

Research Article

Recognition of Chinese Legal Elements Based on Transfer Learning and Semantic Relevance

Dian Zhang^{ID},¹ Hewei Zhang^{ID},¹ Long Wang^{ID},¹ Jiamei Cui^{ID},¹ and Wen Zheng^{ID}^{1,2}

¹Institute of Public-Safety and Big Data, College of Data Science, Taiyuan University of Technology, Taiyuan 030060, China

²Center for Healthy Big Data, Changzhi Medical College, Changzhi, Shanxi 046000, China

Correspondence should be addressed to Wen Zheng; zhengwen@tyut.edu.cn

Received 29 August 2021; Revised 2 March 2022; Accepted 13 April 2022; Published 30 April 2022

Academic Editor: Yan Huang

Copyright © 2022 Dian Zhang et al. 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.

In recent years, LegalAI has rapidly attracted the attention of AI researchers and legal professionals alike. Elements of LegalAI are known as legal elements. These elements can bring intermediate supervisory information to the judicial trial task and make the model's prediction results more interpretable. This paper proposes a Chinese legal element identification method based on BERT's contextual relationship capture mechanism to identify the elements by measuring the similarity between legal elements and case descriptions. On the China Law Research Cup 2019 Judicial Artificial Intelligence Challenge (CAIL-2019) dataset, the final result improves 4.2 points over the method based on the BERT model but without using similarity metrics. This research method makes full use of the semantic information of text, which is essential in the judicial field of document processing.

1. Introduction

At the heart of the law is language, and natural language processing technologies have long been used in the legal field to support many tasks that would benefit from structured data and automated legal reasoning, such as better search and information retrieval, compliance checking and decision support, and better presentation of legal information to professional and nonprofessional stakeholders [1]. In July 2016, the National Strategy for the Development of Information Technology (NSIDC) proposed the construction of “smart courts” one of the core objectives of providing intelligent assistance to judges in handling cases, including case pushing, sentencing assistance, and document generation. The core goal of the construction is to provide intelligent assistance to judges in handling cases, including case pushing, sentencing assistance, and document generation [2]. This development for automated analysis, indexing, etc. creates opportunities for new approaches to improve the legal system's efficiency, comprehensibility, and consistency. For example, one approach is to extract syntactic elements from

legal documents that satisfy specific semantics. These meanings can be everyday and specific to legal information and have particular significance for practitioners in private practice, government, public administration, education, and research.

Legal documents contain descriptions of the facts of the case. To improve the efficiency of the legal system, we can extract key elements from a particular sentence of the case description. The main goal of the research in this paper is to extract the key elements from the case description. The process is shown in Figure 1. The key elements are extracted from the law by legal professionals and are easy to understand. We extract the relevant elements from the case descriptions with the help of deep learning algorithms and the final results are used in the decision-making process.

The method proposed in this paper is suitable for text data and extracts the key information in the text. The extraction result is used as intermediate auxiliary information. On the one hand, it can provide work basis for judicial staff, on the other hand, it can be used as basic information for downstream tasks and finally applied to real life, as shown in

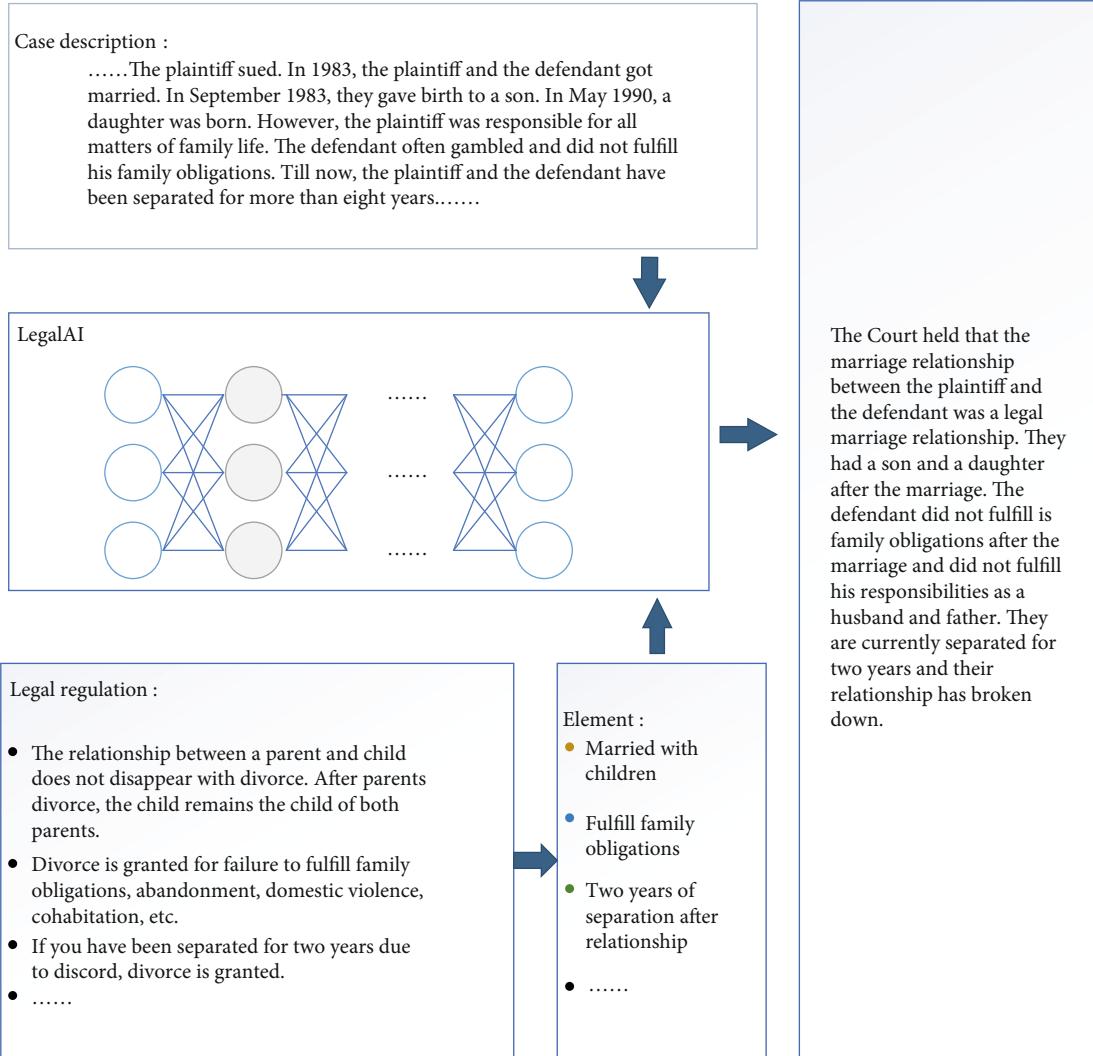


FIGURE 1: A block diagram of the proposed research.

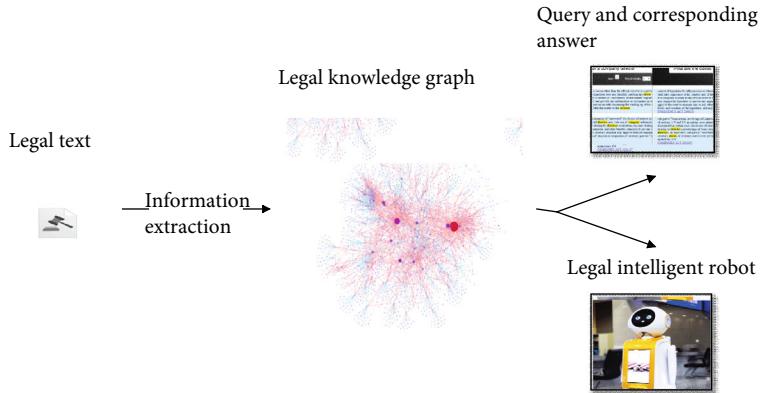


FIGURE 2: The research results in this paper can eventually be used as supporting information for the construction of the above application scenarios. This includes legal intelligent consulting systems and for legal intelligent robots, etc. [3, 4].

Figure 2. Furthermore, we outline the application directions of this study as shown in Figure 3. However, the situation in actual case descriptions is relatively complex. A sample of

data often expresses multiple semantics of great complexity, and the relevance of sentences and elements in the legal domain is vital for understanding the interpretation and

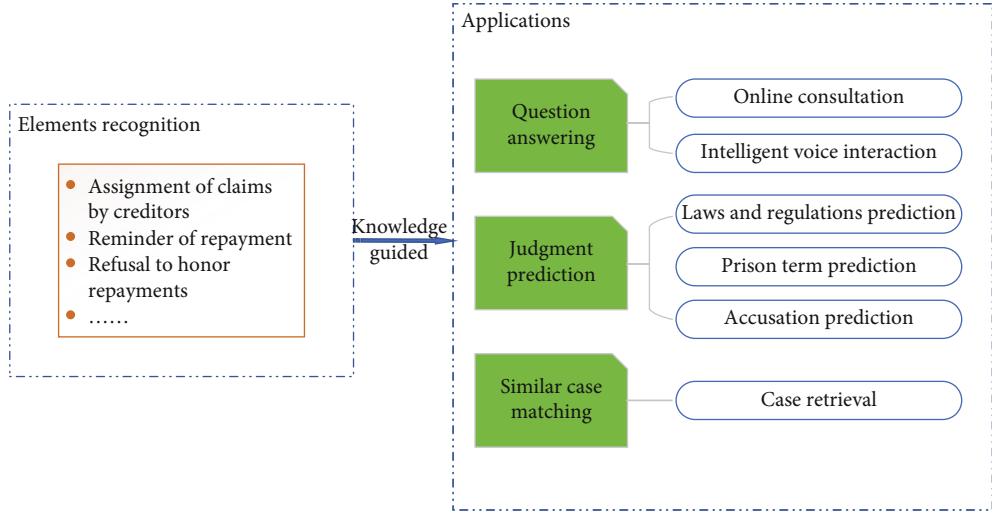


FIGURE 3: An overview of the application directions of this research.

application of the law [5]. In the process of building a “smart court,” the element identification task plays an important role, aiming to extract key elements from legal documents automatically. The purpose is to extract key elements from legal documents automatically. For example, “The plaintiff Zou Mou A claims: I gave birth to a son named Zou Mou B with the defendant Pan Mou on June 6, 2010; on November 1, 2010, both parties registered for divorce and agreed that The defendant would raise Zou Mou B and I would pay monthly alimony.” Support, payment of alimony, and payment of alimony monthly. This information has a significant reference value for the trial outcome and is critical in the trial process. Trial experts generate the critical elements based on case analysis. The extracted results can be used in practical operations in the judicial field, such as case summary, interpretable class case push, and related knowledge recommendation. This paper proposes a new method for identifying legal elements, which can learn the complex semantic information in the documents and help uncover the key elements of legal documents, and is of outstanding technical significance and practical importance for promoting the development of “smart court.”

2. Related Work

Legal research is usually the process of finding the information needed to support legal decisions. In practice, this usually means searching for content relevant to the particular matter at hand [6]. LegalAI also has its exclusive notation, named legal elements. The extraction of legal elements focuses on extracting key elements, such as whether someone has been killed or whether something has been stolen. These elements are known as the constituent elements of a crime, and we can directly convict an offender based on the outcome of these elements. These elements bring intermediate supervisory information to the judgment prediction task and make the model’s predictions more interpretable. In natural language processing, several classical coding models are used to extract features [7–10] and perform ele-

ment recognition tasks. They can automatically extract key information from text and apply it to downstream tasks. Typical algorithms include ML-KNN [11] based on the KNN algorithm, ML-DT [12] based on the decision tree algorithm, and Rank-SVM [13] based on the SVM algorithm; Yildirim et al. [14] analyzed text to extract noun phrases from it and trained a support vector machine classifier to determine whether the noun phrases were genuine attributes; Chen et al. [15] compared conditional random fields with a variety of other methods for element recognition, including hidden Markov models and association rule-based statistical methods, and the experiments show that conditional random fields are more suitable for element recognition tasks. However, machine learning methods often use phrases as features for classification, ignoring contextual information.

In recent years, deep learning has achieved perfect results on text classification task. Kim [16] used the TextCNN method to perform text feature extraction after obtaining vector representations of text, which can capture local semantic information; Liu et al. [17] used the deep learning method RNN to capture global contextual information by passing information from one moment to the next, which can also achieve good results. However, the position of words in the sentence in the case description of legal documents has a more critical impact on the case outcome. Although we can capture local or global contextual information by CNN and RNN, we cannot express the position information of words and the interrelationship of other words in the text simultaneously [17, 18]. The above algorithms require a large amount of labeled data for training and are very costly. To overcome this limitation, a method called “transfer learning” (in this paper, fine-tuning is used as a transfer learning technique) is introduced in this paper [19, 20]. The basic rule of this approach is to reuse the model trained for a specific task as a starting point for the model trained for the target downstream task. In the last two to three years, migration learning has shown extraordinary results in most computer vision tasks [21, 22], and in today’s

TABLE 1: Results of elemental analysis.

Case description:	
.....the plaintiff sued. In 1983, the plaintiff and the defendant got married. In September 1983, they gave birth to a son. In May 1990, a daughter was born. However, the plaintiff was responsible for all matters of family life. The defendant often gambled and did not fulfill his family obligations. Till now, the plaintiff and the defendant have been separated for more than eight years.	
Married with children	Yes
Fulfill family obligations	No
Two years of separation after discord in a relationship	Yes
Case result: (the plaintiff won the case.) divorce granted	

TABLE 2: Sample data.

Sentence	Label list
Sentence 1	[]
Sentence 2	["DV1"]
Sentence 3	["DV1", "DV4", "DV2"]

practice, researchers rarely train deep learning models from scratch. Transfer learning, previously restricted to CV tasks, can now also be performed in the natural language processing domain, introducing recent language representation models [23–25], and the latest technology, Google’s BERT [26]. Transfer learning performs well in natural language understanding tasks.

For the element recognition task of legal documents, this paper proposes a semisupervised migratory learning approach based on BERT’s contextual relationship capturing mechanism to accomplish element recognition by measuring the similarity between legal elements and case descriptions. The core idea of this approach is to use the transfer learning capability of the BERT pretrained language model to represent each sentence of the text and Chinese element features and finally use cosine similarity to measure the similarity between the sentence feature vector and the Chinese element feature vector, to predict whether the Chinese label is a critical element of the text.

3. Methods of Legal Element Recognition

3.1. Element Recognition. In the judicial domain, the primary purpose of the case element recognition task is to automatically extract critical factual descriptions from the case description. The results of the case element recognition can be used in practical business requirements in the judicial field, such as case summaries, interpretable case pushing, and relevant knowledge recommendation. After the relevant paragraph of a given judicial document, the system analyzes and judges each sentence to identify the critical case elements. Table 1 shows that the recognized elements can determine the final judgment results. It shows that legal elements are important for downstream tasks.

Table 2 shows the three sentences in the instrument. It is important to note that each sentence in the instrument cor-

TABLE 3: Chinese semantic information of character tags.

English characters	Chinese semantics
DV1	Semantics1
DV2	Semantics2
DV4	Semantics3

responds to a variable number of category labels, which may contain one or more or even no critical elements at all. The labels are represented using English characters and have corresponding Chinese semantics, as in Table 3. The Chinese semantics is given by the relevant experts in the field and is a vital reference for the orientation of judicial outcomes.

This task covers three areas in total, including marriage and family, labor disputes, and loan contracts. Before the element identification task, the English labels in the dataset needed to be replaced using Chinese labels. The Chinese semantics of all tags needed to be judged and whether there was a contextual connection with each sentence. If two sentences are contextually related, they are tagged, indicating that the tag belongs to the critical elements of the sentence and is stored in the list; if it is not tagged, it means that the tag does not belong to the critical elements of the sentence and no other processing is done.

3.2. BERT Model. The BERT model is a language representation model based on a large amount of unsupervised data trained by Google released in 2018, using transformer’s powerful information extraction capabilities to build pretrained language models with strong generalization capabilities using massive amounts of data. The BERT model uses the encoder part of transformer [27], and its model structure is shown in Figure 4.

The transformer is a self-attention based model that uses the encoder-decoder structure in seq2seq, where both the input and output are sequences. The decoder decodes this fixed-length vector into a variable-length output sequence.

This idea is based on RNNs, unlike RNNs, where the gradient of each input depends not only on the computation of the current step but also on the data of the previous time steps. Therefore, RNNs have the disadvantage of not being parallel and running slowly. Elements also take into account the relationship between any two elements of the input [28].

The structure of the encoder in the transformer model is shown in Figure 5. The input to the encoder is a word embedding representation of a sentence, which is fed into the self-attention layer with the location information of each word in the sentence. Self-attention computes the relationship of each word in a sentence to other words, embedding the semantic information of the whole sentence into the word vector, and is the most powerful module in the encoder part.

Previous work on pretraining representations, such as in OpenAI, GPT, and ELMo, is one-way and shallow bidirectional, respectively, whereas BERT is deep bidirectional. Bert removes the constraints provided by the one-way approach by using a masked language model as a pretraining target. BERT uses three embedding layers in its input

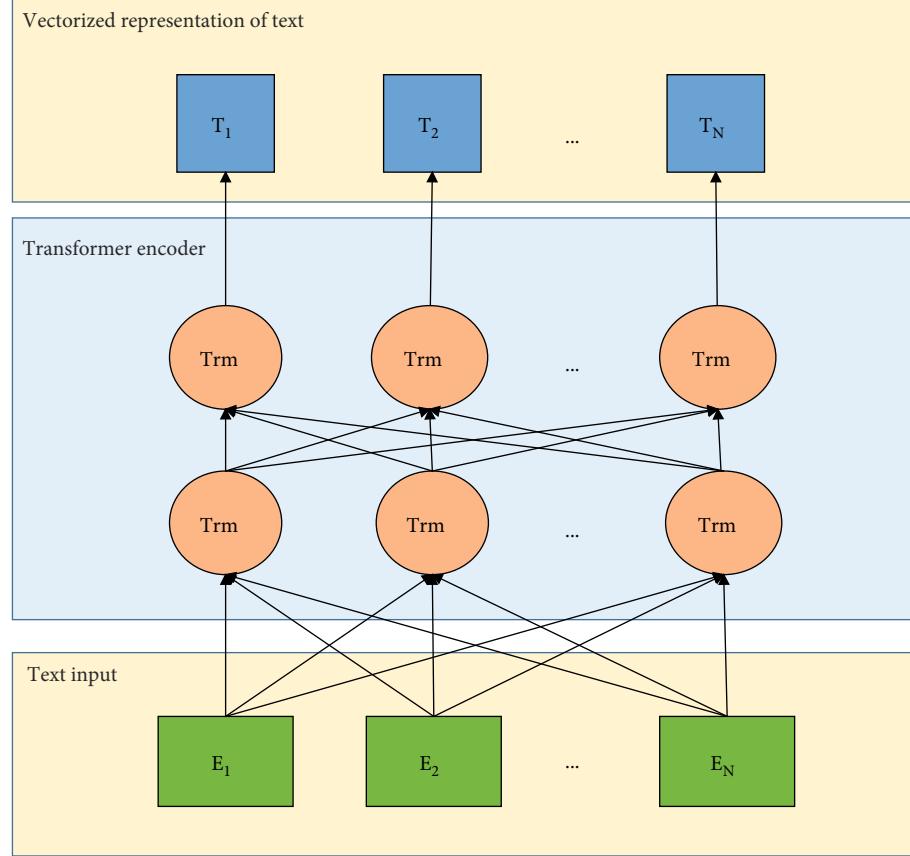


FIGURE 4: BERT model structure.

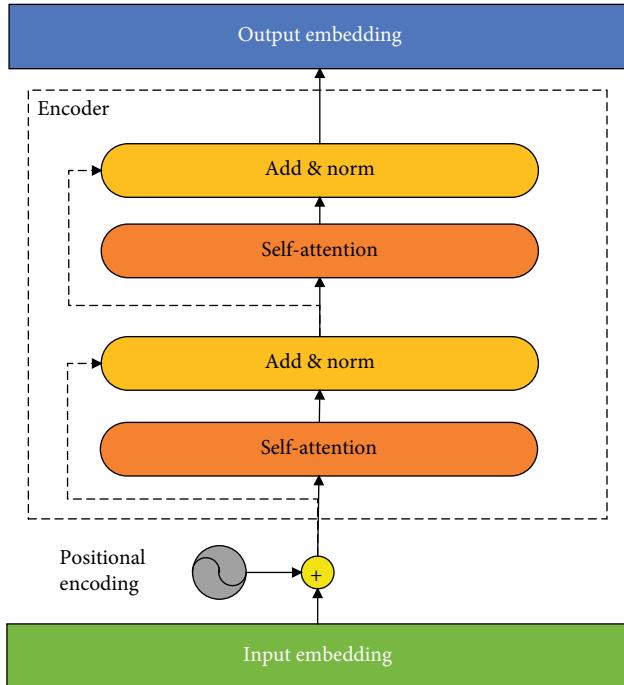


FIGURE 5: Transformer encoder structure.

representation, where the token embedding converts words into a fixed 768-dimensional vector representation of the [CLS] and [SEP] tokens are appended to the beginning and end of the tokenized sentence; they are used as the input representation for the classification task and to separate the input text. If given a pair of text inputs, the segmentation embedding of BERT can perform the text classification task. An example of this is to classify two texts if they are semantically similar. Thus, the texts are concatenated and fed to BERT, which differentiates the texts with the help of segmented embeddings. The positional embedding in BERT uses the learned positional embedding; this is done using the function used to compute the positional encoding in the transformer. The positional embedding used here understands the relative position, not just the absolute position. This is achieved by adding a sine function to a 768-dimensional vector representation of the word, which depends on the position of the word i in the sentence sequence, and the position of j in the embedding feature position, which produces slightly of the same word in different positions. The final input embedding is the sum of the three embeddings.

BERT uses a novel approach to use bidirectionality by pretraining the “masked language model” and “next sentence prediction” rather than pretraining the language model. All word point tokens are masked 15% of the time, with words replaced by [MASK] tokens 80% of the time,

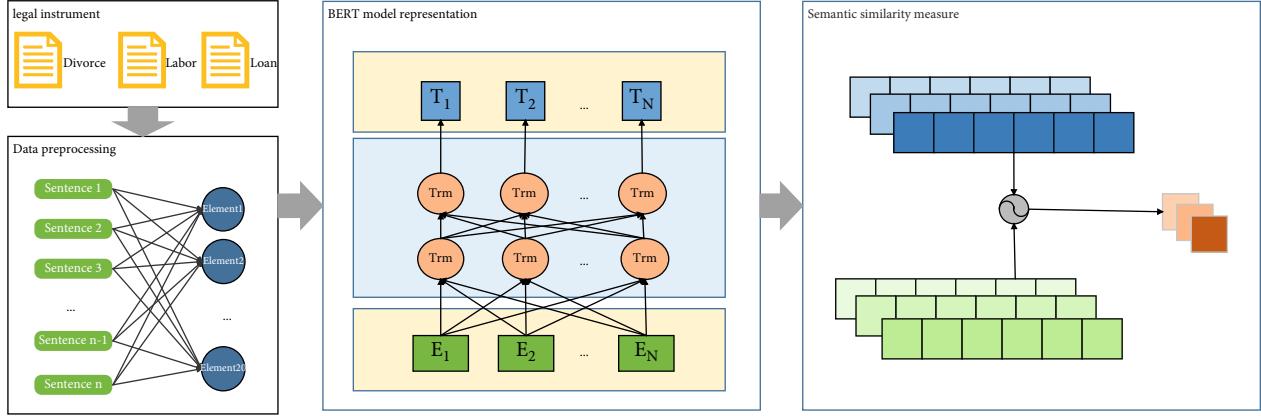


FIGURE 6: Element recognition framework diagram.

Input: a list of each sentence and corresponding label in the legal document doc, $data\{sen_1 : [lab_{11}, lab_{12}, \dots], sen_2 : [lab_{11}, \dots], \dots, sen_n[lab_{11}, lab_{12}, \dots]\}$;
Output: Each sentence of the legal instrument and all combinations of elements, and a 0, 1 label. If the label is 0, then the sentence does not contain the element; if the label is 1, the sentence contains the element. $T = [\{l, (sen_1, lab_1)\}, \{l, (sen_1, lab_2)\}, \dots, \{l, (sen_1, lab_{20})\}, \dots, \{l, (sen_2, lab_1)\}, \dots, \{l, (sen_n, lab_1)\}, \dots, \{l, (sen_n, lab_{20})\}]$

1. $T = \emptyset$
2. For sen in doc do
3. For i=1 to 20
4. Algorithm 1 Data preprocessing if lab_i in $data[sen]$: l=1
5. Else: l=0
6. T.attend($\{l, (sen, lab_i)\}$)
7. End for
8. End for
9. Return T

ALGORITHM 1: Data preprocessing.

Input: Combination T of sentences and elements of legal instruments
Output: List of elements for each sentence

1. $result = \emptyset$
2. For t in T do
3. Using the BERT model for sentence t.sen and element t.lab representations, it obtains the sentence vector bs, bl
4. Calculating similarity using cosine similarity
5. Calculation using schematic functions:

$$X = \text{similarity}(ds, dl) = ds \cdot dl / \|ds\| \times \|dl\| = \sum_{i=1}^n ds_i \times dl_i / \sqrt{\sum_{i=1}^n ds_i^2} \sqrt{\sum_{i=1}^n dl_i^2}$$
6. If $I(x)=1$: Result[t.sen].Attend(t.lab)
7. Else: Continu
8. End for
9. Return result

ALGORITHM 2: Element identification method for Chinese legal documents.

random words 10% of the time, and words remaining unchanged the rest of the time. The model attempts to predict the correct value of the masked words based on the con-

text given by the unmasked words in the sequence. Technically speaking, predicting the output words requires three steps:

TABLE 4: Marriage and family domain element label.

Marriage and family Element label	Chinese semantics
DV1	Married with children
DV2	Maintenance of a child with restricted capacity
DV3	With a community of property
DV4	Payment of alimony
DV5	Division of real property
DV6	Live apart after marriage
DV7	Second divorce proceedings
DV8	Monthly child support payments
DV9	Granting of divorce
DV10	Have joint debts with spouse
DV11	Premarital personal property
DV12	Legal divorce
DV13	Failure to meet family obligations
DV14	Existence of children born out of wedlock
DV15	Appropriate help
DV16	Failure to comply with the divorce agreement
DV17	Compensation for damages
DV18	Two years of separation after discord in a relationship
DV19	The children live with a non-parent
DV20	The personal property after the marriage

- (1) A classification layer is added to the top of the encoder layer
- (2) The embedding matrix is multiplied by the output vector to convert it into the dimensionality of the vocabulary
- (3) The probability of each word in the vocabulary is calculated using softmax

MLM captures the relationship between two sentences, vital for questions and answers and natural language inference tasks. To better understand the relationship between two sentences, the BERT authors used next sentence prediction, simply a classification task to discover whether sentence B follows sentence A. 50% of the training examples were correct, and the rest were randomly selected to generate a pair of incorrect sentences.

3.3. Legal Element Recognition Method. In summary, this paper proposes a legal element recognition method (BERT-LER) based on the BERT model and completes the characterization of text data by using the transfer learning ability of the BERT model. Then, this paper extracts the elements of sentences by calculating the cosine similarity of sentences and elements, as shown in Figure 6. First, we preprocess the legal documents into a single sentence. Each sentence and all the elements form a sentence combination. Second, the sentence combinations need to be represented vectorially and by calculating the semantic similarity. We identify the criti-

cal elements of the sentences based on their relevance and finally output all the key elements of each sentence.

The specific process is as follows. Algorithm 1 introduces the data preprocessing part, and Algorithm 2 introduces the method proposed in this question.

4. Experimentation and Analysis Subheadings

The results and discussion may be presented separately or in one combined section and may optionally be divided into headed subsections.

4.1. Experimental Data. The dataset used in this paper is from the 2019 China Law Research Cup Judicial Artificial Intelligence challenge and is selected from legal documents publicly available on the Chinese judicial documents website. Each row of the data represents the result of a sentence of a part of a paragraph extracted from a judgment document, as well as a list of the sentence's element labels. The three main areas of adjudication are marriage and family, labor disputes, and loan disputes, with 2,740 cases, including 1,269 marriage and family cases, 836 cases of labor disputes, and 635 cases of loan disputes. The data were all annotated by professionals with a legal background, and each of the three fields has 20 element labels and the Chinese semantics they represent, as shown in Tables 4–6.

4.2. Evaluation Indicators. The evaluation metrics used in this paper include microaverage $F1$ values and macroaverage $F1$ values.

The $F1$ value is an indicator that combines P and R and is calculated as follows.

$$F_{\beta} = \frac{(\beta^2 + 1) * P * R}{\beta^2 * P + R}. \quad (1)$$

In particular, at this point, the F value is based on a weighted summation average of P and R . $\beta > 0$ measures the relative importance of R to P and is usually taken to be $\beta = 1$, at which point the formula degenerates to the standard $F1$ value, as shown in the following equation.

$$F1 = \frac{2 * P * R}{P + R}. \quad (2)$$

Macroaveraged $F1$ values (“Macro_ $F1$ ”) are obtained by first counting the indicator values for each class and then averaging them arithmetically over all classes, as shown in the following equation.

$$\text{Macro_}F1 = \frac{1}{k} \sum_{i=1}^n \frac{2 * P_i * R_i}{P_i + R_i}, \quad (3)$$

where “ k ” is the number of categories, and “ i ” is the “ i ” category.

The microaverage $F1$ value (“Micro_ $F1$ ”) is a global confusion matrix created by counting each instance in the dataset without classification and then calculating the

TABLE 5: Labeling of elements in the field of labor disputes.

Labor dispute Element label	Chinese semantics
LB1	Dissolution of labor relations
LB2	Payment of wages
LB3	Payment of financial compensation
LB4	Failure to pay total remuneration for work
LB5	Existence of labor relations
LB6	No labor contract signed
LB7	Sign a labor contract
LB8	Payment of overtime wages
LB9	Compensation for payment of double wages for unsigned labor contracts
LB10	Paying compensation for injuries at work
LB11	Not initiated at the labor arbitration stage
LB12	Non-payment of compensation for unlawful termination of labor relations
LB13	Economic redundancies
LB14	No bonus payment
LB15	Illegal collection of property from workers
LB16	Special types of work
LB17	Payment of death benefits funeral benefits pensions
LB18	Cancellation by the employer with prior notice
LB19	The legal person status has been lost
LB20	With mediation agreements

TABLE 6: Labeling of elements in the area of borrowing disputes.

Borrowing disputes Element label	Chinese semantics
LN1	Assignment of claims by creditors
LN2	The amount borrowed x million yuan
LN3	With proof of borrowing
LN4	The lender has a relationship with the financial institution
LN5	Repayment of a loan
LN6	Company unit other organizations loans
LN7	Joint and several warranties
LN8	Reminder of repayment
LN9	Interest payments
LN10	Conclusion of warranty contracts
LN11	Have a written commitment to repay the loan
LN12	The security contract is invalid withdrawal release
LN13	Refusal to honor repayments
LN14	Exclusion of the guarantor's warranty
LN15	The guarantor is not liable for the guarantee
LN16	The pledgee is associated with the company
LN17	The lender does not provide the loan on the agreed date amount
LN18	Many people borrowing
LN19	Assignment of debts by the debtor
LN20	The agreed interest rate is unknown

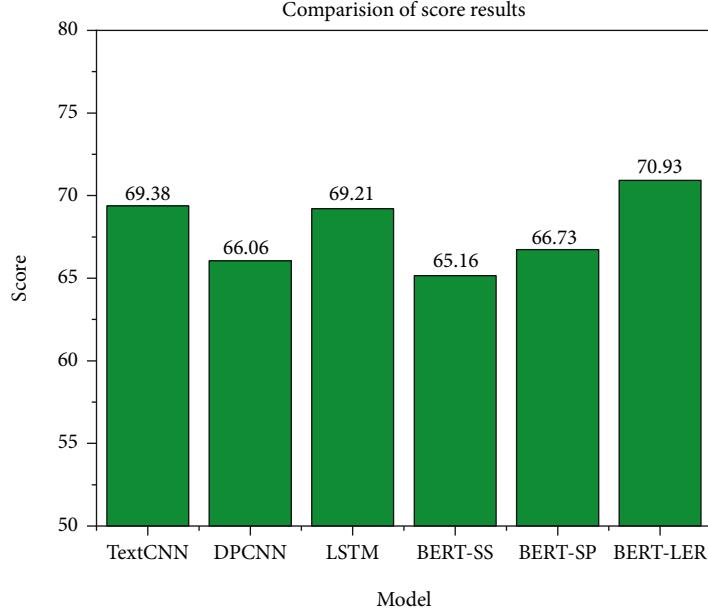


FIGURE 7: Comparison of experimental models.

TABLE 7: Comparative results of different methods.

Model	Divorce		Labor		Loan	
	MiF1	MaF1	MiF1	MaF1	MiF1	MaF1
TextCNN	78.7%	65.9%	76.4%	54.4%	80.3%	60.6%
DPCNN	81.3%	64.0%	79.8%	47.4%	81.4%	42.5%
LSTM	80.6%	67.3%	81.0%	52.9%	80.4%	53.1%
BERT-SS	83.3%	69.1%	76.8%	43.7%	78.6%	39.5%
BERT-SP	82.8%	67.6%	79.2%	47.1%	80.0%	43.7%
BERT-LER	83.1%	69.6%	80.9%	54.6%	81.0%	56.5%

corresponding metric, as shown in the following equation.

$$\text{Micro_F1} = \frac{2 \times \text{Micro_P} \times \text{Micro_R}}{\text{Micro_P} + \text{Micro_R}}. \quad (4)$$

The Macro_F1 value treats each category equally, so rare categories mainly influence its value, while the Micro_F1 value considers each document in the document set equally, so common categories more influence its value. The performance criteria of the model in the 2019 China Law Research Cup Judicial AI Challenge are evaluated using the score, which is calculated using “Micro_F1” and “Macro_F1”, as shown in the following equation.

$$\text{Score} = \frac{\text{Macro_F1} + \text{Micro_F1}}{2}. \quad (5)$$

4.3. Experimental Results. To verify the effectiveness of the new method, we conducted six experiments. TextCNN, DPCNN, and LSTM [29] were used as baselines. As shown in Figure 7, the migration learning method is used to achieve text vector representation, and cosine similarity is used to achieve element recognition with good results. The results show that the recognition of Chinese legal elements based

on transfer learning and semantic relevance is feasible for element recognition in the judicial domain.

We conducted experiments using three types of cases in the judicial field, including 1,269 marriage and family documents, 836 labor dispute documents, and 635 loan dispute documents. According to Table 7, the prediction results of marriage and family documents are good. Although the Bert model has strong generalization ability, a lot of data is also required for fine-tuning. In the experimental results, TextCNN performs better, because some elements in the data sample can be directly extracted by keyword matching. CNN has the ability of local feature extraction, so it has good performance in dealing with these problems. However, the gradient will disappear after the LSTM sequence length exceeds a certain limit, so it does not achieve the best performance.

To verify the effectiveness of our method, we conduct three independent experiments based on the BERT model. The first two groups use the two main downstream tasks of BERT, the single-sentence classification task, and the sentence-pair classification task. They use the two main downstream tasks of BERT, the single-sentence classification task, and the sentence-pair classification task. Sentence pair classification is done by forming sentence pairs for 20 elements in a legal document and determining the contextual relationship between them. The last set of experiments (BERT-LER) is the method proposed in this paper. Although BERT can learn global semantic information, the learned features are difficult to map to the correct elements. Although BERT-SP also considers the semantic information of elements, this is only reflected in the vector representations. As described in the last paragraph of introduction of BERT model, BERT still matches whether two sentences have a contextual relationship by probability in doing the sentence pair matching task. In this paper, however, by

TABLE 8: Dataset example.

Domain	Sentence	Element
Divorce	The plaintiff and the defendant registered their marriage after getting to know each other, and the foundation of the marriage was relatively good. At the early stage of their marriage, their relationship was good, and they gave birth to a son.	Married with children
Divorce	The defendant paid RMB 300 per month for the maintenance of the children until the children became independent.	Monthly support payments
Loan	Main reasons for layoffs: product structure, the sharp decline in orders, worsening financial situation.	Economic redundancies
Labor	The guarantee period had not yet expired when the Industrial and Commercial Bank of China Foshan Branch announced on August 9, 2005, that it would assume the guarantee liability from the defendant metal company again.	The lender has a relationship with the financial institution

measuring the similarity between sentences and elements, element recognition is really done from the semantic point of view, which effectively improves the accuracy of recognition.

In addition, why are the results in the area of marriage and family significantly better than in the area of labor disputes and lending disputes? As shown in Table 8, the similarity of key information and element labels of sentences is high in marital and family instruments. However, in the field of borrowing disputes, the model has difficulty in recognizing most domain-specific terms such as financial institutions and companies. In the field of employment contracts, more semantic understanding is needed to recognize legal elements. Therefore, the recognition results improved after semantic correlation analysis using cosine similarity, further validating the effectiveness of the method in this paper.

5. Conclusions

In this paper, a new method is proposed to solve the problem of judicial instrument element recognition. The method not only adds global semantic information to the sentence vector representation of the instrument using the BERT pre-training model. Moreover, the semantic information of element labels is utilized in the recognition of elements. On the one hand, the article makes full use of the semantic information of the text; on the other hand, it indirectly solves the problem of unbalanced data distribution. After obtaining global semantic information, we use cosine similarity to analyze the interrelationships between sentences and elements. Finally, we completed the element recognition task. The experimental results show that the BERT pretraining model gives the method good transferability, and the method in this paper has good results under various kinds of data of justice.

We hope that the ideas and methods presented will prove useful for the analysis of many other types of legal cases. Next, we take into account the relevance of the different elements. The result of element extraction will not only be related to the semantics of the sentence. We will determine the final extracted elements of the sentence by judging the correlation between the identified elements based on the probability of their simultaneous occurrence. Finally, the accuracy of recognition is improved.

Data Availability

The data included in this paper are available without any restriction.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Authors' Contributions

Dian Zhang and Hewei Zhang contributed to the work equally and should be regarded as co-first authors.

Acknowledgments

This work is supported by National Natural Science Foundation of China Grants no.11702289, Key core technology and generic technology research and development project of Shanxi Province, no. 2020XXX013, and National Key Research and Development Project.

References

- [1] R. Dale, "Law and word order: NLP in legal tech," *Natural Language Engineering*, vol. 25, no. 1, pp. 211–217, 2019.
- [2] C. J. Yuan, X. Y. Ding, J. Wei, and T. M. Zhang, "Application of artificial intelligence technology in the construction of intelligent courts," *AI-View*, vol. 4, pp. 56–65, 2020.
- [3] J. S. Dhani, R. Bhatt, B. Ganesan, P. Sirohi, and V. Bhatnagar, "Similar cases recommendation using legal knowledge graphs," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing ({EMNLP})*, pp. 1746–1751, Doha, Qatar, 2021.
- [4] S. Pudaruth, R. P. Gunputh, K. M. S. Sojjaudah, and P. Domun, "A question answer system for the Mauritian Judiciary," in *3rd International Conference on Soft Computing & Machine Intelligence (ISCFMI)*, pp. 201–205, Dubai, United Arab Emirates, 2016.
- [5] R. S. Wagh and D. Anand, "Legal document similarity: a multi-criteria decision-making perspective," *PeerJ Computer Science*, vol. 6, p. e262, 2020.
- [6] H. Zhong, C. Xiao, C. Tu, T. Zhang, Z. Liu, and M. Sun, "How does NLP benefit legal system: a summary of legal artificial intelligence," in *Proceedings of the 58th Annual Meeting of*

- the Association for Computational Linguistics*, pp. 5218–5230, Online, 2020.
- [7] L. Deng, C. Zhao, Z. Xu, and W. Zheng, “Critical point of jamming transition in two-dimensional monodisperse systems,” *The European Physical Journal E*, vol. 43, no. 12, pp. 1–5, 2020.
 - [8] Z. Xu, R. Wang, J. Cui, Y. Liu, and W. Zheng, “Effects of short-range attraction on jamming transition,” *Chinese Physics B*, vol. 30, no. 6, article 66101, 2021.
 - [9] F. Hou, Z. X. Cheng, L. Y. Kang, and W. Zheng, “Prediction of gestational diabetes based on lightgbm,” in *Paper presented at the Proceedings of the 2020 Conference on Artificial Intelligence and Healthcare*, pp. 161–165, New York, 2020.
 - [10] R. Wang, F. Fang, J. Cui, and W. Zheng, “Learning self-driven collective dynamics with graph networks,” *Scientific Reports*, vol. 12, no. 1, 2022.
 - [11] W. M. Shaban, A. H. Rabie, A. I. Saleh, and M. A. Abo-Elsoud, “A new COVID-19 patients detection strategy (CPDS) based on hybrid feature selection and enhanced KNN classifier!,” *Knowledge-Based Systems*, vol. 205, article 106270, 2020.
 - [12] H. F. Zhou, J. W. Zhang, Y. Q. Zhou, X. J. Guo, and Y. M. Ma, “A feature selection algorithm of decision tree based on feature weight,” *Expert Systems with Applications*, vol. 164, article 113842, 2021.
 - [13] G. Wu, R. Zheng, Y. Tian, and D. Liu, “Joint ranking SVM and binary relevance with robust low-rank learning for multi-label classification,” *Neural Networks*, vol. 122, pp. 24–39, 2020.
 - [14] S. Yildirim, D. Jothimani, C. Kavaklıoglu, and A. Başar, “Classification of “hot news” for financial forecast using NLP techniques,” in *2018 IEEE International Conference on Big Data (Big Data)*, pp. 4719–4722, Seattle, WA, USA, 2018.
 - [15] L. Chen, L. Qi, and F. Wang, “Comparison of feature-level learning methods for mining online consumer reviews,” *Expert Systems with Applications*, vol. 39, no. 10, pp. 9588–9601, 2012.
 - [16] Y. Kim, “Convolutional neural networks for sentence classification,” in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, pp. 1746–1751, Doha, Qatar, 2014.
 - [17] P. Liu, X. Qiu, and X. Huang, “Recurrent neural network for text classification with multi-task learning,” in *IJCAI’16: Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp. 2873–2879, New York, USA, 2016.
 - [18] M. E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U. R. Acharyad, “ABCDM: an attention-based bidirectional CNN-RNN deep model for sentiment analysis,” *Future Generation Computer Systems*, vol. 115, pp. 279–294, 2021.
 - [19] J. Yosinski, J. Clune, and Y. Bengio, “How transferable are features in deep neural networks?,” in *advances in Neural Information Processing Systems*, vol. 27, 2014.
 - [20] V. S. Bawa and V. Kumar, “Emotional sentiment analysis for a group of people based on transfer learning with a multi-modal system,” *Neural Computing and Applications*, vol. 31, no. 12, pp. 9061–9072, 2019.
 - [21] F. Zhuang, Z. Qi, K. Duan et al., “A comprehensive survey on transfer learning,” *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021.
 - [22] I. Ramirez, A. Cuesta-Infante, J. J. Pantrigo et al., “Convolutional neural networks for computer vision-based detection and recognition of dumpsters,” *Neural Computing and Applications*, vol. 32, no. 17, pp. 13203–13211, 2020.
 - [23] M. Peters, M. Neumann, M. Iyyer, M. Gardner, and L. Zettlemoyer, “Deep contextualized word representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 2227–2237, New Orleans, Louisiana, USA, 2018.
 - [24] R. Wang, J. Cui, Y. Zhang, and W. Zheng, “Research on collective movement prediction based on graph network,” *Journal of University of Electronic Science and Technology of China*, vol. 50, no. 5, pp. 768–773, 2021.
 - [25] Y. Zhang, J. Ren, R. Wang, F. Fang, and W. Zheng, “Multi-step sequence flood forecasting based on MSBP model,” *Water*, vol. 13, no. 15, p. 2095, 2021.
 - [26] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “Bert: pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1, pp. 4171–4186, Minneapolis, Minnesota, 2019.
 - [27] P. Shaw, J. Uszkoreit, and A. Vaswani, “Self-attention with relative position representations,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 2, New Orleans, Louisiana, USA, 2018.
 - [28] Z. Niu, G. Zhong, and H. Yu, “A review on the attention mechanism of deep learning,” *Neurocomputing*, vol. 452, pp. 48–62, 2021.
 - [29] F. Hou, Y. Zhang, X. Fu, L. Jiao, and W. Zheng, “The prediction of multistep traffic flow based on AST-GCN-LSTM,” *Journal of Advanced Transportation*, vol. 2021, 10 pages, 2021.