A Legal Judgment Prediction Method Based on Knowledge-integrated Event Evolutionary Graphs and Causal Inference

Chunling Dong†, Mingsen Li, Xinyi Liu

School of Computer and Cyber Sciences, Communication University of China, Beijing 100024, China

Keywords: Causal event evolutionary graph; Legal judgment prediction; Causal inference; Multi-path root cause inference; Legal intelligence

Citation: Dong C.L., Li M.S., Liu X.Y.: A Legal Judgment Prediction Method Based on Knowledge-integrated Event Evolutionary Graphs and Causal Inference. Data Intelligence, Vol. XX, Art. No.: 2025??XX, pp. 1–27, 2025. DOI: https://doi.org/10.3724/2096-7004.di.2025.0092

ABSTRACT

Existing methods for predicting legal judgments typically focus on the analysis of contextual relationship of word level in the legal case descriptions, rather than the global causality at event level. The approaches are susceptible to interference from irrelevant information in lengthy texts, thereby reducing the accuracy of similar case retrieval. Moreover, the complex causal relationships in legal documents are rarely paid attention to and utilized, which diminishes the interpretability of the judgment outcomes. To address these issues, this paper proposes a similar case retrieval algorithm by the fusion of textual features and graph structural features derived from the causal event evolutionary graph (CEEG), in which the EventBERT and relational graph convolutional network (R-GCN) model are developed to improve the accuracy and completeness of similar case retrieval. Therefore, a multi-path root cause inference (MRCI) algorithm is proposed for charge prediction, by means of uncertain causality representation and causal inference on the legal knowledge-integrated CEEG for cases. The legal judgment prediction method is thus proposed by integrating the algorithms of similar case retrieval and charge prediction. Experimental comparative analysis reveals that the proposed method intuitively represents the internal logical and causal relationships in the legal cases, and significantly improves both the accuracy and interpretability of legal judgment prediction.

† Corresponding author: Chunling Dong (E-mail: dongcl15@tsinghua.org.cn).

© 2025 Chinese Academy of Sciences. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

1. INTRODUCTION

In recent years, researches on judicial prediction tasks have mostly been conducted on the basis of pre-trained language models fine-tuned by specific issues in the legal field. However, large models often exhibit weak capabilities in structured representation of legal domain knowledge and internal logical causality, which hinders the learning and reasoning on fundamental principles of judicial prediction [1] and diminishes the interpretability of predicted results. Therefore, relying solely on big data and large models still has certain limitations in resolving this task. To address the challenges, the methods based on knowledge graph and causal inference methods which offer structured knowledge representation, enhanced interpretability and improved computational accuracy, present a promising solution. In particular, causal event graph can effectively describe and quantify the global causalities among legal entities, establishing an efficient and interpretable basis for predictive reasoning.

Therefore, this paper proposes a novel approach to legal judgement prediction that integrates causal inference with similar case retrieval methods. Based on the construction of knowledge-integrated causal event evolutionary graph (K-CEEG) for legal cases, a fusion mechanism of textual and graph structural information (FTGS) and a multi-path root cause inference (MRCI) algorithm are proposed. The solutions are intended to facilitate more accurate and interpretable legal judgment reasoning, by establishing a consistent framework for structured representation of legal knowledge and inherent causalities underlying legal cases.

The remainder of this paper is structured as follows: Section 2 briefly reviews the method of case retrieval and legal judgement prediction, Section 3 describes the detailed reasoning process of our method, Section 4 provides a detailed presentation of the experimental results and analysis, and Section 5 concludes the paper.

2. RELATED WORK

In judicial practice, similar cases are important basis for judgments, which helps to maintain the stability of the law and judicial fairness, and improve judicial efficiency. Similar case retrieval refers to the analysis of the current case by referring to the judgment results or reasons of previous cases that have similarities in facts, legal relationships, etc., in order to guide the judgment or ruling of the current case.

Significant progress has been made in the research of implementing similar case retrieval methods in recent years. Shao et al. [2] utilized the BERT-PLI model to capture semantic relationships at the paragraph level and aggregated paragraph information interaction for similar case retrieval in order to address the issues related to the unique characteristics of legal text. Xiao et al. [3] further proposed Lawformer, a pre-trained language model specifically tailored for the Chinese legal domain to enhance the model's ability to deal with lengthy legal texts, and adapted it for similar case retrieval tasks. Li et al. [4] developed SAILER, a novel structure-aware pre-trained language model designed for legal case retrieval to leverage the structural information of legal documents to improve retrieval accuracy. However, relying solely

on language models often fails to capture the complete structural information within legal documents. For that, Zhang [5] and Tang et al. [6] proposed the CaseGNN, a similar case retrieval model based on graph neural networks, which utilizes the structural information of legal documents and overcomes the limitations of language models regarding text length. Gao et al. [7] further improved this approach by replacing graph neural networks with graph matching networks and proposed a Legal-GMN model, which improves the interpretability of case matching predictions. The EMV method proposed by Zhang et al. [8] incorporates event knowledge and leverages a heterogeneous graph structure for legal retrieval, further demonstrating the critical role of structural and knowledge enhancement in improving retrieval performance in the legal domain.

Liu et al. [9] proposed a predictive model that integrates BERT with LSTM-CNN and is capable of handling multi-label features based on the characteristics of legal texts. However, this model focuses solely on extracting pure legal information and neglects the integration of legal background knowledge. Luo [10] and Zhang et al. [11] introduced the attention mechanisms and highlighted the significance of legal knowledge, by demonstrating the correlation between case descriptions and legal statutes. Nonetheless, these approaches encounter difficulties in differentiating complex cases and identifying low-frequency charges, particularly in the context of limited data. Hu et al. [12] achieved accurate predictions for low-frequency charges under small sample conditions by distinguishing attributes. He et al. [13] addressed the imbalance in charges by constructing a sequence-enhanced capsule model incorporating an attention mechanism. While these studies solely utilize language models for single-layer text analysis, they do not incorporate knowledge graphs. Zhao et al. [14] improved the effectiveness of charge prediction tasks by incorporating relevant legal statutes as external knowledge and integrating this information through graph fusion. While this approach enhances task performance, it does not explore the causal relationships between case developments and judgment outcomes. Zhou et al. [15] proposed the LK-IB framework by utilizing legal knowledge and combining an interpretable model and a black-box model to balance the interpretability and prediction performance. Although it performs well on simple cases, further investigation is required to address more complex scenarios. Recent advances have emphasized the importance of modeling label dependencies and task-level consistency, as exemplified by GJudge [16], which constructs legal label graphs and introduces graph-based consistency constraints to enhance multi-task performance. To better capture the structural correlations between legal facts and statutes, LF-HetG [17] further proposed a heterogeneous graph framework that integrates case facts, keywords, and law articles, enabling more accurate representation and disambiguation of similar provisions. To further enhance legal text modeling, Meng et al. [18] incorporated domain knowledge into large language models via deontic prompt learning and multi-granularity graph fusion, significantly improving legal text understanding and law article prediction.

Similar case retrieval can provide decision-making basis for judgment prediction, therefore, there is an inherent correlation between the two tasks rather than independence. However, the aforementioned studies mostly focus on improving the accuracy of individual tasks and overlook the intrinsic correlation. To bridge this gap, Guo et al. [19] developed a novel judgment prediction approach, TenLa, which introduces case similarity as a key factor by utilizing a controllable tensor decomposition algorithm

combined with an optimized Lasso regression model. A CaseLink [20] model based on inductive graph learning is proposed to utilise the intrinsic case connectivity for legal case retrieval, a novel Global Case Graph is incorporated to represent both the case semantic relationship and the case legal charge relationship. Similarly, Zhou et al. [21] introduced a similar case-based prison term prediction method that integrates case similarity into the judgment prediction process, enabling more accurate sentencing outcomes based on similar cases. Li et al. [22] differentiated the value of different similar cases by designing a legal judgment prediction model called MAtching GuIding Cases (MAGIC), which enhances the weight of typical and guiding cases in the prediction process. Yao et al. [23] proposed a novel Gated Hierarchical Multi-Task Learning Network, GHE-DAP, to tackle multiple subtasks of Judicial Decision Prediction simultaneously by extracting deep semantic information from fact descriptions from various perspectives. Further, Yao et al. [24] proposed a novel multi-task learning network, CSDNet, to jointly tackle multiple subtasks incorporating the properties including commonalities, specificities, and dependencies. A denoising module is developed to minimize noise interference between sub-tasks and enhance prediction accuracy. However, these models overlook the intrinsic structural information within similar cases. Dong et al. [25] proposed a framework named Graph Contrastive Learning with Augmentation (GCLA), which conducts fine-grained structural processing of similar cases by constructing trainable document-level graphs for fact descriptions and capturing local and global contexts through sentence-level subgraphs. However, this approach does not address the modeling and processing of complex causal relationships. To improve the interpretability of judgment outcomes, Wang et al. [26] developed a reinforcement learning principal representation model by integrating an attention mechanism for regression prediction (PRRP), which captures distinguishing features of cases from multiple perspectives and achieves a more interpretable sentencing prediction through feature-enhanced fact representation.

Despite rapid research progress in the field, existing methods for predicting judgments still face some challenges. Firstly, case descriptions often contain a significant amount of information irrelevant to the judgment. Although attention mechanisms can highlight relevant parts and diminish the importance of irrelevant words, they cannot completely discard unrelated information. Even if each unrelated word has a low weight, their cumulative weight can significantly impact the semantic representation of the case description [27]. Secondly, these models rarely focus on the logical reasoning and causal relationships embedded in the case details, making it difficult to accurately infer complex cases through causal chain analysis. Thirdly, while some methods have improved model interpretability, they lack in-depth causal analysis and a transparent reasoning process, which undermines the consistency of the prediction results and their explanations.

To address the above issues, this paper proposes a legal judgment prediction method based on the knowledge-integrated causal event evolution graph (K-CEEG) based on Dynamic Uncertain Causality Graph (DUCG) [28-30], which can compactly represent and precisely reason about uncertain causalities. The uncertain causality representation mechanism is introduced in the construction of K-CEEG to enhance the interpretability of the model. Moreover, a novel similar case retrieval algorithm FTGS and a charge prediction algorithm MRCI are proposed by means of uncertain and causal inference schemes on the K-CEEG.

3. METHODOLOGY

This paper presents a legal judgement prediction method based on knowledge-integrated causal event evolutionary graphs. As illustrated in Figure 1, this method is mainly composed of four parts: (1) The construction of K-CEEG; (2) FTGS similar case retrieval algorithm; (3) MRCI charge prediction algorithm; and (4) Causal inference-based sentencing prediction algorithm.

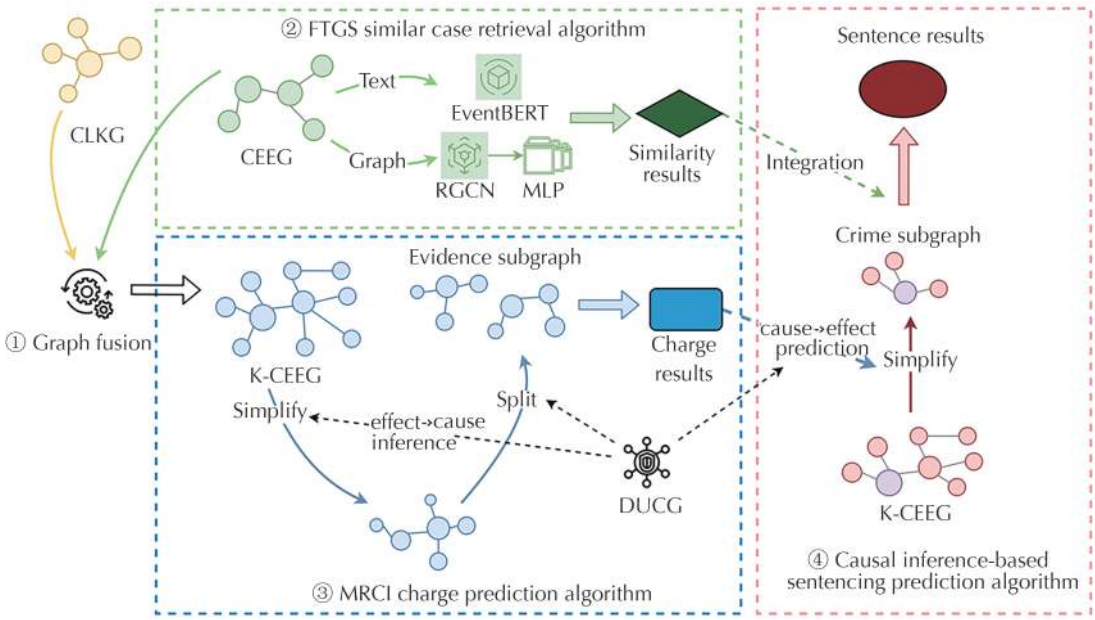


Figure 1. The framework of judgment prediction method.

3.1 The construction method of K-CEEG

The construction of K-CEEG consists of three parts: the construction of CEEG and CLKG, and the graph fusion and simplification. The construction process is illustrated in Figure 2.

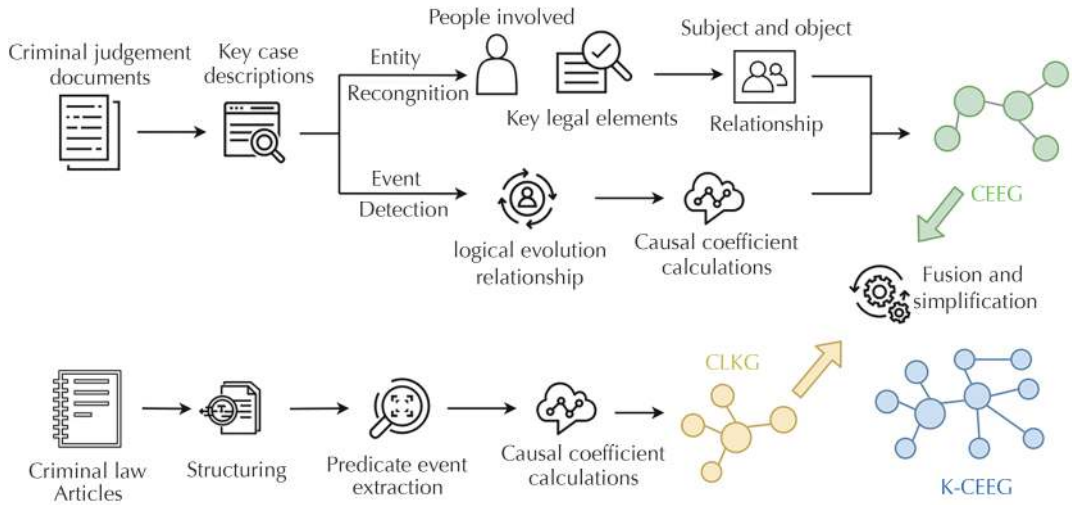


Figure 2. Construction process of K-CEEG.

3.1.1 The construction of CEEG

CEEG is proposed to represent the large-scale criminal judgment documents. The process of constructing a CEEG begins with data preprocessing, aiming at extracting key case descriptions from the judgment documents. Next, based on ontology layer definitions, legal entity recognition is accomplished through methods such as constructing regular expressions, matching keyword dictionaries, and analyzing syntactic structures. Concurrently, the OmniEvent toolkit is utilized for event detection tasks. Finally, the PMI score is combined with attention mechanisms to calculate causal coefficients, enabling the extraction of logical evolutionary relationships and completing the construction of the CEEG.

(1) CEEG ontology layer design

The ontology layer design serves as the backbone and foundation of the graph construction task, abstracting the content in the domain texts. To delineate the domain of formal case details while refining each field and describing the logical relationships between the concepts behind the cases, additional definitions for concept types and relationship types are provided.

Definition of CEEG Concept Types: To comprehensively capture the critical information involved in a case and achieve effective structured representation of legal documents, three types of concepts are defined in CEEG: 1) the individuals including defendants, victims, and other related individuals appearing in the case; 2) the abstract events involved in the case, and six types of events and behaviors exist in a specific case according to the Chinese legal event detection data set LEVEN [31], including general behaviors, prohibited behaviors, judicial-related behaviors, harmful consequences, accidents, and natural disasters (totally 300 abstract events can be defined); 3) key legal elements, including the aggravating factors, mitigating factors, amounts involved, and injury descriptions. Some of the key elements that are

difficult to identify through event detection but will have an impact on the verdict. Figure 3 illustrates the definition of CEEG concept types.

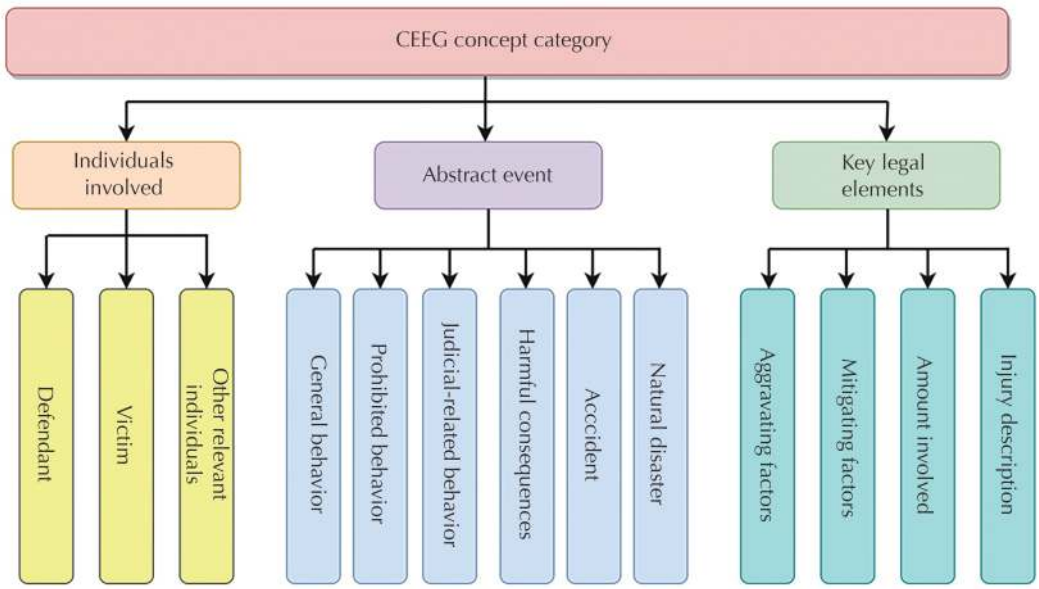


Figure 3. Definition of CEEG concept types.

Definition of relationship types: Logical evolution relationships primarily include causality and sequential relationships. Sequential relationships can be viewed as a generalized causality with weaker causal strength. In the adjudication process of judicial cases, the causal connection between the behavior of the perpetrator and the resulting harm is particularly emphasized, as it is a crucial basis for determining legal responsibility and directly decides whether the perpetrator should be held accountable for a specific outcome.

Subject-object relationships focus on the connections between involved individuals, abstract events, and key legal elements. This relationship is simplified by analyzing the relative positions of entities within the text. Specifically, the judgment documents are segmented at the sentence level. If the role of an involved person appears before the current focal event or legal element, the relationship between them is marked as a subject relationship. Conversely, if the person's role appears after the event or legal element, the relationship is defined as an object relationship. This differentiation helps distinguish the behaviors of different individuals, enabling effective extraction of the connections between involved individuals and the events related to their actions in legal documents, thereby facilitating a better understanding of case development.

(2) Logical relationship evolution and causality coefficient calculation

Due to the unique stylistic nature of judicial documents, extracting causal relationships solely based on explicit causal trigger words is incomplete. Therefore, this paper proposes a causality coefficient calculation method that combines Pointwise Mutual Information (PMI) values with a self-attention mechanism. By setting a threshold, it distinguishes between causality and sequential relationships. PMI is a metric that measures the co-occurrence frequency of two events related to the occurrence frequencies, as defined in Eq. (1).

$$\mathrm{PMI}(x,\ y) = \log_2\left(\frac{\mathrm{P}(x,\ y)}{\mathrm{P}(x)\mathrm{P}(y)}\right) \tag{1}$$

In which $\mathrm{P}(x,\ y)$ represents the joint probability of events x and y, while $\mathrm{P}(x)$ and $\mathrm{P}(y)$ represent the individual probabilities of events x and y. The PMI value is calculated by the normalization process defined in Eq. (2).

$$\mathrm{normPMI}(x,\ y) = \frac{\mathrm{PMI}(x,\ y) - \mathrm{minPMI}}{\mathrm{max\,PMI} - \mathrm{minPMI}} \tag{2}$$

Attention weights are calculated using a self-attention mechanism combined with the softmax function. The self-attention mechanism captures the strength of associations between different events, and the softmax function converts these association strengths into a probability distribution. The specific process is: 1) input the node features $\mathrm{h} = \left\{\vec{h}_1,\ \vec{h}_2,\ \ldots,\ \vec{h}_N\right\}$, where N is the total number of nodes in the graph, and 2) train a weight matrix for each node, and perform the self-attention mechanism calculation.

If there are explicit causal trigger words between two events, the weight between these two arcs should be directly set to 1 to reflect the explicitly stated causal relationships in the texts. Finally, the causality coefficient is obtained by integrating the PMI values and attention weights, which represents the strength of causal relationship between events, as defined in Eq. (3).

$$\mathrm{CausalCoefficient}(x,\ y) = \frac{\mathrm{normPMI}(x,\ y) + \mathrm{AttentionWeight}(x,\ y)}{2} \tag{3}$$

(3) The example of CEEG

This section selects the case description in "*Chen committed the crime of intentional injury*" and uses this as an illustrative example of the CEEG construction, which is shown in Figure 4.

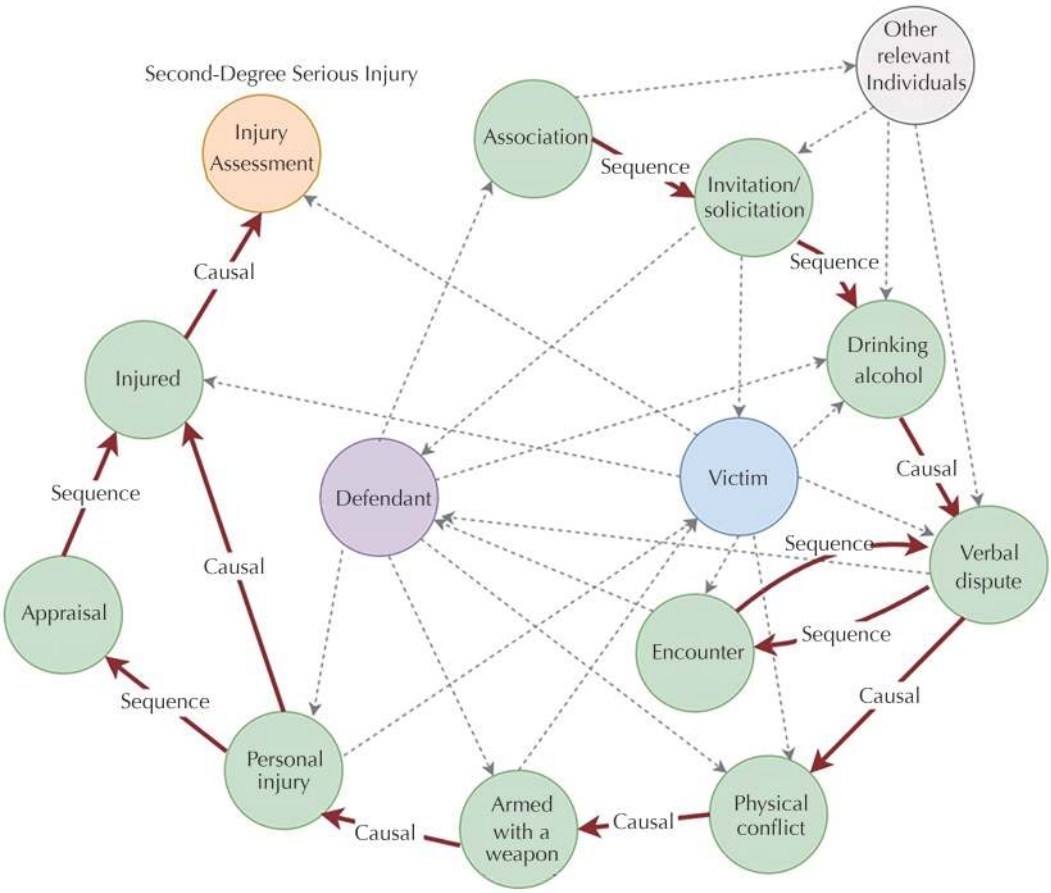


Figure 4. An illustrative example of the construction of CEEG.

3.1.2 The construction of CLKG

CLKG is constructed based on the articles of the Criminal Law of the People's Republic of China, as follows.

(1) Approach of extraction of criminal acts based on syntactic structure analysis

The behavior descriptions within the articles of the law have a decisive impact on the determination of the charges and sentencing. These descriptions are usually detailed and complex, and directly represent them as nodes in the knowledge graph would hinder subsequent processes of graph alignment and integration. Therefore, a concise and effective approach is proposed for identifying and extracting the involved behaviors. Firstly, the Baidu LAC tool is used to segment the content of the charges and label parts of speech. The syntactic analysis focuses on three types of words with high prevalence and significant importance: common verbs, nominal verbs, and adverbs. This approach effectively extracts key behaviors

and elements from the descriptions of charges. Further, considering that the CEEG focuses on abstract events, only verb-related words are retained as entities under the charge types in the CLKG. This effectively establishes the connection between the two types of graphs, better aligning the behavior descriptions in the charges with the abstract events in the CEEG.

(2) Injection of causal relationships and weight calculation

The constructed knowledge graph focuses on three basic elements of the Criminal Law: charges, descriptions of the criminal acts, and statutory penalties, which are interconnected by intrinsic causal logical relationships. Charges serve as the qualitative characterization of criminal behaviors, providing a high-level legal abstraction. In contrast, the descriptions of the criminal acts and statutory penalties provide the concrete and quantitative expressions of this abstraction. The detailed elements within the descriptions of the criminal acts directly impact the determination of statutory penalties. Statutory penalties, as the legal consequences of charges, are tightly linked to both charges and descriptions of the criminal acts. Therefore, causality is introduced to define the relationships among entities in the CLKG, and the relationship weights are defined as causal coefficients.

The calculation of causal coefficients in the knowledge graph integrates a statistics-based method and an expert knowledge fine-tuning method. As shown in Eq. (4), for each pair of entities in the graph, we first count the number of all downstream entities directly connected to the head entity in the current triplet and take the reciprocal as the initial weight coefficient of the relationship. Then, we count the number of occurrences of the current triplet and multiply it by the initial weight coefficient to obtain the final probability value.

$$\text{Weight}(h,\ t) = \left(\frac{1}{\text{N}(h)}\right) \times \left(\text{C}(h,\ t)\right) \tag{4}$$

In which, N(h) represents the number of all downstream entities directly connected to the head entity h. C(h, t) denotes the number of occurrences of the current head entity h and the tail entity t. This approach allows for the preliminary quantification of the strength of causal relationships between each pair of entities.

However, the statistical method alone cannot fully account for the specificity and complexity of the legal domain. Therefore, it is necessary to incorporate expert knowledge for fine-tuning, which can effectively address the limitations of a single model in dealing with diverse real-world problems. To ensure the rationality of the CLKG construction and the setting of causal coefficients, domain experts can be invited to evaluate and revise the relationships within the knowledge graph as Eq. (5). CausalCofficient (h, t) is the final causal coefficient of this edge, δ is the adjustment amount of the weight made by the domain expert.

$$\text{CausalCofficient}(h,\ t) = \text{Weight}(h,\ t) + \delta \tag{5}$$

(3) The example of CLKG

Figure 5 illustrates an example of CLKG construction based on four types of criminal charges under the Criminal Law: murder, negligent homicide, negligent causing of serious injury, and intentional injury.

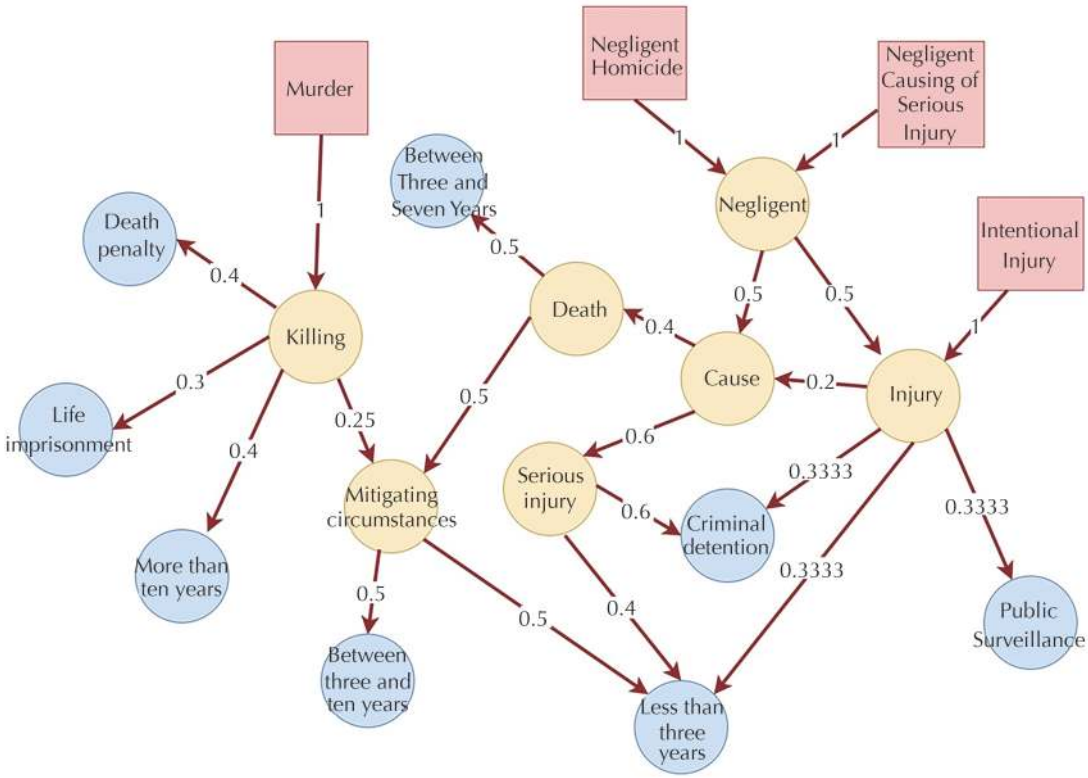


Figure 5. An illustrative example of the CLKG construction.

3.1.3 The graph fusion algorithm of CEEG and CLKG

The above-mentioned two types of graphs are integrated by means of similar node fusion, in order to generate the K-CEEG. Nodes that coexist in the two graphs primarily include the verb words related to the charge type in the knowledge graph and the abstract events in the CEEG for the case evolutionary process. By comparing the verb-related words with the corresponding trigger words of the abstract events in the text, the identical shared nodes from two graphs are merged.

However, just merging the identical shared nodes in two graphs is not enough. A text similarity comparison algorithm for the nodes is further proposed to further enhance the effectiveness of the graph fusion by three steps: 1) word vector acquisition, 2) similarity calculation, and 3) node merging. Firstly, the BERT-Base-Chinese model is utilized to obtain the word embeddings for the nodes in the graph. Secondly,

cosine similarity is employed to evaluate the similarity between word vectors, as shown in Eq. (6). Finally, based on the similarity calculation results, a threshold τ is set to decide whether to merge two nodes or not. The experimental results reveal that the threshold τ can be set to 0.8 in practical applications to ensure correct merging of similar nodes. An example of K-CEEG construction is illustrated in Figure 6.

$$Similarity = \cos(\theta) = \frac{A \cdot B}{|A||B|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}} \tag{6}$$

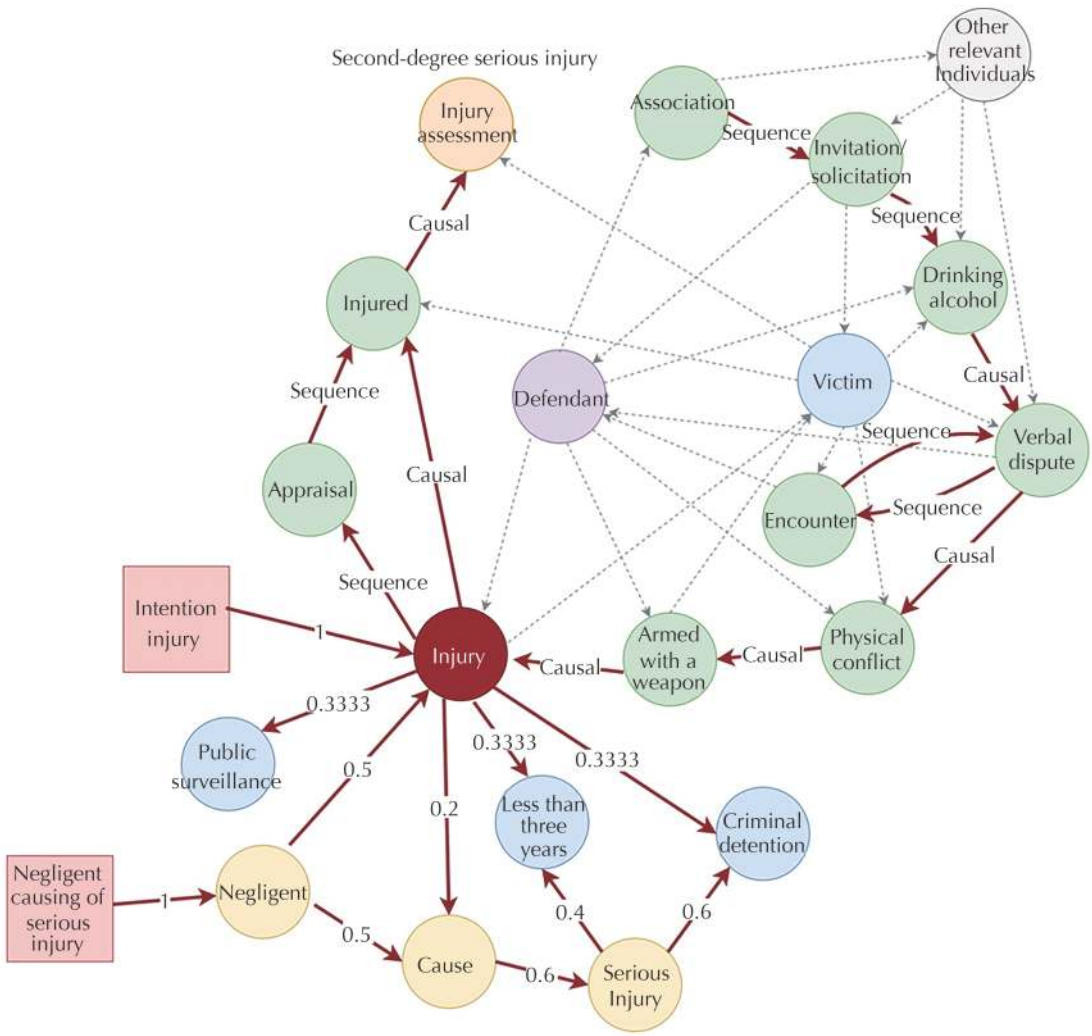


Figure 6. An illustrative example of the K-CEEG construction.

3.2 Similar case retrieval algorithm FTGS

A similar case retrieval algorithm FTGS is proposed, as shown in Figure 7. Initially, summaries of case descriptions are derived from the triplets in the CEEG. The EventBERT model is then applied to incorporate event information from the graph, enabling the comparison of textual similarities. Simultaneously, the R-GCN model, coupled with an MLP, is employed to obtain the similarity between the graph structures by analyzing the structural relationships. Finally, the outcomes are integrated to generate a similarity ranking for candidate cases.

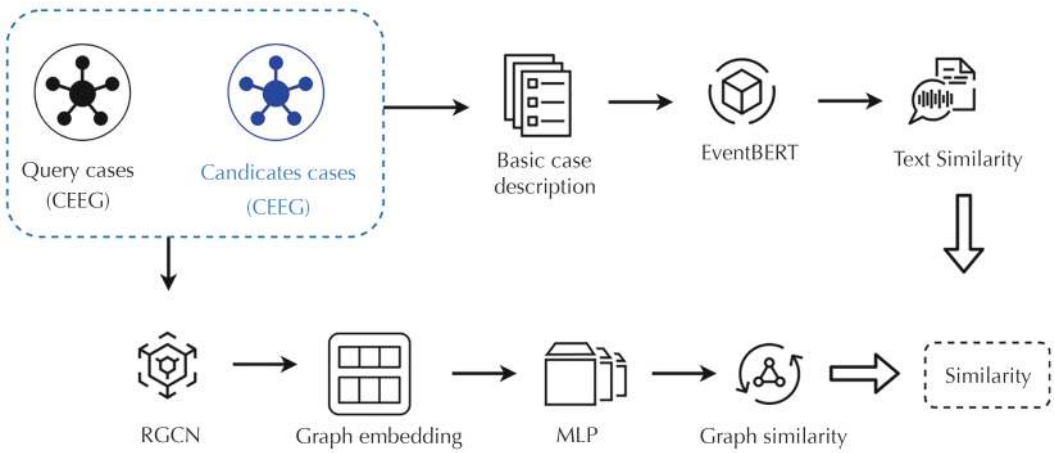


Figure 7. FTGS similar case retrieval algorithm.

3.2.1 Improved EventBERT model

In judicial documents, case descriptions are usually lengthy and filled with extraneous details. For that, the CEEG is employed to convert extracted triplets into concise text summaries, thereby distilling the essential information and core elements of the cases. The objective is to succinctly capture the key content of legal documents to support subsequent case retrieval tasks and balance retrieval efficiency with accuracy.

Once the basic case description summaries are obtained, the EventBERT model is utilized to integrate event information within the case descriptions. As shown in Figure 8, EventBERT enhances the original BERT model by incorporating the *Event Type Embeddings* in addition to the existing *Token Embeddings*, *Segment Embeddings*, and *Position Embeddings*.

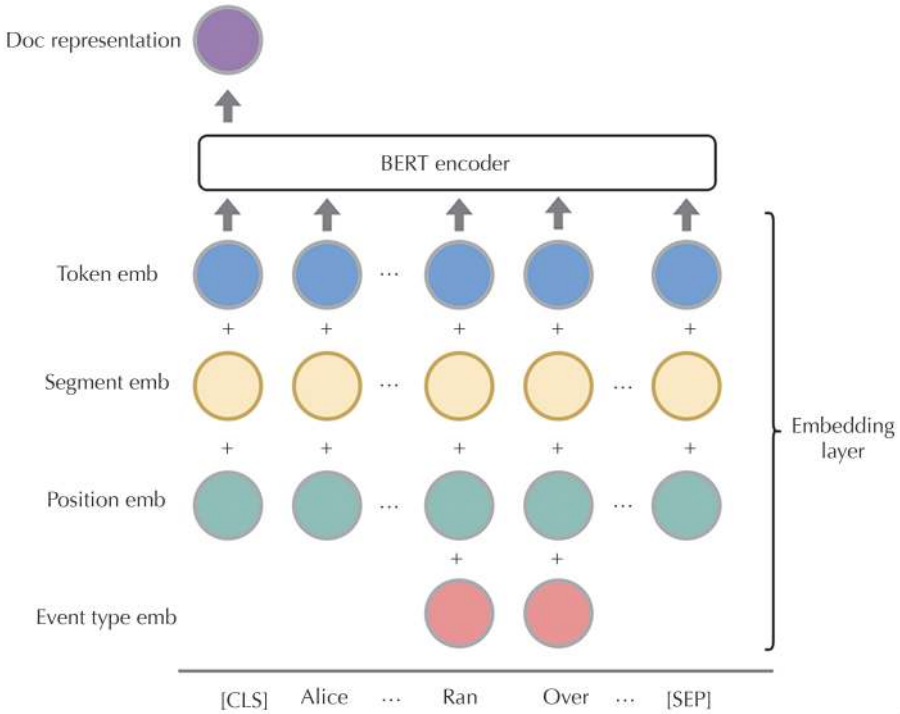


Figure 8. Embedding layer structure of EventBERT model.

During the construction of the CEEG, 300 distinct types of abstract events are identified and assigned unique IDs. These abstract events serve as trigger tokens within the input text summaries, where each event is represented with an Event Type Embedding, and its corresponding event type ID is added as an embedding to the input. For non-trigger tokens, the sum of their token embeddings and position embeddings is directly input into the encoder. Based on this structure, the EventBERT model is trained using a sentence pair classification framework. The input sequence is composed of concatenated query and candidate cases, and the model classifies the relationship between the two. The output is a two-dimensional probability distribution, indicating the confidence level that the query case and the candidate case are either related or unrelated.

3.2.2 The similarity calculation method of R-GCN

This study proposes a multi-relational graph convolutional network (R-GCN) method for similarity calculation, with the detailed process illustrated in Figure 9.

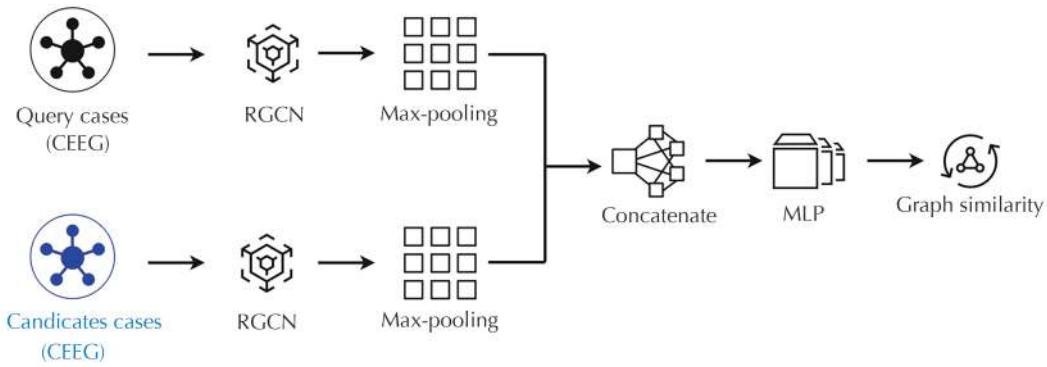


Figure 9. Similarity calculation method of R-GCN.

The similarity calculation considers both the node and arc similarities. During the graph fusion phase, the word embedding information obtained via the BERT-Base-Chinese model is used as the vector representation of the nodes, transforming abstract events into a form comprehensible to the model. The input graph is defined as G = (*V*, *E*, *R*), where *V* is the set of nodes, storing the word vector of each node; *E* is the set of arcs, storing the graph structure information of the CEEG through an adjacency matrix; and *R* is the set of relationships in the graph, primarily including subject-object relationships, sequential relationships, and causal relationships. The forward update formula for nodes in the R-GCN model is shown in Eq. (7).

$$h_i^{(l+1)} = \sigma\left(\sum_{r \in \mathcal{R}} \sum_{j \in \mathcal{N}_i^r} \frac{1}{C_{i,r}} W_r^{(l)} h_j^{(l)} + W_0^{(l)} h_i^{(l)}\right) \tag{7}$$

In which, $h_i^{(l+1)}$ represents the embedding of node *i* in the *l*-th layer of the graph convolutional network. $C_{i,r}$ is a normalization constant, set to the number of neighbors that node *i* can find through relation *r*. $W_r^{(l)}$ is the weight matrix for relation *r* in the *l*-th layer, while $W_0^{(l)}$ is the self-loop weight matrix, which is continuously updated during the training process. The ReLU function is utilized as the activation function to compute the embedding $h_i^{(l+1)}$ of node *i* in the (*l* + 1)-th layer.

$$\text{GraphPred}_{query,\ candidates} = \text{softmax}\left(\sigma\left(W_i\left(\sigma\left(W_0(g_{query} \oplus g_{candidates}) + b_0\right)\right) + b_i\right)\right) \tag{8}$$

Finally, the results obtained through GCN are linearly combined with the results predicted by EventBERT at the textual level to generate the similarity score for the current query case and candidate cases. The similarity scores of all candidate cases are then compared to obtain the final ranking result as:

$$\text{SimilarityPred}_{query,\ can} = \alpha \times \text{GraphPred}_{query,\ can} + \beta \times \text{AbstractPred}_{query,\ can} \tag{9}$$

In which α and β are adjustable parameters that can be tuned to enhance or diminish the influence of textual and graph similarities on the final result.

3.3 The algorithm of MRCI charge prediction

The MRCI algorithm is proposed to predict potential judgment outcomes by incorporating graph simplification, subgraph partitioning, and root cause inference modules. Take Figure 10 as an example, which illustrates a fused graph for the case of "*Wang Moutao's Public Brawl*". In the graph, red square nodes represent the charges associated with the case, green circular nodes represent abstract events and key legal elements mentioned in the case description, yellow nodes represent the descriptions of charges in related legal statutes, and purple nodes represent nodes that appear in both the knowledge graph and the event-effect graph. The node types corresponding to each node ID are listed in Table 1.

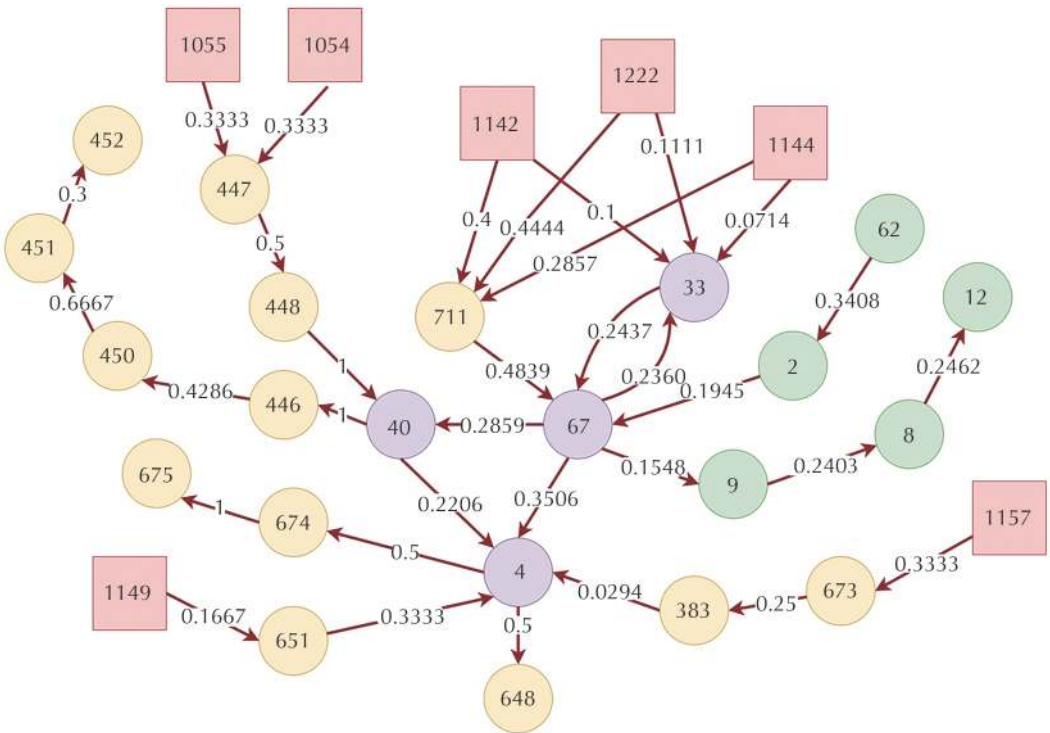


Figure 10. An illustrative example of fused graph for case.

Table 1 Node ID and Type.

ID	Type	ID	Type
1054	Crime of fire safety liability accident	67	Brawling
1055	Crime of falsely reporting a safety incident	383	Conduct
1142	Murder	446	Take
1144	Intentional injury	447	Fire safety
1149	Illegal detention	448	Supervise
1157	Abuse of a person under supervision	450	Correction
1222	Crime of organizing or participating in group brawls	451	Refusal
2	Contact	452	Execution
4	Personal injury	648	Humiliate
8	Injury description (Minor injury)	651	Possess
9	Injury description (Second-degree minor injury)	673	Supervision
12	Forgiveness	675	Abuse
33	Armed with a firearm/weapon	711	Assemble
40	Notify/Remind	674	Corporal punishment
62	Verbal dispute		

3.3.1 The method of graph simplification and subgraph partitioning

Based on the constructed K-CEEG, the DUCG causality graph simplification and subgraph partitioning is employed to eliminate event nodes that are unrelated to the root cause charge and to disentangle the causal chains in complex cases. This process effectively identifies evidence nodes related to each charge.

(1) Process of graph simplification

Starting from each charge B, a depth-first search (DFS) traversal is conducted. During traversal, all downstream evidence nodes accessible from B are considered to be covered by this root cause. After traversal, the nodes X that cannot be explained by B are deleted.

(2) Process of subgraph partitioning

After simplifying the graph G to G', we further partition G' into multiple subgraphs based on the DFS traversal results.

3.3.2 The method of root cause inference based on multi-path expansion

MRCI represents causal relationships between evidence nodes through relatively independent random events. For each subgraph obtained from the partitioning process, the algorithm infers the posterior probability $\Pr\{H_{k,j}|E\}$ for a given applicable charge B_k, which is the prediction result for root cause of the charge, as defined in Eq. (10),

$$h^s_{k,j} \equiv \Pr\left\{H_{k,j} \mid E\right\} = \frac{\Pr\left\{H_{k,j}E\right\}}{\Pr\{E\}} = \frac{\Pr\left\{H_{k,j}\prod_i E_i\right\}}{\Pr\left\{\prod_i E_i\right\}} \quad .(10)$$

In which $H_{k,j}$ represents the hypothesis event corresponding to the candidate charge B_k in K-CEEG, and E indicates the set of observed events (case facts), while E_i represents a specific observed event within that set, $E = \prod_i E_i$. Subsequently, the obtained subgraphs are compared and the prediction results are normalized according to Eq. (11), $\tilde{h}^s_{k,j}$ denotes the normalized score for the hypothesis corresponding to the candidate charge B_k under subgraph j, derived from the causal event knowledge graph. The original score $h^s_{k,j}$ is scaled by a normalization factor ξ_i, which is defined as the probability of observing the event E in subgraph i, normalized over all subgraphs. Specifically, $\xi_i = \Pr\{E\} / \sum_i \Pr\{E\}$ ensures that each subgraph's contribution is weighted proportionally to the likelihood of the observed evidence, thereby enabling fair comparison across multiple candidate subgraphs.

$$\tilde{h}^s_{k,j} = h^s_{k,j} \cdot \xi_i, \ \xi_i = \frac{\Pr\{E\}}{\sum_i \Pr\{E\}} \quad (11)$$

As shown in Figure 11, considering the subgraph partitioning results for case B_{1055}, the observed 11 event nodes can be treated as evidence, $E = E_{452}E_{451}E_{450}E_{446}E_{40}E_{448}E_{447}E_{675}E_{674}E_4E_{648}$. Then the logical expansion is performed on each of the evidence event, e.g., E_4, E_{446}, and E_{674} are respectively expanded as following.

$$E_4 = X_{4,1} = A_{4,1;40,1}A_{40,1;448,1}A_{448,1;447,1}A_{447,1;1055,1}B_{1055,1}$$

$$E_{446} = X_{446,1} = A_{446,1;40,1}A_{40,1;448,1}A_{448,1;447,1}A_{447,1;1055,1}B_{1055,1}$$

$$E_{674} = X_{674,1} = A_{674,1;4,1}A_{4,1;40,1}A_{40,1;448,1}A_{448,1;447,1}A_{447,1;1055,1}B_{1055,1}$$

Note that the causal coefficients calculated during the graph construction are defined as the probabilities $A_{n,k;i,j}$ of event occurrence. The prior probability of charge prediction, $B_{1055,1}$, is defined by combining the frequency of its occurrence in a large-scale corpus of judicial documents with expert judgment. For the purpose of simplicity and in order to mitigate the impact of uneven distributions of high- and low-frequency charges on the results, we assume that the prior probabilities of all charges are equal. In the final calculation of the B_{1055} charge prediction task, the outcome is derived as the product of causal coefficients across all logical pathways generated through the multi-path expansion of the root cause.

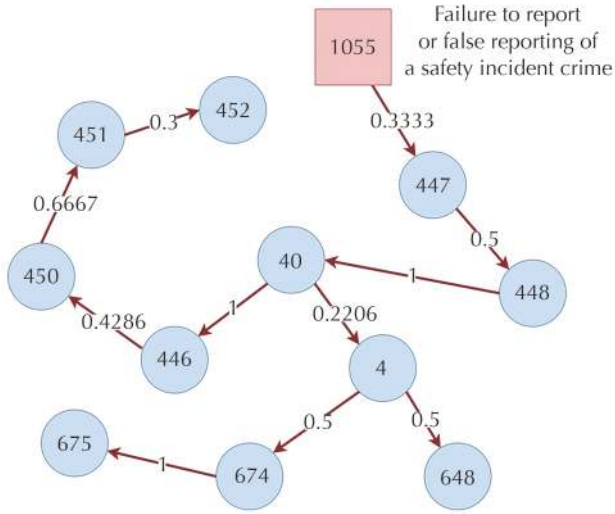


Figure 11. Root cause subgraph partitioning results for case B_{1055}.

Furthermore, to refine and optimize the charge ranking results, the explanatory coverage of the root cause over relevant downstream nodes is considered. By calculating the proportion of downstream nodes associated with each root cause within each subgraph relative to all nodes in the complete graph, predictions not meeting a predetermined threshold are removed. This ensures that the remaining root causes more accurately reflect a broader range of critical nodes in the integrated graph, eliminating charge predictions with low relevance to the actual case. The threshold is adjustable and is set to 60% in this study for subsequent analysis.

The subgraphs corresponding to other charges in the example are computed using the same method. Based on this approach, the final charge prediction ranking for the example case is derived. Charges from the integrated graph that are associated with the case description but have low relevance, such as "*Crime of fire safety liability accident (ID: 1054)*", "*Crime of falsely reporting a safety incident (ID: 1055)*", "*Illegal Detention (ID: 1049)*", and "*Abuse of a Person under Supervision (ID: 1057)*", do not meet the requirement of covering over 60% of downstream nodes. Therefore, these charges are not included in Table 2.

Table 2. Sample case charge prediction results.

Charge	MRCI prediction result
Crime of Organizing or Participating in Group Brawls (ID: 1222)	0.4498
Intentional Injury (ID: 1144)	0.3643
Murder (ID: 1142)	0.1859

3.4 The causal inference-based sentencing prediction algorithm

The core concept of the MRCI charge prediction algorithm is performing the casual reasoning based on the observed evidence and predict the applicable charges for the case. The algorithm process is as follows.

Step 1: Simplification of the integrated graph. After determining the applicable charges for the case using the MRCI algorithm, remove other root causes from the K-CEEG and simplify the graph by eliminating event nodes unrelated to the applicable charges.

Step 2: Connection of sentencing nodes. In previous charge prediction task, the sentencing nodes were simplified or omitted, as they were not essential to the objective. However, for the purpose of sentence length inference in this task, these nodes need to be re-integrated into the graph structure.

Step 3: Adjustment of causal coefficients based on similar cases. Considering the importance of similar cases in the sentencing process, integrate the sentencing results from the top 30 most similar candidate cases and adjust the causal coefficients of arcs between event and sentencing nodes accordingly.

For the example cases, approximately 83.33% of similar cases obtain sentencing results within the range of less than 3 years, while the 16.67% cases fall within the interval of 3 to 10 years. The statistics are combined with the original causal coefficients and divide by 2 to adjust the coefficient, as shown on the arcs in Figure 12.

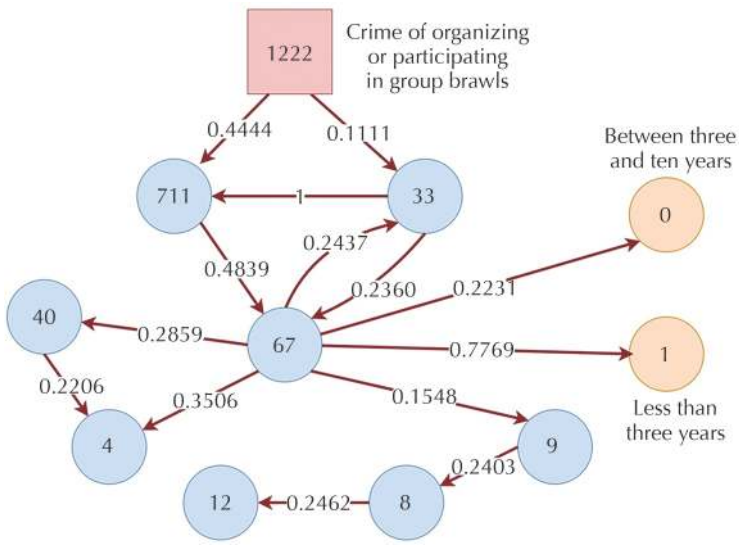


Figure 12. Simplified results of example charge prediction.

Nodes 0 and 1 in the Figure 12 represent the potential sentencing ranges for *Crime of Organizing or Participating in Group Brawls*, that is applicable to this case. Further, the probabilities of the two sentencing outcomes are calculated by Eq. (12).

$$\Pr(X_i) = \Pr\left\{\sum_i \sum_{j_i} \left(\frac{r_{n;\,i}}{r_n}\right) A_{n,\,k;\,i,\,j_i} V_{i,\,j_i}\right\} \tag{12}$$

In which, X_i represents the sentencing nodes to be predicted. Take Figure 11 as an example, in which all the nodes connected to the sentencing nodes are observed events, and the inference is performed as Eq. (13).

$$X_{n,\,k} = \sum_i \sum_{j_i} \left(\frac{r_{n;\,i}}{r_n}\right) A_{n,\,k;\,i,\,j_i} V_{i,\,j_i} \tag{13}$$

The calculated probabilities of X_0 and X_1 indicate that the sentencing range for this case is "less than three years." In the actual court ruling, the case was sentenced to 8 months, demonstrating the method's effectiveness.

$$\Pr(X_0) = \Pr(X_{0,\,1}) = 0.3462$$
$$\Pr(X_1) = \Pr(X_{1,\,1}) = 0.7769$$

4. EXPERIMENTS

4.1 Experiments on similar case retrieval

In order to assess the practical effectiveness of the proposed FTGS similar case retrieval algorithm, the comparative experiments are conducted on three widely used baseline models: TF-IDF, BM25, and BERT. Additionally, the EventBERT model is introduced to evaluate the performance gains achieved by integrating graph neural networks, which processes only textual information in the evaluation.

The experiments use the publicly available LeCaRD dataset [31] for the similar case retrieval task track in the 2022 CAIL competition. It contains 107 query cases and over 10,000 candidate cases. For each query case, 30 candidate cases are selected, with the similarity ranking criteria provided by domain experts. The experimental evaluation metrics include mean average precision (MAP), normalized discounted cumulative gain (NDCG@k), and precision (P@k), which are commonly used in the field of information retrieval.

The 5-fold cross-validation is employed in evaluation by evenly and randomly dividing the original dataset into five subsets. Each subset is used as a test set in turn, with the remaining four subsets serving as the training set. The models are trained and evaluated on the test set, and this experimental method helps to better analyze the generalization ability of the models. Each subset served as the test set once, and the performance on the test set is measured using several metrics. The final evaluation result for the models is obtained by averaging the results of the five tests. This approach reduces the randomness in model performance evaluation and provides more reliable results by averaging multiple random splits of the dataset. The performance results of each model algorithm are shown in Table 3 and Figure 13.

Table 3. Performance of models on the LeCaRD dataset.

Models	MAP	NDCG@5	NDCG@10	NDCG@20	P@5	P@10	P@20
TF-IDF	40.17	69.13	74.59	78.83	23.60	24.32	23.82
BM25	42.45	63.57	68.67	72.53	31.52	32.61	31.09
BERT	45.64	68.90	74.37	79.70	34.55	38.18	36.59
EventBERT	45.45	75.72	77.74	82.49	40.91	38.64	35.68
FTGS	**48.12**	**78.43**	**80.54**	**83.68**	**41.82**	**40.00**	**37.27**

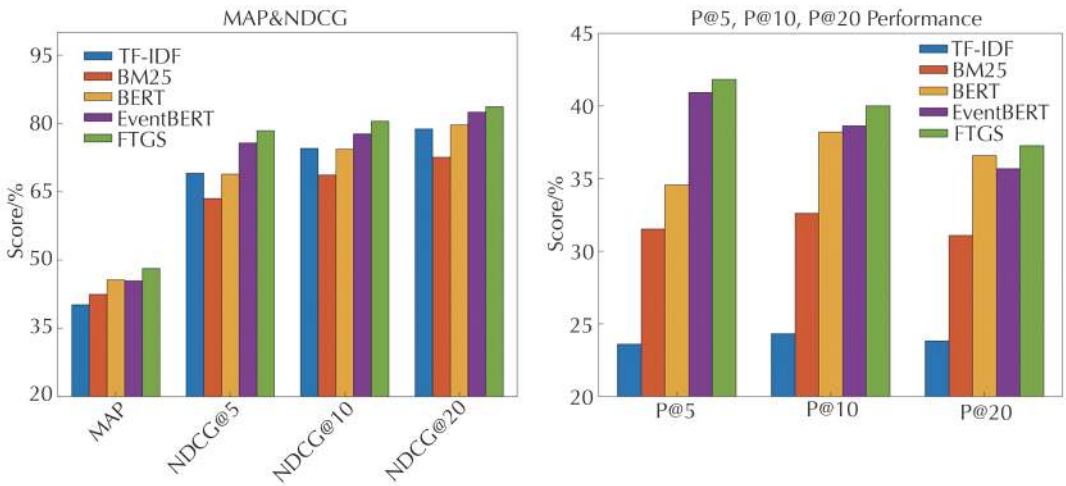


Figure 13. Results of each model for similar case retrieval.

Analysis of the charts indicates that the FTGS similar case retrieval algorithm outperforms all baseline models across various metrics. Compared to the EventBERT model, which solely considers textual features, the integration of graph neural networks significantly improves performance by effectively leveraging graph structural information in the similar case retrieval task. Consequently, this experiment validates the effectiveness of combining textual and structural features for more robust legal judgment predictions.

4.2 Experiments on charge prediction

In order to validate the model's effectiveness in charge prediction task, we conduct multi-label charge prediction experiments. Given that inaccurately predicting a charge may lead to misunderstandings of case judgment results, it is critical to prioritize accuracy. However, focusing solely on accuracy may result in the rare charges with fewer samples being overlooked, exacerbating the effects of data imbalance on the model's predictive capability. Furthermore, considering the complexity of legal decision-making and the pursuit of justice, it is essential to account for all potential judgment outcomes comprehensively. Therefore,

we use recall and F1-score as two metrics, along with the interpretability of the reasoning process, to evaluate the algorithm's performance from three perspectives. The baseline models chosen for this comparative experiment include commonly used algorithms in text classification tasks, such as Support Vector Machine (SVM), Convolutional Neural Network (CNN), and Bidirectional Long Short-Term Memory Network (BiLSTM).

The dataset for this task is also derived from LeCaRD, with 1,200 samples randomly selected to simulate a small sample environment for the experiments. Since the dataset lacks true and accurate charge judgment labels, we utilize the ChatGLM3-turbo large model API to generate pseudo-labels by inputting case names into a constructed template. These generated labels are then manually reviewed and revised to ensure the accuracy of the case judgment outcomes. Totally 10% of the data is extracted as the test set, and the remaining 90% is used as the training set. Table 4 and Figure 14 demonstrate the performance of the algorithms in the experiments.

Table 4. Performance of each model in the charge prediction task.

Models	Recall	F1-Score	Interpretability
SVM	20.56	16.05	×
CNN	31.07	44.75	×
BiLSTM	35.73	50.12	×
MRCI	**74.70**	**52.87**	√

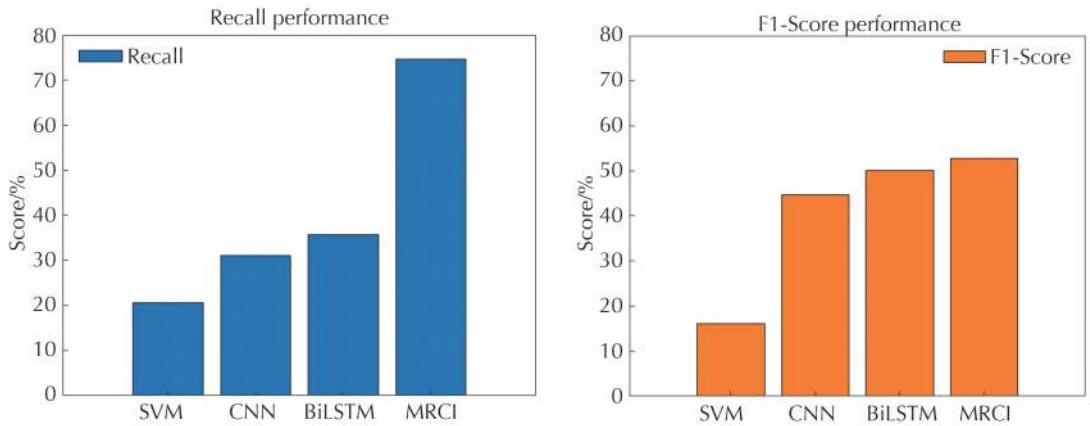


Figure 14. Results of each model for charge prediction.

The MRCI charge prediction algorithm achieves a higher recall rate compared to the other three models, demonstrating superior coverage of potential judgment outcomes, even with a limited sample size. It also attains a high F1-score, indicating that the algorithm effectively balances sensitivity and precision in charge

prediction, reducing both false positives and false negatives. This balance ensures comprehensive capture of all potential charge judgment outcomes. Furthermore, the MRCI algorithm excels in interpretability, as it is based on the K-CEEG framework. The causal reasoning process is clearly delineated along the event logic evolution chain, providing a transparent explanation of the causal relationships and sentencing procedures involved in the judgment process.

5. CONCLUSIONS

This study proposes a novel method for legal judgment prediction based on knowledge-integrated event evolutionary graphs and causal inference methods. The developed K-CEEG framework, which incorporates the FTGS similar case retrieval and MRCI charge prediction algorithms, addresses the limitations of existing methods by enhancing both the interpretability and accuracy of predictions. Experimental results validate the effectiveness of the proposed method in capturing the nuanced logic and causal structures inherent in legal documents. Future research will focus on refining sentencing prediction techniques, enhancing causal inference mechanisms, and extending the application of this model to a broader spectrum of legal scenarios. These ongoing efforts are intended to further strengthen the model's robustness and utility as a dependable tool for legal decision-making.

AUTHOR CONTRIBUTIONS

Chunling Dong designed the research framework, and Minsen Li and Xinyi Liu were responsible for the implementation of the approach and the analysis of the results. All the authors have made meaningful and valuable contributions in revising and proofreading the resulting manuscript.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 62176240, 62572446), the Beijing Education Science 14th Five Year Plan Project (Grant No. CDDB25231), the Fundamental Research Funds for the Central Universities (Grant No. CUC25GT04), and Public Computing Cloud, CUC.

DECLARATIONS

Competing Interests: All the authors declare that they have no competing financial interests or personal relationships that could influence the work reported in this paper.

AUTHORS CONTRIBUTION STATEMENT

Ethical and informed consent for data used: This study adheres to ethical standards, and informed consent is obtained from all participants. Data are anonymized and managed securely to protect confidentiality.

Data availability and access: The data used to support the findings of this study is available from the corresponding author upon request.

REFERENCES

[1] H. Chen, L. Zhang, Y. Liu, et al., "Rethinking the Development of Large Language Models from the Causal Perspective: A Legal Text Prediction Case Study," in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, no. 19, pp. 20958–20966, 2024.

[2] Y. Shao, J. Mao, Y. Liu, et al., "BERT-PLI: Modeling paragraph-level interactions for legal case retrieval," in Proceedings of the International Joint Conference on Artificial Intelligence, pp. 3501–3507, 2020.

[3] C. Xiao, X. Hu, Z. Liu, et al., "Lawformer: A pre-trained language model for Chinese legal long documents," AI Open, vol. 2, pp. 79–84, 2021.

[4] H. Li, Q. Ai, J. Chen, et al., "SAILER: Structure-aware pre-trained language model for legal case retrieval," in Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1035–1044, 2023.

[5] K. Zhang, C. Chen, Y. Wang, et al., "Cfgl-lcr: A counterfactual graph learning framework for legal case retrieval," in Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, vol. 1, no. 2, pp. 3332–3341, 2023.

[6] Y. Tang, R. Qiu, Y. Liu, et al., "CaseGNN: Graph neural networks for legal case retrieval with text-attributed graphs," in Proceedings of the European Conference on Information Retrieval, pp. 80–95, 2024.

[7] S. Gao, Y. Li, F. Ge, et al., "Match and retrieval: Legal similar case retrieval via graph matching network," in Proceedings of the 2023 IEEE International Conference on Data Mining Workshops, pp. 227–234, 2023.

[8] Y. Zhang, S. Zhai, Y. Meng, et al., "Event is more valuable than you think: Improving the similar legal case retrieval via event knowledge," Information Processing & Management, vol. 61, no. 4, p. 103729, 2024.

[9] L. Liu, D. An, Y. Wang, et al., "Research on legal judgment prediction based on BERT and LSTM-CNN fusion model," in Proceedings of the 3rd World Symposium on Artificial Intelligence, pp. 41–45, 2021.

[10] B. Luo, Y. Feng, J. Xu, et al., "Learning to predict charges for criminal cases with legal basis," in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2727–2736, 2017.

[11] H. Zhang, Z. Dou, Y. Zhu, et al., "Contrastive learning for legal judgment prediction," ACM Transactions on Information Systems, vol. 41, no. 4, pp. 1–25, 2023.

[12] Z. Hu, X. Li, C. Tu, et al., "Few-shot charge prediction with discriminative legal attributes," in Proceedings of the International Conference on Computational Linguistics, pp. 487–498, 2018.

[13] C. He, L. Peng, Y. Le, et al., "SECaps: A sequence enhanced capsule model for charge prediction," in Proceedings of the 28th International Conference on Artificial Neural Networks, pp. 227–239, 2019.

[14] Q. Zhao, T. Gao, S. Zhou, et al., "A survey of the state of the art," Applied Sciences, vol. 12, no. 5, pp. 2531–2534, 2022.

[15] X. Zhou, Q. Liu, Y. Wu, et al., "LK-IB: A hybrid framework with legal knowledge injection for compulsory measure prediction," Artificial Intelligence and Law, vol. 32, no. 3, pp. 1–26, 2023.

[16] S. Tong, J. Yuan, P. Zhang, et al., "Legal judgment prediction via graph boosting with constraints," Information Processing & Management, vol. 61, no. 3, p. 103663, 2024.
[17] Y. Kong, Y. Wang, et al., "LF-HGRILF: A law-fact heterogeneous graph representation and iterative learning framework for legal judgment prediction," Knowledge-Based Systems, vol. 274, p. 114083, 2025.
[18] C. Meng, Y. Todo, C. Tang, et al., "DPFSI: A legal judgment prediction method based on deontic logic prompt and fusion of law article statistical information," Expert Systems with Applications, vol. 272, p. 126722, 2025.
[19] X. Guo, H. Zhang, L. Ye, et al., "TenLa: An approach based on controllable tensor decomposition and optimized lasso regression for judgement prediction of legal cases," Applied Intelligence, vol. 51, no. 4, pp. 2233–2252, 2021.
[20] Y. Tang, R. Qiu, H. Yin, et al., "Caselink: Inductive graph learning for legal case retrieval," in Proceedings of the 47th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 2199–2209, 2024.
[21] S. Zhou, Y. Liu, Y. Wu, et al., "Similar case based prison term prediction," in Proceedings of the Chinese Association for Artificial Intelligence International Conference on Artificial Intelligence, pp. 284–297, 2022.
[22] H. Li, S. Cai, Z. Ming, "Legal judgment prediction incorporating guiding cases matching," in Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computin, pp. 511–523, 2023.
[23] F. Yao, X. Sun, H. Yu, et al., "Gated hierarchical multi-task learning network for judicial decision prediction," Neurocomputing, vol. 411, pp. 313–326, 2020.
[24] F. Yao, X. Sun, H. Yu, et al., "Commonalities-, specificities-, and dependencies-enhanced multi-task learning network for judicial decision prediction," Neurocomputing, vol. 433, pp. 169–180, 2021.
[25] Y. Dong, X. Li, J. Shi, et al., "Graph contrastive learning networks with augmentation for legal judgment prediction," Artificial Intelligence and Law, pp. 1–24, 2024.
[26] P. Wang, X. Zhang, H. Yu, et al., "Interpretable prison term prediction with reinforce learning and attention," Applied Intelligence, vol. 53, no. 2, pp. 1306–1323, 2023.
[27] Y. Feng, C. Li, V. Ng, "Legal judgment prediction: A survey of the state of the art," in Proceedings of the International Joint Conference on Artificial Intelligence, pp. 5461–5469, 2022.
[28] Q. Zhang, C. Dong, Y. Cui, et al., "Dynamic uncertain causality graph for knowledge representation and probabilistic reasoning: Statistics base, matrix, and application," IEEE Transactions on Neural Networks and Learning Systems, vol. 25, no. 4, pp. 645–663, 2014.
[29] C. Dong, Q. Zhang, "The cubic dynamic uncertain causality graph: A methodology for temporal process modeling and diagnostic logic inference," IEEE Transactions on Neural Networks and Learning Systems, vol. 31, no. 10, pp. 4239–4253, 2020.
[30] C. Dong, J. Zhou, "A new algorithm of cubic dynamic uncertain causality graph for speeding up temporal causality inference in fault diagnosis," IEEE Transactions on Reliability, vol. 72, no. 2, pp. 662–677, 2023.
[31] F. Yao, C. Xiao, X. Wang, et al., "LEVEN: A large-scale Chinese legal event detection dataset," in Proceedings of the Association for Computational Linguistics, pp. 183–201, 2022.

AUTHOR BIOGRAPHY



Chunling Dong is a professor and Ph.D. supervisor with Communication University of China, and a Secretary General of the Technical Committee for Causality and Uncertainty in Artificial Intelligence of the Chinese Association for Artificial Intelligence.
She was a Post-Doctoral Fellow with the Department of Computer Science and Technology, Tsinghua University, China, and she received the Ph.D. degree in computer science and technology from Beihang University, China, in 2015. Her research interests include causality representation and inference under uncertainties, and diagnostic reasoning and decision-making methods for complex system.



Mingsen Li received the bachelor's degree in computer science and technology from Zhengzhou University in 2023 and is currently pursuing the master's degree in computer science and technology from Communication University of China.



Xinyi Liu received the M.S. degree in electronic information from Communication University of China, China, in 2024, and the B.S. degree in digital media technology from Beijing Language and Culture University, China, in 2022.