International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 3604-3613

https://doi.org/10.70517/ijhsa463304

Design of Intelligent Review Mechanism for Legal Compliance in Public Security Evidence Collection System Based on Artificial Intelligence Algorithm

Chenyue Hui[1,*]
[1] Shaanxi Police College, Xi'an, Shaanxi, 710021, China
Corresponding authors: (e-mail: 15509185235@163.com).

Abstract With the increasing complexity and refinement of public security forensics, the legal compliance review mechanism plays an important role in the process of public security forensics.In this paper, we propose a semantic role annotation and legal domain oriented entity-relationship extraction method based on BERT-BiLSTM-CRF. In legal text processing, this paper introduces the BERT model with powerful semantic understanding capability on the basis of BiLSTM-CRF model, which further enhances the semantic role annotation model's ability to understand the terminology of semantic structure of legal text. In addition, the models of legal information enhancement module, legal potential relationship and global correspondence model and decoder are constructed for entity relationship extraction in legal domain. The study shows that the semantic role labeling algorithm in this paper has different degrees of improvement in *F1*, *P* and *R* indicators, while the entity relationship extraction method has an extraction accuracy of more than 78% in multiple cycles, and the extraction accuracy is close to 100% on individual legal relationships. And the application of legal knowledge graph under the method of this paper in public security forensics provides rich legal entity relations for public security forensics, reduces the time of manual review and improves the reliability of the review results.

Index Terms BERT model, semantic role annotation, entity relationship extraction, legal information enhancement

I. Introduction

In recent years, the gradual emergence of self media and network media, everyone is a reporter, at any time in the interview of the national supervision environment is gradually established, the network of public opinion on the wrongful conviction of a wide range of reports, there are real, there are also fish in troubled waters, ill-intentioned. But this situation, but also from the side reflect the original criminal procedure process "heavy confession, light evidence" drawbacks [1]-[3]. With the reform of the judicial system and the trial of the substance of the comprehensive promotion of evidence in criminal proceedings in the pillar role is more and more prominent, is the foundation of the entire litigation activities, is the important means to exploit the cocoon, expose the truth of the case, but also the current embodiment of the justice of the powerful means [4], [5].

High-tech means of crime, to the public security organs of the detection of the case has brought a lot of difficulties, especially in the detection of cases involving the network, in accordance with the previous criminal procedure, as long as the suspect confessed to their own criminal behavior, the public security organs in the investigation of the veracity of the case can be transferred to prosecution, into the litigation process [6]-[8]. However, in the current procedure, the court needs the public security to provide detailed evidence of illegal and criminal behavior in the cross-examination link, so the electronic evidence becomes the most powerful means to be able to confirm the suspect's illegal behavior [9], [10]. At the same time, it is also the only means, the acquisition of electronic evidence or not, has become the key to determine the facts of cases of illegal and criminal offenses related to the Internet and to maintain the safety of people's property [11].

Whether the chain of evidence is complete, legal and effective often determines the success or failure of a case, but also determines the life of a suspect, the acquisition of traditional evidence often exists witnesses or suspects refused to cooperate with the public security organs, this time the investigation is in a deadlock [12], [13]. However, because e-discovery has strict security protection in the process of generation and circulation [14]. In addition to human destruction and tampering, electronic evidence is difficult to fade, more objective than traditional evidence, and more able to expose and restore the nature of the case [15]. At the same time, electronic evidence because of its electronic characteristics, once obtained that can be preserved for a long time and not easy to wear out, as long as in the hands of the judiciary, it is difficult to be misrepresented and misdirected [16], [17]. Even if part of the electronic data in the investigation process is destroyed, but always leave certain traces, through the investigators meticulous test, restore can still be used, maximize the guilty suspect to get a fair trial, so that the innocent people get justice [18].

In order to enhance the knowledge comprehension of legal compliance text, this paper implements semantic role annotation of legal compliance text based on deep learning model, which customizes legal semantic labels on one hand, and adopts BIO sequence annotation to annotate legal text on the other hand. Then a multi-level semantic role labeling model for legal compliance text containing BERT pre-training layer, BiLSTM layer and CRF layer is designed. In order to improve the accuracy of legal entity relationship extraction, a legal information enhancement module is designed, and a joint extraction method based on potential relationship and entity alignment is proposed to avoid the overlapping problem of legal neighborhood entity relationship extraction. The legal knowledge graph combining the methods of this paper is applied to the legal compliance review of public security forensic system, which opens up a new path for the application of legal neighborhood AI algorithms.

II. Legal Compliance Entity Relationship Extraction Based on Semantic Role Annotation

II. A. Deep Learning Based Text Semantic Role Annotation

The construction of the knowledge graph data layer requires the translation of the sentences in the legal normative text into the ternary form, and understanding the real semantics of the normative text sentences is the key to realizing the information extraction and converting them into the ternary form, so this paper adopts the natural language processing technology based on deep learning to firstly annotate the semantic roles of the legal normative text [19] in order to realize the shallow semantic analysis of the legal normative text. The data form of knowledge in the legal normative text is generally forms and text, and has the linguistic characteristics of no ambiguity and clear language, so it is not necessary to go through the step of disambiguation again in the process of information extraction and processing of normative text data. The deep learning-based semantic information annotation of legal normative text mainly includes data preprocessing, construction of annotated dataset, model training and learning, analysis and output results.

II. A. 1) Textual data annotation for legal compliance

Defining semantic role labels refers to labeling the semantic labels of words in a sentence of a legal normative text, which are used to indicate the semantic information of the words (since the subsequent aim is to parse them into triples, the semantics of entities, attributes and relations in the triples are indicated here). In order to improve the accuracy of the standard model of semantic roles, the semantic labels are categorized and generalized, and six semantic role labels are finally proposed as shown in Table 1.

Table 1: Semantic role tags and definitions

Tags	Define
JD	Review object
DS	Object attribute
SS	Data attribute
SXZ	Attribute value
BJ	Comparison word
LJ	Logical word

The problem of sequence labeling refers to labeling each element in a sequence. A sentence represents a sequence, and the words in the sentence are the elements in that sequence. Sequence labeling can be further divided into primitive and joint labeling. In primitive labeling, each element is labeled with a single label, while joint labeling refers to labeling each segment with the same label.BIO sequence labeling approach is a common joint labeling method, and is also a commonly used labeling strategy in deep learning tasks.

The BIO sequence annotation method converts the problem of joint labeling into the problem of original labeling. BIO consists of three tags: "B", "I", "O", where "B" is the start, "I" is the middle position, and "O" is not belonging to any tag type. In this way, the elements in the sequence are marked in the form of "B-XX", "I-XX", and "O".

The corpus of the semantic role annotation model of the legal normative text comes from the rules and provisions of various professional normative documents, and needs to be processed into a data format that can be processed by the Bert model, or in the form of "citizens have their own obligations." This sequence is processed and its format becomes "B-JD I-JD I-LJ O B-JD I-JD O I-JD I-JD O". It should be noted that punctuation marks are also elements in the sequence, and they need to be marked as "O" marks, and a blank line should be used at the end of each sentence before the next sequence is marked.

The canonical text is processed in the above format to construct the training set, development set and test set required for subsequent deep learning models. Among them, the training set is used to train the algorithm, and the development set is mainly used to realize the adjustment of parameters and the selection of features. The test set, on the other hand, is used to evaluate the trained optimal model and has no influence on the training of the algorithm as well as the parameters. In this study, the dataset of the modeling algorithm is divided according to the ratio of 8:1:1, and it is ensured that the development set and the test set obey the same distribution in order to obtain a faster iteration speed.

II. A. 2) Semantic Role Labeling Based on BERT-BiLSTM-CRF

Based on the dataset constructed in the previous section, this section will complete the annotation of semantic roles of legal specification texts based on deep learning models. At present, Bi-LSTM-CRF [20] model has achieved better results in sequence annotation tasks such as named entity recognition in the general domain. As for the domain of legal norms in this paper, due to the variety and specialization of named entities, it is difficult to achieve sufficiently high accuracy based on the Bi-LSTM-CRF model, a classical approach, to achieve the task of named entity recognition for legal norms. Compared with named entity recognition, semantic role annotation task uses fewer tags, which can achieve better recognition effect on smaller scale datasets and facilitate subsequent parsing into ternary data. Therefore, for the specific domain of legal normative texts, this paper transforms the legal normative knowledge graph data layer information extraction task into the semantic role annotation task for legal normative texts. The generalized semantic labels are defined, and the Bi-LSTM-CRF model is used to realize the semantic role annotation of legal texts. Meanwhile, considering the difficulty of acquiring word context information for this model, in order to obtain better semantic features, a BERT pre-training model is introduced, and finally a BERT-Bi-LSTM-CRF model is constructed to realize semantic role annotation.

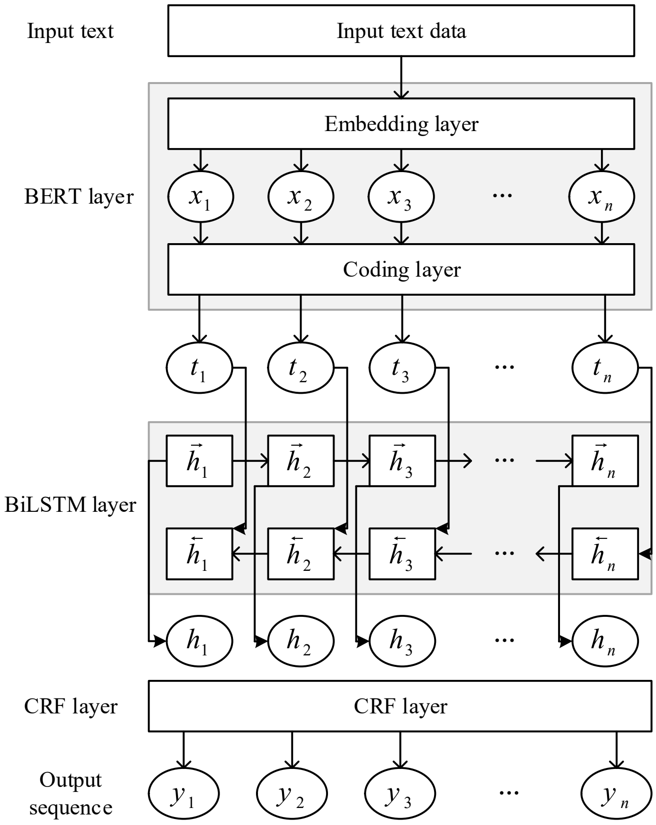

Figure 1: Structure diagram of the BERT-BILSTM-CRF model

The structure of BERT-BiLSTM-CRF model is shown in Fig. 1, the model is divided into three layers, the first is the BERT pre-training layer, the embedding layer of BERT converts the input text data into word vectors, which are superimposed by the word vector layer, the segmentation vector layer and the position vector layer to get the vector representation of the input text data $X = (x_1, x_2, ..., x_n)$, obtain the representation vectors of each character of the input text data by inputting the BERT coding layer, and learn the character context information, and output the coding sequence vector $T=(t_1,t_2,...,t_n)$.

The sequence vector T output by the BERT layer is again input to the BiLSTM layer for extracting semantic features from the input text data. The forward LSTM and backward LSTM are used to learn the semantic features of the text combined with the contextual information from left to right and right to left, respectively, to obtain the forward output sequence $\vec{h}_i$ and the reverse output sequence $\bar{h}_i$, which can then be spliced to obtain the final output sequence vector $hi=[\overrightarrow{hi},\overline{hi}]$. The output sequence hi indicates the probability value of each location being recognized as a marker for each BIO.

Once the sequence vector hi is obtained, it is then fed into the final CRF layer to obtain the label of the sentence. For example, the text before the text of the corresponding label "I-XX" corresponds to the "O" label, which is an incorrect labeling that does not conform to the law of the labeling strategy. Therefore, the CRF layer will adjust and optimize the sequence vector h to improve the logic and inaccuracy of the annotation, and finally output the maximum possible annotation sequence $Y=(y_1,y_2,...,y_n)$.

II. B.Entity-relationship extraction methods for legal domains

In this study, the legal information enhancement module is constructed. Meanwhile, in order to solve the problem of entity overlapping, a legal potential relationship and global correspondence (LPRGC) model is proposed for relationship extraction in legal domains, and the model mainly contains the Chinese legal BERT, legal information enhancement module and decoder.

II. B. 1) LegBERT-Chinese

BERT [21] is a deep neural network architecture that uses a multilayer bi-directional Transformer [22] encoder to learn contextual representations of words from large text corpora. While traditional unidirectional language models can only see the preceding words and are therefore limited by the preceding contextual information when predicting the following words, BERT uses a bi-directional Transformer model, which can take into account the contextual information at the same time, enabling it to perform better language understanding. The LegBERT-Chinese used in this study is a specific improved version of the BERT model, which is specifically trained or optimized for Chinese texts in the legal domain, which is used as the encoder of the model and the coding sequence of the model is obtained.

II. B. 2) Legal information enhancement layer

To enhance the domain knowledge of the model, this study designs a legal information enhancement layer to be more suitable for the task. Since the accuracy of recognizing entities is critical to the performance of ternary extraction, this study constructs a legal entity dictionary $Legal_{word}$ as legal features. First, a feature dictionary containing legal terms and entities that are used to augment the domain knowledge of the model is constructed. For the word sequence $S=(w_1,w_2,\cdots,w_i,\cdots,w_N)$, match S with $Legal_{word}$ to find all the subsequences that may form a noun expression in the dictionary, where w_t is the t th word in the input sequence. N is the length of the sequence. Define $S_{ij}=(w_i,w_{i+1},w_{i+2},\cdots,w_j)$. This study utilizes a mask matrix $M_D=(m_{ab})$ of size $N\times N$ to represent the legal features of the nouns, where D is the dimension of the matrix. m_{ab} is the expression for whether the subsequence S_{ij} is a legal noun:

$$m_{ab}=\begin{cases}0, & S_{ij}\in Legal_{word}\\ 1, & \text{other}\end{cases} \tag{1}$$

A legal domain-specific representation of the input sentence is computed using the Transformer coding layer. First, the input sentence H_L is linearly transformed to obtain the matrices Q_{Dh},K_{Dh} and V_{Dh}. Where L is the embedding vector or representation of the input sequence. h is the hth head of the multi-head attention. The sizes of Q_{Dh},K_{Dh} and V_{Dh} are $N\times d_q$, $N\times d_k$, and $N\times d_v$, respectively, where d_q,d_k and d_v are the dimensions of the query vector, key vector and value vector, respectively. The legal feature mask matrix is combined with the self-attention function to obtain a representation of the fused legal features. The available legal feature mask matrix M_D of the self-attention function is counted as:

$$A_{Dh} = softmax\left(M_D \times \frac{Q_{Dh} K_{Dh}}{\sqrt{d_k}} \right) V_{Dh} \tag{2}$$

where softmax is the normalized exponential function.

The results Q_{Dh}, K_{Dh} and V_{Dh} computed by each attention head are connected and the results are passed through the feedforward sublayer to finally output the features integrated with the lexicon, denoted as H_D. Finally, the representation H_L and the feature fusion representation H_D obtained from the coding layer are weighted and averaged to obtain a legal feature-enhanced representation as:

$$H_E = \gamma H_L + (1-\gamma) H_D \tag{3}$$

where γ is the weighting parameter.

II. B. 3) Decoders

The decoder includes a potential relation prediction component, a relation-specific sequence labeling component and an entity alignment component. For a given sentence consisting of n words encoded through the coding layer, the sequence can be represented as $h \in R^{n \times d}$. The set of possible relations P_{rel} in the sentence is computed using Eqs. (4) and (5) representations, and irrelevant relations are filtered to minimize the computation.

$$h_{avg} = p_{avg}(h) \in R^{d \times 1} \tag{4}$$

$$P_{rel} = \sigma(W_{rel} h_{avg} + b_{rel}) \tag{5}$$

where p_{avg} is the average pooling operation. $W_{rel} \in R^{1 \times d}$ is the trainable weights. σ is the sigmoid function. b_{rel} is the corresponding bias term.

The potential relation prediction component proposed in this study treats relation prediction in sentences as a multi-label binary classification task. Sequential labeling of specific relations is applied only to the predicted relations and not to all relations. This is because, for a given text, only some of the relations may exist and not all possible relations. Therefore, assigning labels only to potential relations that are predicted to exist as relations reduces the number of labels and thus improves classification efficiency.

After determining the potential relationships in a sentence, two sequence labeling operations are next performed to extract the subject and object separately. Separate extraction of subject and object is to deal with the problem of overlapping entities in triples. To ensure the simplicity of the model and computational speed, a fully connected neural network is chosen for sequence labeling, and two independent sequence labeling operations are performed for each relation to extract subject and object entities respectively, which are calculated as follows:

$$P_{i,j}^{sub} = softmax(W_{sub}(h_i + u_j) + b_{sub}) \tag{6}$$

$$P_{i,j}^{obj} = softmax(W_{obj}(h_i + u_j) + b_{obj}) \tag{7}$$

Among them, $P_{i,j}^{sub}$ is the probability distribution of the i token and the j relation is predicted as the subject, $P_{i,j}$ is the probability distribution of the i token and j relation is predicted as the object, and $h_i \in R^{1 \times d}$ is the encoded representation of the i token. $u_j \in R^{1 \times d}$ is the j th relation representation in the trainable embedding matrix $U \in R^{d \times n_{rel}}$, and n_{rel} is the size of the full set of relations. W_{sub} and $W_{obj} \in R^{d \times 3}$ are the trainable weight matrices, where 3 corresponds to the 3 labels of the tag set {B, I, O}. b_{sub} and b_{obj} are bias vectors.

After the sequence labeling, all possible subjects and objects about the sentence relation are obtained, and then a global correspondence matrix is used to determine the correct subject and object pairs.

Given a sentence with n tokens, its global correspondence matrix M has the shape $R^{n \times n}$ where n is the number of tokens in the sentence. M_{ij} is the score of the i th token as subject and the j th token as object. Each element of this matrix is the starting position of the subject and object with respect to a pairing and represents the confidence level of a subject-object pair; the higher the value, the higher the confidence that the subject-object pair belongs to a triad. After extracting all possible SUBJECTS and OBJECTS except those in the sentence for a particular type of relation, the global association matrix is used to determine the correct subject-object pair, calculated as:

$$P_{i\_sub,j\_obj}=\sigma(W_g\left[h_i^{sub};h_j^{obj}\right]+b_g) \tag{8}$$

where $h_i^{sub},h_j^{obj}\in R^{2d\times 1}$ is the vector representation of the i th and j th tokens in the input sentence through the coding layer, both of which form a potential subject and object pair. $W_g\in R^{2d\times 1}$ is a trainable weight. b_g is a trainable bias vector and g is a parameter of the global alignment matrix.

II. B. 4) Loss function

The model is trained using a joint, which optimally combines the objective function during training and shares the parameters of the encoder. The total model loss (L_{total}) consists of a loss L_{rel} for relation judgment, a loss L_{seq} for sequence labeling, and a loss $L_{global3}$ for entity alignment, all of which are cross-entropy loss functions, with the expression:

$$L_{total}=L_{rel}+L_{seq}+L_{global} \tag{9}$$

$$L_{rel}=-\frac{1}{n_{rel}}\sum_{i=1}^{n_{rel}}\left[y_i\ln P_{rel}+\left(1-y_i\right)\ln\left(1-P_{rel}\right)\right] \tag{10}$$

$$L_{seq}=-\frac{1}{2\times n\times n_r^{pot}}\sum_{t\in sub.obj}\sum_{j=1}^{n_r^{pot}}\sum_{i=1}^{n}y_{i,j}^t\ln P_{i,j}^t \tag{11}$$

$$L_{global}=-\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}(y_{i,j}\ln P_{i_{sub},j_{obj}}+(1-y_{i,j})\ln(1-P_{i\_sub,j\_obj})) \tag{12}$$

where n_{rel} is the size of the full set of relations. y_i is the actual probability of the i th class of relations. P_{rel} is the predicted probability of the i th class of relations. n_r^{pod} is the size of the subset of potential relations for the sentence. $y_{i,j}^t$ is the actual probability that a word from i to j is an entity t. $P_{i,j}^t$ is n the predicted probability that a word from i to j is an entity t. n is the size of the global matrix. $y_{i,j}$ denotes the actual probability that i is the subject and j is the object. P_{i_sub,j_obj} denotes the predicted probability of i being the subject and j being the object.

III. Experimental results and analysis

III. A. Data sets

In this chapter, the CONL-2009 dataset is used to train the semantic role annotation model, and the self-built complex entity annotation dataset is used to train the complex entity annotation model. The first focus is on the part about SRL in the CONL2009 dataset. For the CoNLL-2009 dataset, the data can be divided into a training set, a validation set and a test set, with the ratio of the training set to the validation set close to 10:1, where the training set is used for model training.

III. B. Experimental results of semantic role annotation

In this section, three evaluation metrics, *P*, *R*, and *F1*, are taken to measure the classification accuracy. In order to verify the effectiveness of the proposed improved algorithm for SRL task, this paper conducts multi-group comparison experiments on CoNLL-2009 dataset. The input data for the experiments in this section is the token group, i.e., the real text that has been sliced and diced, in the CoNLL-2009 dataset.

The control groups for the benchmark experiments are BERTGCN, RoBERTa_GCN that replaces the embedding layer encoding of BERTGCN with RoBERTa, Path_LSTM model, and GCNs_Denpendency model, respectively. The specific experimental results are shown in Table 2.

In comparison, compared with the benchmark model, the algorithm proposed in this paper not only achieves 3.1% and 4.5% improvement in *F1*, *P* respectively under the same experimental environment, but also has no inferior effect compared with the Path_LSTM and GCNs_Denpendency models that introduce syntactic dependency features.

Table 2: The semantic role annotation algorithm is compared with the whole

Model	CoNLL-2009		
	F1	*P*	*R*
BERT_GCN	83.0	83.1	83.8
RoBERTa_GCN	83.3	82.6	83.8
Path_LSTM	74.7	74.0	76.2
GCNs_Denpendency	83.7	83.1	84.8
Ours	86.3	86.8	86.3
Ascending/%	3.1	4.5	1.8

In order to verify the influence of the number of network iteration layers on the model output, this paper takes four layer settings, 2, 3, 4 and 5, as an example, to observe the process of model Loss curves under different iteration layers. The Loss curve during the training and validation process is shown in Fig. 2, and it can be seen that when the number of iterative layers is taken as 3, the Loss converges faster and has a smaller value.

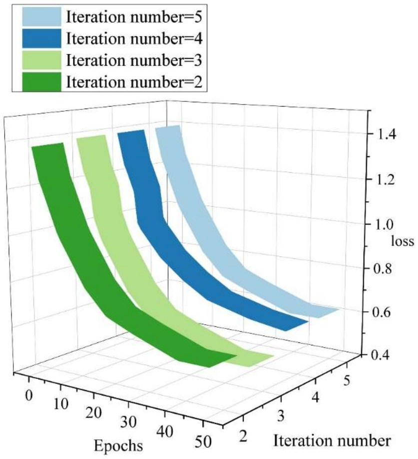

Figure 2: Loss curve

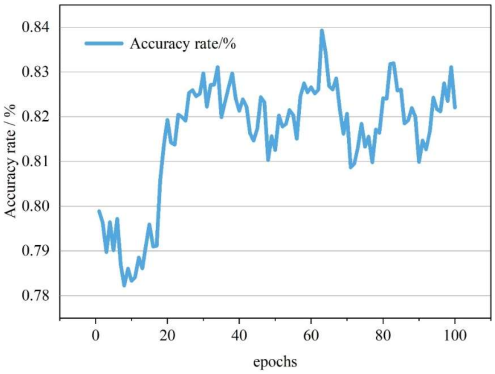

Figure 3: Accuracy of physical correlation

III. C. *Legal Compliance Entity Relationship Extraction Experiment Results*

The final output of this experiment is one of the eight relationships, so the experimental results have only accuracy, not recall. Because the experimental data is 14451 entries and the batch size batch_size is 100, 100 rounds of iterations are needed for each complete epoch for rounding purposes. Figure 3 shows the variation of relational extraction accuracy for 100 complete cycles.

As the number of epochs increases, the accuracy curve gradually changes from underfitting to overfitting, and the highest accuracy exists between epoch times of 60-70, and Fig. 4 shows the experimental results for the accuracy between epoch times of 60 and 70. The model has the highest accuracy when epoch=63.

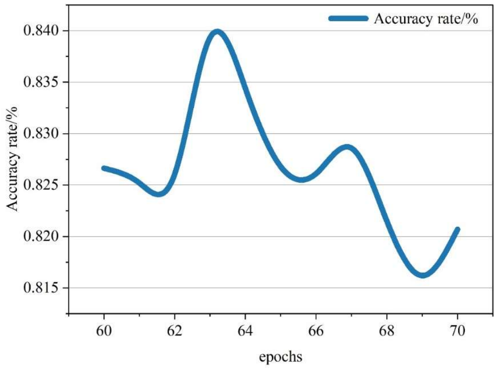

Figure 4: Accuracy of physical correlation

When epoch=63, the accuracy rates of different kinds of relationships are shown in Table 3. The crash and driving relationship has the highest accuracy rate, which is close to one hundred percent, probably because the characteristics of the relationship between the person's name and the vehicle information can be easily detected. Aggravated and casualty status had lower accuracy rates, and the lowest accuracy rate was for the same-defendant relationship, which was less than half of the accuracy rate. Through the observation of the data, two possible reasons were obtained: one is that the amount of data is too small, which leads to the relationship prediction as the relationship between two other names during the model training process, and the other is that there are some errors in the data of same-defendant relationship, for example, due to the inputting error of the court staff in a certain case, the victim is written as defendant so-and-so, or the two defendants are the same person, only that the inputting process will be the defendant LinMou2 into Lin, resulting in the same defendant being turned into two defendants when processing the data. Such input errors may lead to predicting same-defendant relationships as injury relationships during model training and testing.

Table 3: Accuracy of physical correlation

Relational type	Correct number	Accuracy rate
Casualties condition	454	76.05%
Hurt	488	81.68%
Co-defendant	92	56.48%
Same damage	531	89.46%
Bruise	599	99.67%
Driving	590	98.26%
Aggravated punishment	439	82.26%
Penalize	506	88.13%

Finally, the experimental results of the LegBERT-Chinese model are compared with the experimental results of the model without the BERT mechanism, and it is found that the best accuracy of the model in this paper is 84.00%, which is better than that of the model without the BERT mechanism, which is 82.55%.

IV. Compliance review applications based on legal knowledge mapping

IV. A. Content Knowledge Graph Schema Layer

When fusing different data sources to form a domain knowledge graph, the nodes in the graph mainly include: entities, i.e., some kind of things that can be distinguished and exist independently, such as a specific person, a city, a unit. Concepts, i.e., a collection of entities of a certain type, such as a computer, a country, a nation. Attribute, i.e., the entity points to the value of the attribute, such as the person (attribute) for Zhang San (attribute value). Relationships, i.e. functions that map different graph nodes to Boolean values. Character objects in the content of the legal text include both plaintiffs and defendants. Institutions include four categories: governmental institutions, corporate institutions, business institutions, and media institutions. Geography mainly involves domestic and foreign countries. Objects contain the subjects involved in the content, including people, objects, and things of 3 kinds. Theme is the subject that the content is divided into, which can include social, life, culture, politics, economy, science and technology categories. The legal events involved in the subject matter are reflective of the impact brought about by legal communication, including the 3 categories of public opinion, emergencies, and dissemination. Dissemination, on the other hand, indicates specific behaviors, such as legal or illegal comments, reposts, retweets, views, likes, and other behaviors involving the law.

IV. B. Content Compliance Identification

This paper will elaborate on how to use the knowledge graph for the compliance identification of legal content in the public security forensics system by constructing an example, and select one of the crimes of intentional dissemination of false information in the law as a rumor about food, entitled "Hot lemonade (water temperature below 60°C) will save you for a lifetime!" The specific content of the post reads, "Frozen lemonade contains only vitamin C, and when heated, lemonade turns into 'alkaline water', lemons have been shown to be able to remedy all types of cancer, and its killing effect on cancer cells is 10,000 times stronger than chemotherapy". The knowledge graph model is used to extract the knowledge of this piece of information, and the knowledge integration and review analysis are carried out through the legal compliance review graph system.

Hot lemon water for cancer knowledge mapping example shown in Figure 5, through the knowledge management framework model to build an integrated knowledge map of the legal knowledge base, and the use of data interfaces to connect with the knowledge base, by topic to extract the objective knowledge in the knowledge base, and with the required review of the knowledge of the automated knowledge matching, here it is found that alkaline water, a knowledge that has no relevance to the treatment method of cancer, it is considered that there is a mismatch of knowledge. When there is a mismatch in knowledge integration, the impact produced by the information can be judged, and further threshold judgments can be made on the legal compliance event attributes. Early warning rule setting is mainly based on the degree of inconsistency of knowledge matching, the degree of violation of the law, for the matching of inconsistent knowledge, if a certain impact is reached, it will give false information risk dissemination of illegal warning signals. The threshold value can be set according to the platform experience. The legal compliance knowledge graph construction engine can continuously extract social events from the information management systems of major public security organs and save them in the knowledge graph, so that the platform can carry out an intelligent review of the legal compliance content of the newly evidenced social events, and can give illegal or legal disposal opinions if the risk threshold is reached.

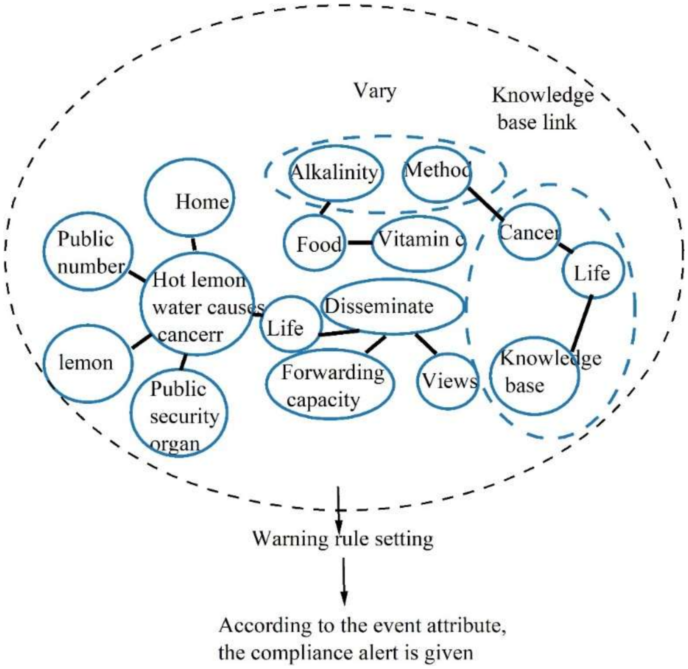

Figure 5: The knowledge map of hot lemon water cancer

V. Conclusion

A deep learning algorithm based semantic role annotation and legal neighborhood oriented entity relationship extraction method is proposed to be applied to the intelligent review of legal compliance in public security forensic system.

The semantic role labeling method for legal compliance in this paper performs better in terms of semantic understanding, and comparing with benchmark models such as BERT_GCN, the proposed semantic role labeling algorithm based on BERT-BiLSTM-CRF achieves higher than 3% improvement in F1,P, which contributes to the design of an intelligent review mechanism for legal compliance with reliable technical support.

The extraction accuracy of the entity relationship extraction method based on the legal latent relationship and global correspondence model is above 78% in all 100 complete cycles. When epoch=63, the accuracy of this paper's model is the highest, 84.00%, and the recognition accuracy on the legal relationship of abrasion reaches 99.67%, which is extremely accurate.

The application in the public security evidence collection system shows that the intelligent review mechanism based on legal knowledge mapping in this paper can reasonably follow the procedure to gradually identify the legal compliance issues, and provide scientific and objective review recommendations for law enforcement officers, the design of the method in this paper provides technical support for public security evidence collection, reduces the work pressure of the public security evidence collection staff, and has a valuable application value.

References

[1] Riekkinen, J. (2019). Electronic Evidence in Criminal Procedure: On the Effects of ICT and the Development towards the Network Society on the Life-cycle of Evidence. Digital Evidence & Elec. Signature L. Rev., 16, 6.

[2] Simonato, M. (2014). Defence rights and the use of information technology in criminal procedure. Revue internationale de droit pénal, 85(1), 261-310.

[3] Dmitrieva, A. A., & Pastukhov, P. S. (2023). Concept of Electronic Evidence in Criminal Legal Procedure. Journal of Digital Technologies and Law, 1(1).

[4] Lapshin, V. F., Korneev, S. A., & Kilimbaev, R. V. (2020, December). The use of artificial intelligence in criminal law and criminal procedure systems. In IOP Conference Series: Materials Science and Engineering (Vol. 1001, No. 1, p. 012144). IOP Publishing.

[5] Zgoliński, I. (2025). Technology and Criminal Proceedings: An attempt to systematize the issue and determine the main directions of implementation. Europejski Przegląd Prawa i Stosunków Międzynarodowych, (1/2025/73), 99-113.

[6] Ulrich, S. I. E. B. E. R. (2016). The paradigm shift in the global risk society: From criminal law to global security law–an analysis of the changing limits of crime control. Journal of Eastern European Criminal Law, (01), 14-27.

[7] Dzhanadilov, O. M., & Azhibayev, M. G. (2019). Problems of countering criminal offenses in information and communication networks. Journal of Advanced Research in Law and Economics, 10(1 (39)), 134-143.

[8] Faqir, R. S., Sharari, S., & Salameh, S. A. (2014). Cyber Crimes and Technical Issues under the Jordanian Information System Crimes Law. J. Pol. & L., 7, 94.

[9] Tatale, S., & Bhirud, N. (2016). Criminal data analysis in a crime investigation system using data mining. J. Data Mining Manag., 1(1), 1-13.

[10] Dzhansarayeva, R. E., Bissengali, L., Bazilova, A. A., Akbolatova, M. E., & Bissenova, M. K. (2013). Problems of formation of the concept of criminal policy of state in the theory of criminal law. Middle-East Journal of Scientific Research, 14(4), 508515.

[11] Depauw, S. (2018). Electronic evidence in criminal matters: How about e-evidence instruments 2.0?. Eur. Crim. L. Rev., 8, 62.

[12] Kasper, A., & Laurits, E. (2016). Challenges in collecting digital evidence: a legal perspective. The future of law and eTechnologies, 195-233.

[13] Murzo, Y., & Halchenko, V. (2023). Electronic evidence as a means of proof during the pillage investigation. Scientific Journal of the National Academy of Internal Affairs, 3(28), 48-57.

[14] Losavio, M. M., Pastukov, P., Polyakova, S., Zhang, X., Chow, K. P., Koltay, A., ... & Ortiz, M. E. (2019). The juridical spheres for digital forensics and electronic evidence in the insecure electronic world. Wiley Interdisciplinary Reviews: Forensic Science, 1(5), e1337.

[15] Yeboah-Ofori, A., & Brown, A. D. (2020). Digital forensics investigation jurisprudence: issues of admissibility of digital evidence. Journal of Forensic, Legal & Investigative Sciences, 6(1), 1-8.

[16] Brown, C. S. (2015). Investigating and prosecuting cyber crime: Forensic dependencies and barriers to justice. International Journal of Cyber Criminology, 9(1), 55.

[17] Zhou, S., & Dai, L. (2023). Similar Cases Retrieval for Illegal Electronic Data in Criminal Proceedings and Suggestions for Improving. Chinese Studies, 12(1), 80-90.

[18] Casey, E. (2011). Digital evidence in the courtroom. Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet,, 49-84.

[19] Onan Aytuğ. (2023). SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization. Journal of King Saud University - Computer and Information Sciences,35(7),

[20] Asma Mekki,Inès Zribi,Mariem Ellouze & Lamia Hadrich Belguith. (2024). Named Entity Recognition of Tunisian Arabic Using the Bi-LSTM-CRF Model. INTERNATIONAL JOURNAL ON ARTIFICIAL INTELLIGENCE TOOLS,33(02),

[21] Zizhao Zhang,Xinyue Yang,Liping Sun,Yu Sun & Jichuan Kang. (2025). Research on constructing and reasoning the collision knowledge graph of autonomous navigation ship based on enhanced BERT model. Expert Systems With Applications,278,127429-127429.

[22] Wenbin Tan,Li Zhang,Yiwang Huang,Kaibei Peng & Yanyun Qu. (2025). M2DETR: A Multi-band Multi-scale Detection Transformer for Pest Detection. Computers and Electronics in Agriculture,235,110325-110325.