommendation is a method of different lengths of root input. The description of the facts of the case in the legal documents predicts the relevant laws involved in this case. In this way, it helps judicial personnel to carry out judicial assistance according to the results recommended by the law.

3.3. Recommended Model for Converged CNN-GRU Networks.

The structure of the converged CNN-GRU network model proposed in this chapter is based on the characteristics of the convolutional neural network (CNN) and the long-term short-term memory network (LSTM) network model. The new converged CNN-GRU model structure is proposed according to the actual law recommendation problem in legal intelligence to be solved by the text. As the convolutional neural network model in the field of text classification has a good performance in extracting text features to represent the case description and the factual part, there is a rich semantic relationship between the case description and the text context of the fact part, and the convolutional neural network has a rich semantic relationship between the case description text information features.

The ability to obtain feature representation of the relationship between the before and after of the case text information is weak, and the neural network model GRU network model has a good effect on learning the sequence information that has a dependency relationship over long distance. Hence, this paper proposes that the GRU network model extracts the semantic relationship between the case

description and the fact part, and the GRU network model and the convolutional neural network are fused into a new one network structure model CNN-GRU. In the experimental comparison process with the commonly used text classification model, the fusion CNN-GRU model structure has obtained better results in solving the problem of legal recommendation in legal intelligence. The structure of the converged CNN-GRU model based on CNN and GRU is shown in Figure 7. The fusion GNN-GRU modular structure is divided into data input layer, embedding layer, convolutional layer, pooling layer, GRU layer, and output layer.

The main process of the structure of the fusion CNN-GRU model based on CNN and GRU is divided into the following parts:

3.3.1. Experimental Process and Result Analysis.

For the use of convolutional neural network as a text classification method, the input data should first be processed, and the text data will be converted into vectors through embedding, followed by the convolutional layer to extract the text feature representation and then through the pooling layer to compress the feature representation to extract the main features through the convolutional neural network table to obtain the effective features of the text data. Then, sequence the model GRU layer. The long-dependent information of the case description text information is carried out, and finally, the classification function is calculated to obtain the target category. The purpose of text data classification is achieved, and then the recommended task of legal intelligence laws is achieved. The specific process of the convolutional neural network for case description and fact part data processing is as shown in Figure 8.

3.3.2. Convolutional Layer.

The main function of the convolutional layer is to extract the characteristics of the case description and the fact department, and the main structure of the convolutional layer mainly has the characteristics of value sharing, local connection, and composition of multiple convolutional kernels of different sizes. Local connections and weights share the primary.

To use the case description and the word embedding matrix for the factual part of the information, the role of multiple convolutional kernels of different sizes is main. The function to be used is to extract the characteristics of different granular information for case description information because the extraction of information with a different granularity is to make the characteristics of the case description and the fact part of the text information extracted more in line with the original meaning of the expression. For convolutional layers with multiple convolutional kernels, one of which can be expressed, at height h and width k, for h word convolution operations, the characteristic output is expressed as follows:

$$s_i = f(w * x_{ii+h-1} + b), \tag{5}$$

where w is the convolutional kernel weight matrix parameter, $x + 1$ represents the feature representation of the input, s
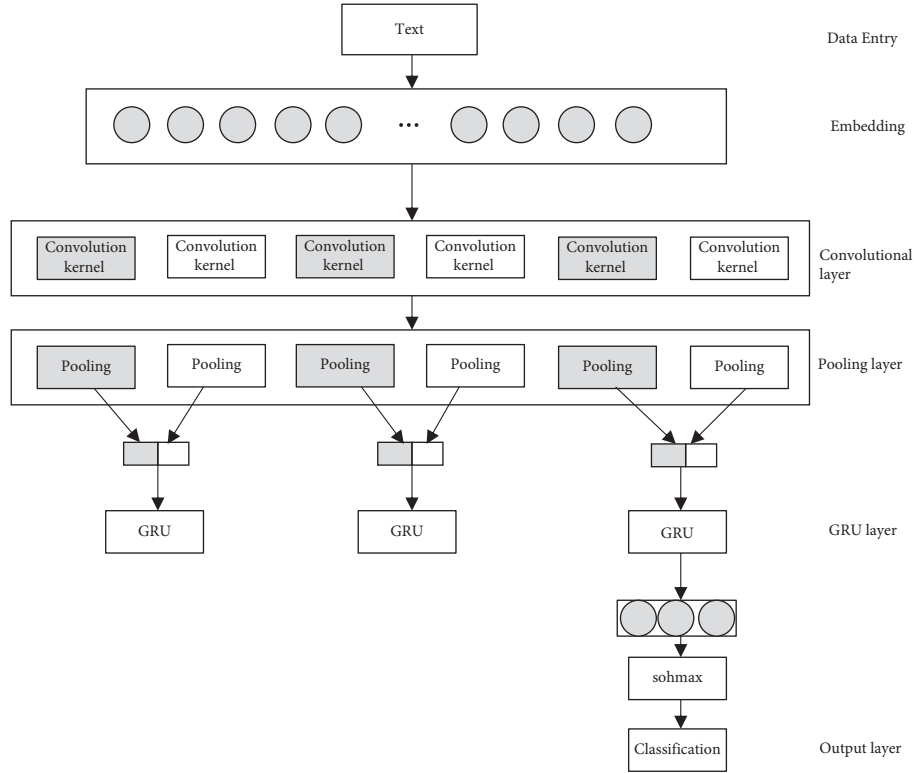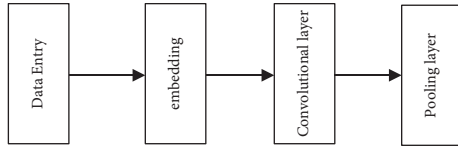
FIGURE 7: Fused CNN-GRU model structure.

FIGURE 8: Convolutional neural network text classification process.

represents a feature representation of the output, b represents a bias term, and f represents a nonlinear activation function that is used when the model is trained to learn.

The ReLU function is used as an activation function to speed up convergence, and it is expressed as follows:

$$f(x) = \max(0, x).\qquad(6)$$

The convolutional kernel w and the input layer are convoluted one-time for each convolution, i.e., the corresponding feature map ($s \in R"$-ht) expression is as follows:

$$S = [s_1, s_2, \ldots, s_{m-h+1}].\qquad(7)$$

3.3.3. Pooling Layer.
In the case description and factual part of the legal document, the text vector is convolved in each convolution kernel, and the local characteristic vector using each convolutional kernel is obtained. The CNN-GRU network model adopts the maximum pooling operation. In the process of text classification, as the main extraction of important words can represent the entire text representation, the use of maximum pooling operation can be achieved

is this function to obtain the most important features of the case description and fact part.

3.3.4. GRU Layer.
After the operation of the convolutional layer, the output of the pooling layer is used as the input of the GRU layer, and each pooling input vector can be regarded as the current moment t data input x in the GRU module. The data output of GRU at the t moment is as shown in Equation 8.

$$h_t = (1 - z_t) * h_{(t-1)} + z_t * h_t.\qquad(8)$$

3.3.5. Data Output Layer.
The vector output by the GRU network module is sent to the Softmax function classifier for classification.

3.4. Experimental Results and Analysis.
The main content of this section is to introduce the dataset used in the experiment, the division of the training set, the validation set, and the test set, and the specific form of the dataset storage. During the experiment, the parameters of the CNN-GRU neural network model of the fusion model are set, and the experimental results of the fusion network model are compared and analyzed with the results of other models.

3.4.1. Experimental Dataset.
The experimental datasets in this chapter are based on the datasets of the first phase of the 2018 China "Legal Research Cup" Legal Intelligence Challenge.

```
act: fact description
meta: annotation information, which includes:
Criminals: defendants (each contains only one defendant)
punish_ of_ money: fine (unit: yuan)
ACCUSATION : guilty
relevant_ _articles: relevant laws
term_ of_ imprisonment: sentence
Format of sentence (unit: months).
death_ penalty: whether to die for death
life_ imprisonment: indefinite
imprisonment: imprisonment for imprisonment
A simple piece of data is as follows:
{
"fact":
```

ALGORITHM 1: Interval division algorithm.

```
"meta":
{
"relevant_ articles": [234],
"Accusation": ["intentional injury"],
"punish_ of_ money": 0,
"criminals": ["Zhou X"],
"term_ of_ imprisonment":
"death_ penalty": false,
"imprisonment": 6,
"life_ imprisonment": false
}
}
```

ALGORITHM 2: Logistic regression algorithm.

CAIL2018-small data contain 196,000 instrument samples, including 150,000 training sets, 16,000 validation sets, and 30,000 test sets. In these data, a total of 183 laws are involved. Data are stored in json format, with one piece of data for each behavior, and each piece of data is a dictionary. The data are in the following form:

"The People's Procuratorate of Ningxiang County, Hunan Province, charged: at about 17 : 00 p.m. on August 1, 2016, the defendant,

Zhou *X*, and the victim, Luo *X* X, are in the field of Shuangjiangkou Town, Ningxiang County, because of farmland irrigation. During the conflict, defendant Zhou xxx injured the nose of the victim Luo xxx with his fist. After identification: Luo Moumou's injury is a minor injury of the second degree. After the case occurred, defendant Zhou XX compensated the victim Luo XX for various losses of 9,000 yuan and obtained forgiveness from the victim Luo XX.

3.4.2. Experimental Parameter Settings. The specific experimental environment configuration in this section is given in Tables 1 and 2.

(1) Preprocessing data, after removing pause words, personal names, place names, and other relevant

information, the length of text data is basically within 500, and the length of each text data is set to 500 in the experiment. The end of the end is not enough to make up 0, so that each case description and fact part are unified, a length for the effective input of the neural network structure for training.

(2) The word breaker tool for the text data in this section selects the Chinese word breaker tool based on the Python programming language Jieba word breaker tool.

(3) Word vector: this section experiment uses the Word2Vec tool developed by Google, and the Skip-gram model of the Word2Vec tool processes the text after data preprocessing to get the text after word segmentation, a 300-dimensional word vector for each word in the text.

(4) Convolutional neural network model input data vector dimension is 300. To be able to obtain the characteristics of the case description of more dimensions, the convolutional layer uses three convolutional kernel sizes, which are (3, 4, 5). Each size of the convolutional kernel is 32. The pooling layer uses the maximum pooling operation, and the pooling kernel size is 4. The effective vector output of

TABLE 1: Experimental environment configuration.

Experimental environment	Environment configuration
Mainframe	Aurora workstation
Operating system	Ubuntu16.04 TS
System type	64 Bit
Disk size	1T
Processor	Intel Xeon CPU E5-1603 v3@ 2.80Ghzx4
GPU	GPU Quadro M4000/PCIe/SSE2
Programming languages	Python3.5
Deep learning framework	TensorFlow

TABLE 2: Fused hyperparameters of the CNN-GRU model.

Parameter name	Learning rate	Dropout	Batch_size	Word embedding size
Parameter value	0.001	0.5	128	300

the pooling layer is sent to Softmax. The function classifier is classified and trained by the dropout strategy, and finally, the recommended classification prediction of the case description method is realized. The hyperparameters used in the training process are given in Table 2.

3.4.3. Experimental Results and Analysis. To obtain a comprehensive evaluation of the performance of each model, the prediction accuracy rate (acc)%, microaverage (F-micro) %, and macroaverage (F-macro) % are mainly used as evaluation indicators in the experimental process. Microaverage (F-micro) % and macroaverage (F-macro) % are also used as evaluation indicators for the prediction task of the proposed problem in other model experiments. For the effectiveness and practicability of the evaluation model, FastText was used in the legal recommendation problem task. The results of the TFIDF + SVM, convolutional neural network (CNN), and the fused CNN-GRU network model proposed in this chapter were compared on the same dataset using the evaluation index accuracy rate (acc), micromean (F-micro) %, and macroaverage (F-macro) %. The experimental results of using different models on the same dataset are given in Table 3.

From the above experimental results, it can be seen that when using the accuracy evaluation index, FastText, TFIDF + SVM, CNN, CNN-LSTM, and fused CNN-GRU network models achieve relatively high accuracy because the data are prediction of 183 laws, a multiclassification problem, and this accuracy rate is the correct accuracy of the overall prediction, if not considered. When the accuracy of each category is the case, these models can be used to a certain extent because the imbalance in the amount of case description information applicable to each law in the dataset used requires the use of microaverages and macroaverages as evaluation indicators to evaluate the performance of each model. It can be seen from the experimental results that the

TABLE 3: Each model method recommends predicting experimental results.

Model	Accuracy (acc) %	Microaverage (F-micro) %	Macroaverage (F-macro) %
FastText [12]	94.3	50.9	39.7
TFLDF + SVM [12]	94.0	73.9	56.2
CNN [12]	97.6	37.0	21.4
GRU	85.5	69.2	53.6
CNN-LSTM	97.9	79.6	68.3
CNN-GRU	98.2	83.2	72.6

fusion CNN-GRU network model structure proposed in this chapter is predicted in each category classification. It has improved a lot, improved the practicality and effectiveness of the mold to a certain extent, and made it possible to provide legal recommendations for auxiliary sentencing. To a certain extent, it is expected to become an effective tool for the application of legal intelligence.

According to the prediction experimental results of each model rule recommendation given in Table 3, we can use the line chart to clearly characterize the previous relationship of the results of each model. FastText, TFIDF + SVM, and CNN model are the solutions proposed in the literature to recommend the task; the overall accuracy of the performance is good in the GRU, CNN + LSTM, and the fusion CNN-GRU proposed in this study. The line chart of the prediction experimental results recommended by each model in the model is shown in Figure 9.

It can be seen from the line chart of the prediction experimental results recommended by the various model methods in Figure 9 that in FastText, TFIDF + SVM, CNN, GRU, CNN + LSTM, and the fused CNN-GRU model proposed in this study, when using the convolutional neural network (CNN) model, the experimental results show the lowest microaverage (F-micro) and macro-average (F-macro) accuracy. In the case of high data, it can be concluded that in the case of large amount of data and unbalanced data, the CNN model processes a large amount of text data, and the ability of the CNN model to extract text information features is very strong; however, when there are little data, the ability to extract text information features is weak. Use the GRU network model to process the same dataset. The accuracy rate, micromean (F-micro), and macroaverage (F-macro) values are relatively close (when the text classification is a binary classification, each category corresponds to a recall and accuracy rate, and the evaluation is a single class). In classification accuracy on the other, for n binary classification evaluations, the overall performance of the entire classification system can be evaluated by the accuracy of these individual categories. To comprehensively investigate the accuracy and recall rate on n binary classification confusion matrices, macroaverages and microaverages need to be used on n binary classification confusion matrices. From this, it can be seen that the GRU network model performs well in extracting text information features under the condition of unbalanced data volume. The converged CNN-GRU network model proposed in this paper combines
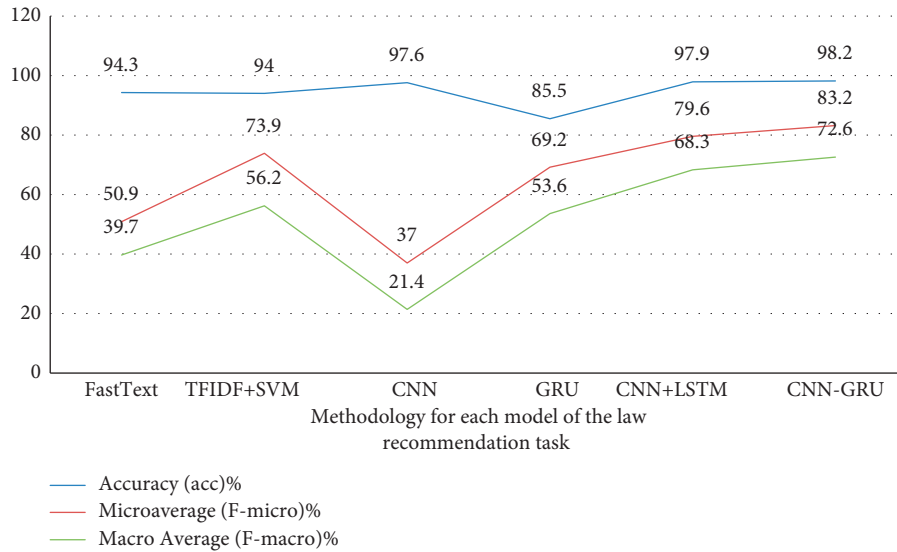
FIGURE 9: The line chart of each model method recommends predicting the experimental result.
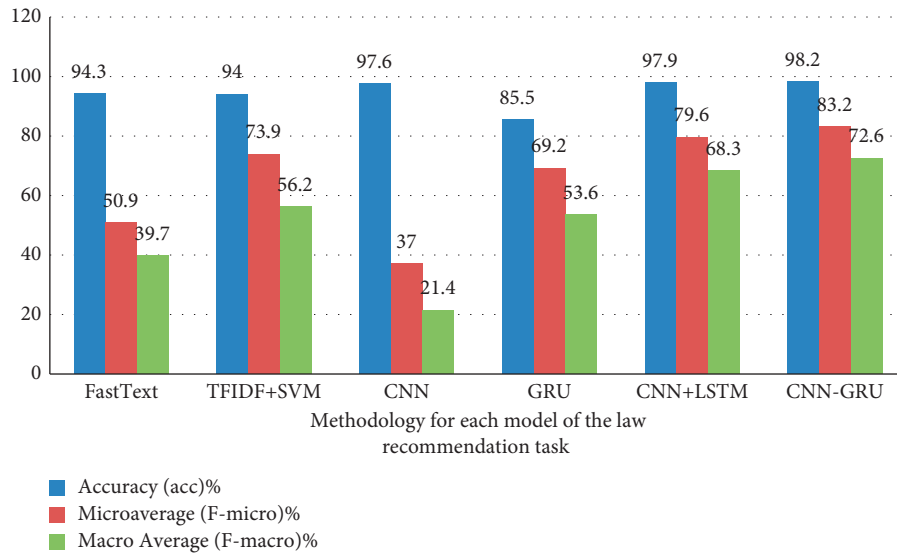

FIGURE 10: Each model method recommends a bar chart predicting experimental results.

the advantages of the two models of CNN and GRU to form a new network structure model as a model to solve the problem of law recommendation. From Figure 9, it can be seen that in the experimental results of each model, fusion is proposed.

The results of the CNN-GRU model experiment achieved the best results. Thus, the effectiveness of the proposed converged CNN-GRU network model is further verified. According to the experimental result data, the column chart recommended by each model method can be obtained, and the difference between the different evaluation indicators of fastText, TFIDF + SVM, CNN, GRU, CNN + LSTM, and the fusion CNN-GRU model proposed in this paper can be clearly seen in terms of accuracy (acc), micromean (F-micro), and macromacro different.

As can be seen from Figure 10, of FastText, TFIDF + SVM, CNN, GRU, CNN + LSTM, and the fused

CNN-GRU models proposed herein, in terms of overall accuracy, FastText, TFIDF + SVM, CNN, CNN + LSTM, and fusion CNN-GRU models are well-behaved in the literature as proposed models. On the TFIDF + SVM combination model to achieve the best results, TFIDF is word frequency–inverse text frequency that has two parts of the composition such as TF and IDF, where TF is the word frequency, which can be understood as the frequency statistics of each word that appears in text data and as a feature of text data. After extracting the features, they are then combined with the SVM algorithm to form the TFIDF + SVM model algorithm, because the SVM algorithm is based on the theoretical basis of statistical learning, and the text data are trained by vectorized text to obtain a higher classification accuracy. Therefore, it can be known that under the premise of extracting the effective characteristics of the text data and then through the good performance of

the classifier, the better classification performance of the model can be achieved. The accuracy rate (acc), micro-average (F-micro), and macros can also be seen from Figure 10.

The average (F-macro) has different evaluation indicators, which more typically solves the task problems proposed in other models; the proposed fusion CNN-GRU network model proposed in this study, according to the characteristics of legal text data, in the extraction of case description and fact part of the information characteristics and the use of extracted case description and fact part of the information characteristics, obtained the best performance indicators. It further shows that the extracted converged CNN-GRU network model has a certain practical value, and to a certain extent, it is expected to help the relevant judicial personnel improve their work efficiency.

4. Conclusion

This paper proposes a fusion CNN-GRU network result model that uses a text classification neural network structure convolutional neural network with a GRU neural network model structure to carry out the task of recommending the law. After experimental verification, the fusion CNN-GRU structure model is proposed, and good results are achieved, which improves the convolutional neural network by only extracting the overall local characteristics to represent the data of the entire text because the case description is before and after, the context is related, and a single convolutional neural network structure cannot be used. In this paper, the case description and fact part of the case are extracted using the convolutional network structure. The semantic representation of text data is extracted using the GRU network structure model, and finally, the Sofmax function is used for effective classification. By training the model through experimental datasets, the results show that the proposed fusion CNN-GRU model structure in this paper can obtain effective feature representations and can obtain context-rich semantic representations, which can achieve better results in multicategory tag classification tasks, which has a certain practical value.

Prospects: artificial intelligence is still in the initial stage of development. Theoretical technology needs to be further developed. Court sentencing problems are related to many kinds of condition factors. Hence, legal intelligence wants to achieve the same sentencing effect as human intelligence. There is still a long way to go and a lot of work tasks to do. This paper on deep learning is based on legal intelligence tasks. Although there is a stage of preliminary results, but there are still many task areas that need to be explored and studied. In the future, we will prepare to further study and explore the following issues:

(1) In the process of text classification, the data first need to be preprocessed. Key information is extracted and then input to the model; If the model is improved so that the data requirements of the model are not so strict, the text classification task can be completed by inputting the original data, which is an important problem for optimizing the neural network model structure or proposing a new classification model in the future.

(2) In the legal intelligence problem, especially in crime prediction, legal recommendation, and sentence prediction task, this research mainly studies the crime prediction and legal recommendation task, and it will further explore and study sentence prediction in the future. The research method tries to use interval division and logistic regression algorithm to solve the problem.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

References

[1] L. Adam, "One small step for 'artificial intelligence' and a giant leap for the Australian patent system? The Federal Court decision in Thaler v Commissioner of Patents," *Journal of Intellectual Property Law & Practice*, vol. 17, no. 2, 2022.

[2] D. S. W Gomes, F. R. Antunes, P. H. S Bermejoa, A. Gersicada, S. Gonçalob, and B. S. Meloc, "Artificial intelligence and speedy trial in the judiciary: myth, reality or need? A case study in the Brazilian Supreme Court (STF)," *Government Information Quarterly*, vol. 39, no. 1, 2022.

[3] G. Elinor, "Balancing the scales in China's smart courts: driving case standardisation through AI," *Peking University Law Journal*, vol. 9, no. 2, 2021.

[4] G. Zheng George, "China's grand design of people's smart courts," *Asian Journal of Law and Society*, vol. 7, no. 3, 2020.

[5] A. D. Dory, "Reiling. Courts and artificial intelligence," *International Journal for Court Administration*, vol. 11, no. 2, 2020.

[6] C. Intahchomphoo, A. Vellino, O. Erik Gundersen, C. Tschirhart, and E. Shaaban, "References to artificial intelligence in Canada's court cases," *Legal Information Management*, vol. 20, no. 1, 2020.

[7] E. Troulinos, "The prospects of artificial intelligence in a court information system," in *Proceedings of the CEUR Workshop*, p. 2844, Copenhagen, Denmark, March 2020.

[8] P. Bhattacharya, P. Mehta, K. Ghosh et al., "Overview of the FIRE 2020 AILA Track: Artificial Intelligence for Legal assistance," in *Proceedings of the CEUR Workshop Proceedings*, p. 2826, December 2020.

[9] M. Chen, "Beijing Internet Court denies copyright to works created solely by artificial intelligence," *Journal of Intellectual Property Law & Practice*, vol. 14, no. 8, 2019.

[10] J. Zeleznikow, "Can artificial intelligence and online dispute resolution enhance efficiency and effectiveness in courts," *International Journal for Court Administration*, vol. 8, no. 2, 2017.

[11] H. Matthew, "Artificial Intelligence Prevails at Predicting Supreme Court decisions," *Science*, 2017.

[12] L. Yu, T. He, Y. Ge, S. Zhang, and H. Wang, "Using Case Facts to Predict Penalty with Deep Learning," in *Proceedings of the Abstracts of the 5th International Conference of Pioneering Computer Scientists,Engineers and Educators(ICPCSEE 2019) Part II*, p. 73, Springer, Guilin, China, September 2019.

[13] C. Wang and Z. Song, "A study on promoting the mechanism of the rule of thumb from the perspective of the smart court in China," *Journal of Studies in Social Sciences*, vol. 18, 2019.

[14] P. Giulia and P. Webb, "Litigation before the international court of justice during the pandemic[J]," *Leiden Journal of International Law*, vol. 34, no. 4, 2021.

[15] C. Patrick and O. H. Palmer, "Explaining power and authority in international courts[J]," *Leiden Journal of International Law*, vol. 34, no. 3, 2021.

[16] J. Florian, K. F. Sarah, K. Ashley, and D. B. Jeffrey, "Analyzing mass attitudes toward the international court of justice," *Foreign Policy Analysis*, vol. 17, no. 2, 2021.

[17] S. Theresa, "Judicial diplomacy: international courts and legitimation," *Review of International Studies*, vol. 47, no. 1, 2020.

[18] H. Desai Bharat and K. Sidhu Balraj, "International courts and tribunals – the new environmental sentinels in international law," *Environmental Policy and Law*, vol. 50, no. 1-2, 2020.

[19] S. Allioui, "Legitimacy and international courts," *International Journal for Court Administration*, vol. 11, no. 1, 2020.

[20] M. Li, *An Intelligent Flood Evacuation Model Based on Deep Learning of Various Flood scenarios*, Kyoto University, Kyoto, Japan, 2021.