Proceedings of Machine Learning Research 260:-, 2024 ACML 2024

A Retrieval-Augmented Contrastive Framework for Legal Case Retrieval Based on Event Information

Changyong Fan TREVO1ZZZ@OUTLOOK.COM
School of Information Science and Technology,
Guangdong University of Foreign Studies, Guangzhou, 510006, Guangdong, China

Nankai Lin NEAKAIL@OUTLOOK.COM
School of Computer Science and Technology,
Guangdong University of Technology, Guangzhou, 510006, Guangdong, China

Dong Zhou✉ DONGZHOU@GDUFS.EDU.CN
School of Information Science and Technology,
Guangdong University of Foreign Studies, Guangzhou, 510006, Guangdong, China

Yongmei Zhou✉ YONGMEIZHOU@163.COM
School of Information Science and Technology,
Guangdong University of Foreign Studies, Guangzhou, 510006, Guangdong, China

Aimin Yang AMYANG18@163.COM
School of Computer Science and Intelligence Education,
Lingnan Normal University, Zhanjiang, 524000, Guangdong, China

Abstract

Similar case retrieval is a crucial aspect of the legal retrieval field, significantly contributing to LegalAI systems. This task aims to retrieve cases that are highly relevant to the query case, thereby enhancing the efficiency of legal practitioners. Recent methods have leveraged the rich semantic knowledge of pre-trained models, greatly improving retrieval performance. However, these methods often overlook key legal elements within the complex language structures of case texts, such as legal event information that can impact case outcomes and judgments. This oversight results in the underutilization of critical case information. To address this issue, we proposed RAEvent, a similar case retrieval contrastive framework augmented by legal event information. This framework utilizes an enhanced case event information database to provide auxiliary information for case retrieval and employs contrastive learning techniques to better extract similar features in cases. In comparison to a range of baseline approaches, the results of our experiments highlight the efficacy of our framework. Moreover, our research provides fresh perspectives and makes a valuable contribution to the ongoing studies in similar case retrieval tasks.

Keywords: Similar Case Retrieval, Legal Case, Event Extraction

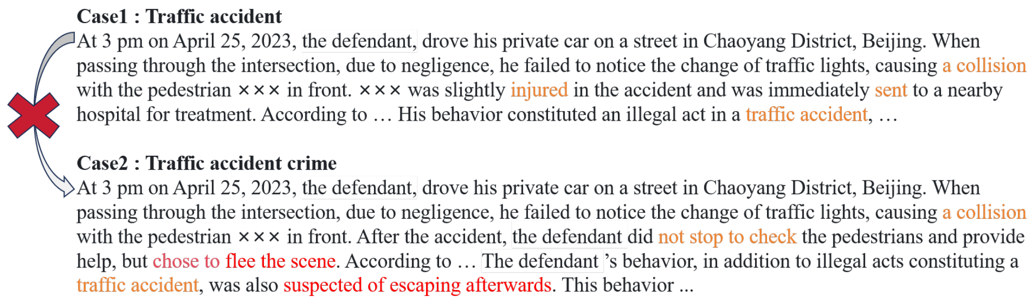

Figure 1: Illustration of an example of case matching. Case 1 and case2 exhibit a high degree of textual similarity. However, due to the presence of the critical legal event "hit and run" in case 2, these two cases cannot be considered a match from a legal perspective.

1. Introduction

With the continuous advancement of the LegalAI (Zhong et al., 2020) field, artificial intelligence technology is poised to assume a more pivotal role in the legal field, bringing forth both opportunities and challenges to legal practices. The process of retrieving similar cases entails finding pertinent cases that align with the query case. Legally speaking, case precedents are historical cases that exhibit comparable legal facts and issues to a particular case. However, as the volume of legal case documents rapidly grows, relying on manual searches for relevant case precedents is becoming increasingly impractical. Consequently, the study of similar case retrieval technology has attracted significant attention from both the legal and information retrieval community (Althammer et al., 2021; Yu et al., 2022). In the task of similar case retrieval, merely transforming the retrieval task into a problem of computing case similarity is insufficient. Whether candidate cases can serve as case precedents for query cases should not depend solely on overall text similarity but also on the key legal event element information between cases. As shown in Figure 1, the texts of two cases may be highly similar, but from a legal perspective, they cannot be matched as similar cases. This discrepancy arises due to significant differences in key elements between the two cases. Firstly, there is a difference in specific behavior. Notably, the defendant's escape behavior in case 2 is an additional and crucial legal event that significantly affects legal judgment and liability determination. Secondly, there is an aggravation of legal liability. In case 2, due to the involvement of post-incident escape, the defendant's legal liability is more severe than in case 1, including criminal liability.

In recent years, similar case retrieval has become an indispensable component of intelligent legal systems. With the continuous exploration of similar case retrieval by scholars, its performance has been further improved. As part of information retrieval, traditional bag-of-words retrieval models still perform excellently (Sparck Jones, 1972; Robertson and Walker, 1994; Ponte and Croft, 2017). With the in-depth research of pre-trained models, many pre-trained models in the legal field have been developed to solve downstream legal tasks, utilizing the learned legal semantics of pre-trained models to convert case texts into

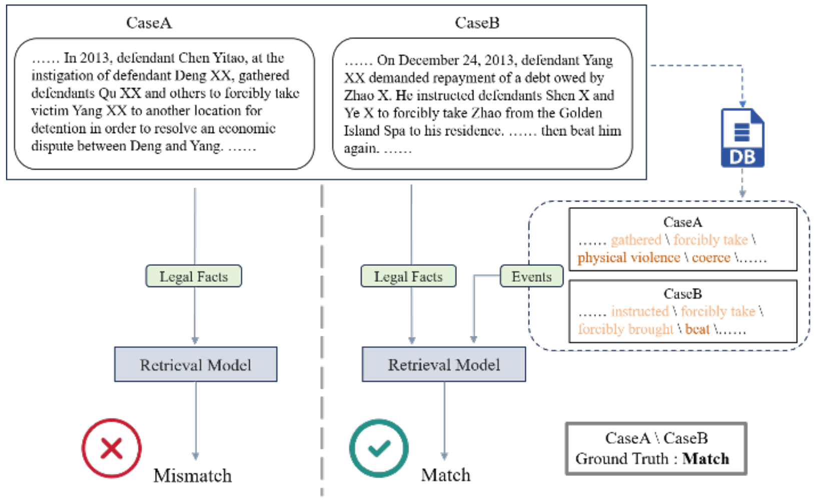

Figure 2: A comparative illustration of case matching methods. **Left**: Encoding the overall information of the case pair and focusing on the overall text similarity for matching. **Right**: Emphasizing finer-grained legal event information within the cases and using it as auxiliary information to further improve the reliability of retrieval results.

vector representations (Chalkidis et al., 2020; Xiao et al., 2021). Additionally, many methods have now developed corresponding neural network model architectures specifically for similar case retrieval tasks, further optimizing the quality of case retrieval (Li et al., 2023b; Xiao et al., 2022; Sun et al., 2024; Tang et al., 2023). Although these methods demonstrate excellent retrieval performance, they often overlook key legal elements in cases with complex linguistic structures, such as legal events that affect case outcomes and judgments, resulting in the underutilization of case information. As shown in Figure 2, previous methods (on the left) tend to calculate the overall similarity of cases, underutilizing finer-grained information such as key event details within the cases. This approach can lead to incorrect results due to the influence of extraneous information.

Noticing the aforementioned issues, we proposed RAEvent, a retrieval-augmented contrastive framework based on legal event information to address this challenge. To fully leverage crucial legal event information in cases, we adopt legal event extraction as an auxiliary task for subsequent case retrieval. To enhance the performance of legal event extraction, we first fine-tune the model on a legal event dataset. This pre-trained model integrates the contextual information of cases and employs attention mechanism to identify the categories and locations of key legal event information. Next, we use this extraction model to build an enhanced legal event information database, serving as auxiliary information for the retrieval model. Using the key facts of the case texts, specifically the event information features, we apply supervised contrastive learning to bring the features of the query case and candidate cases closer, enabling the retrieval model to learn similar cases based on legal event features. Final experimental results indicate that RAEvent surpasses existing baseline methods, showcasing enhanced retrieval capabilities.

To summarize, the key contributions of this work are:

1. We proposed a retrieval-augmented contrastive framework based on legal event information for similar case retrieval, referred to as RAEvent. RAEvent emphasizes key legal elements in cases rather than mere similarity, leveraging legal event information to enhance the retrieval task. Experimental results confirm the effectiveness of RAEvent.

2. We utilize contrastive learning to improve the retrieval model's ability to extract similar features from case pairs, thereby enhancing overall retrieval performance. Ablation experiments have demonstrated the effectiveness of this module.

3. RAEvent demonstrates high flexibility with modular components that complement each other. Each module can be independently swapped out, ensuring that future upgrades can be made without affecting the overall system.

2. Related Work

2.1. Legal Case Retrieval

Traditional retrieval methods, including statistical models, performed well in the retrieval field. For instance, TFIDF (Sparck Jones, 1972), BM25 (Robertson and Walker, 1994), and LMIR (Ponte and Croft, 2017) evaluated the similarity between query cases and candidate cases based on factors such as word frequency, average document length, and inverse document frequency. These methods showed good performance in legal retrieval.

The development of natural language processing, particularly the emergence of pre-trained models, further enhanced retrieval effectiveness. Pre-trained models leveraged their semantic capabilities and linguistic knowledge to understand contextual information and address high-dimensional challenges (Dai and Callan, 2019; Li et al., 2023a). Given the specialization of the legal field, many methods focused on legal-specific models. Liu et al. (2021) proposed a neural network-based case recommendation model that calculated similarity scores using vectors from the legal facts of case texts, recommending cases with the highest scores. Chalkidis et al. (2020) developed a legal language model pre-trained on a large legal corpus. To handle long legal texts, Lawformer (Xiao et al., 2021) combined three attention mechanisms to overcome input limitations. Shao et al. (2020) used a segmented interaction approach, calculating paragraph-level similarity before aggregating results. SAILER (Li et al., 2023b), a structure-aware pre-trained language model, encoded decisions, facts, and reasoning, using the facts part to represent relevant cases.

Despite their success, these methods overlooked key legal elements, such as legal event information, that influence case outcomes. Additionally, pre-trained model encoders did not fully grasp the logical relationships within legal documents due to the abundance of non-essential factual information. We believe the potential application of key event information extraction in retrieval has not been fully realized.

2.2. Event Extraction in Legal Domain

Event extraction is an essential task in information extraction, aimed at identifying event details from textual data, which has garnered considerable attention from scholars. Traditional methods relied on manually designed features (Hong et al., 2011; Ji and Grishman, 2008), such as syntactic, document-level, and entity-level features. With the advent of deep learning, current methods primarily use neural network-based models and other advanced architectures. In the legal field, various studies have defined legal events and built models for automatically extracting these events from factual narratives (Chen et al., 2020; Shen et al., 2020). Feng et al. (2022) introduced a constrained event-driven prediction model that pinpointed key event details to assist in determining judgments, setting a new benchmark for Legal Judgment Prediction. Moreover, researchers have explored deep neural networks in the context of legal documents. Chen et al. (2020) identified entities and semantic relationships in drug-related legal cases. To tackle real-world legal challenges, Li et al. (2020) applied event extraction methods to the case descriptions found in Chinese legal texts.

Advancements in event extraction have significantly promoted the development of legal intelligence. However, previous methods seldom used event extraction to support downstream legal tasks, particularly in similar case retrieval. This study aims to utilize event extraction to extract key legal event information from cases, serving as auxiliary information for downstream retrieval tasks.

3. Methodology

In this section, we begin by outlining our task formulation before detailing the RAEvent framework, as depicted in Figure 3. RAEvent consists of two main components: (1) **Retrieval Augmentation.** Event extraction is employed as an upstream task. For a given query-candidate case set, the fine-tuned event extraction model extracts key legal event information, building an enhanced legal event knowledge base with this information. (2) **Legal Event-Centric Similar Case Retrieval.** The extracted key event information is combined with query-candidate cases as auxiliary information to complete similar case retrieval tasks.

3.1. Task Formulation

The task of Similar Case Retrieval (SCR) involves identifying cases in a candidate case set that are similar to a given query case (Rabelo et al., 2020). In comparison to conventional legal analysis tasks, SCR focuses on judging the similarities between cases rather than other aspects. Formally, given a query case S and a collection of candidate cases $\mathcal{C}$, where both are lengthy texts outlining legal facts, the candidate cases are represented as $\mathcal{C} = \{c_1, c_2, \cdots, c_n\}$. The goal is to identify a set of relevant cases $\mathcal{C}^* = \{c_i^* \mid c_i^* \in \mathcal{C} \wedge rel(c_i^*, S)\}$, where $rel(c_i^*, S)$ indicates that c_i^* is pertinent to the query S. Legally, such relevant cases are considered precedents, which are historical cases that share similar facts and issues with the query. Once the similarity scores for all candidates are calculated, they are ranked accordingly, and the model's performance is evaluated based on this ranking.

To enhance the retrieval task, we introduced legal event extraction as an auxiliary task. The legal event extraction task is defined as follows:

Given a token sequence $\mathcal{T} = \{t_1, t_2, \cdots, t_n\}$, the legal event extraction task identifies trigger words and determines the corresponding event type. Trigger words refer to keywords or phrases that can trigger specific events and are the most important factors in determining the type of event. Event type represents a category of events related to a legal context, such as "theft" or "escape", where the specific event type and number of events are determined by the dataset.

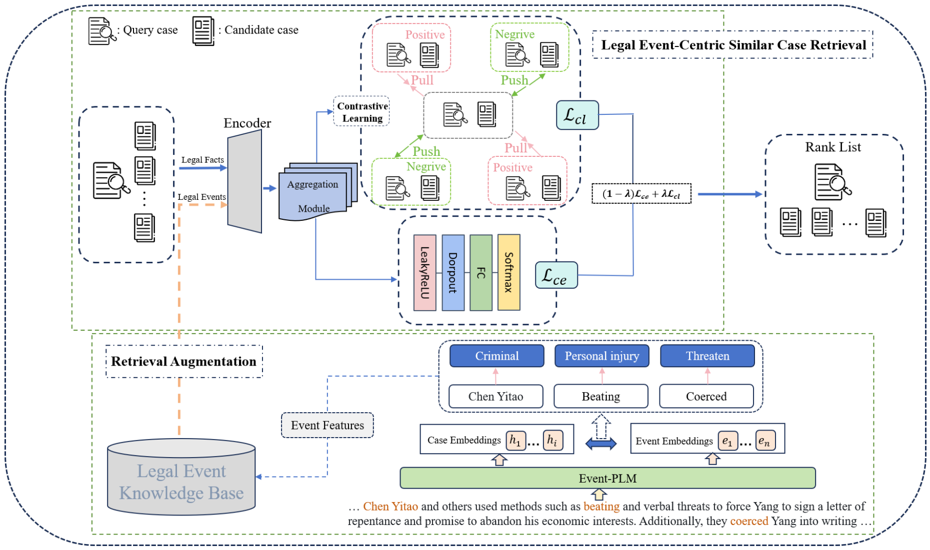

Figure 3: Overview of RAEvent framework. The RAEvent framework is broadly categorized into two main modules. The first module is the retrieval augmentation module, which constructs an augmented case event information database through the extraction of case event information. The second module is the retrieval module, based on supervised contrastive learning of event information.

3.2. Retrieval Augmentation

The key Retrieval Augmentation component can be further subdivided into (i) a legal event information extraction model, and (ii) the construction of an enhanced legal event knowledge base for similar case retrieval.

3.2.1. Legal Event Information Extraction Model

In our similar case retrieval system, accurately extracting and understanding event information in legal texts is key to improving retrieval efficiency and precision. The model for extracting legal event information is designed to forecast the event label e_i corresponding to every token. Our model is built upon the BERT architecture, utilizing its pre-trained bidirectional encoder representations to capture complex dependencies in legal language. Furthermore, a multi-head attention mechanism enhances the model's ability to focus on different semantic subspaces simultaneously, improving its overall information processing capabilities. In this task, the model output is the probability distribution of the event category for each token. Let K be the total number of event categories (including a "no event" category), $y_{i,k}$ be the true category of the i-th token (one-hot encoded), and $\hat{y}_{i,k}$ represents the model-predicted probability that the i-th token belongs to category k. During training, loss function $Loss$ as follows:

$$Loss = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{K} y_{i,k}\log(\hat{y}_{i,k}), \tag{1}$$

where N is the number of tokens in the input text.

We pre-train the model on the large-scale Chinese legal event extraction dataset LEVEN, which is meticulously annotated with event information in legal texts, encompassing both case-related and general events. Given an input text $s = t_1, t_2, \cdots, t_n$, The model captures the representation of cases as follows:

$$input = [CLS]t_1, t_2, \cdots, t_n[SEP], \tag{2}$$

$$h_{input} = h_{[CLS]}, h_1^t, h_2^t, \cdots, h_n^t, h_{[SEP]} = BERT(input), \tag{3}$$

Where $h_{[CLS]}$ and $h_{[SEP]}$ refer to the hidden states corresponding to the $[CLS]$ and $[SEP]$ tokens, respectively. h_i^t represents the hidden state of the input. For the final event detection prediction, a fully connected layer is applied:

$$P_e = softmax(W \cdot h_{input} + b), \tag{4}$$

Where W and b are trainable parameters, $softmax$ acts as the non-linear activation function, and P_e denotes the model's predicted probability distribution, specifically represented as $P_e = [P_{e1}, P_{e2}, \cdots, P_{en}]$.

3.2.2. Enhanced Legal Event Knowledge Base

The extracted event information includes the annotation of event information for each token in the case text, clearly indicating which tokens belong to specific event categories. This greatly enriches the semantic layers of the case text. The case database $\mathcal{W}$ is represented as follows:

$$\mathcal{W} = \{(c_i, T_i, E_i) \mid c_i \in C, T_i \subseteq c_i, E_i \subseteq E\}, \tag{5}$$

Where c_i is the i-th case in the case text collection C, T_i is the set of tokens in the corresponding text, and E_i is the set of event type information extracted from the tokens in T_i, which constitutes a set E containing all event information in C.

To build an enhanced legal event knowledge base containing key case elements, we enrich the case text with detailed incident notes in addition to the original information. This increases the information density and usability of queries, allowing subsequent retrieval models to more accurately match similar cases based on key event information. The semantic information of related case pairs is thus enriched, improving the accuracy and efficiency of retrieval.

3.3. Legal Event-Centric Similar Case Retrieval

In the retrieval stage, the LeCaRD dataset identifies relevant cases by marking those with similar facts to a query case, as assessed by legal professionals. Legal event information is essential in defining the connection between a case and its precedent. If a case has a corresponding precedent, it is likely considered relevant due to similarities in facts, evidence, and judgments.

Civil law cases often have a relatively fixed writing style. For example, the LeCaRD dataset, typically consists of four basic parts: case title, basic case information, judgment analysis process, and judgment result. The case title provides fundamental information,the judgment process analysis explains the reasoning behind the decision, and the judgment result is the final judgment of the case. This clear and concise legal case structure reflects the rigor of the law while posing challenges due to the lengthy text. To prevent redundancy, we chose to retrieve the first two parts of the case, as these contain both basic and detailed information about the case.

We assess whether the legal facts and legal event information in the two cases are logically similar. To achieve this, we employ an encoder to extract both legal facts and event information. Given the strong semantic capabilities of pre-trained Chinese models, we implement a BERT-based paragraph aggregation framework. First, the lengthy text is divided into smaller, manageable paragraphs for BERT processing, after which the semantic encodings of the query and candidate paragraphs are obtained using the pre-trained BERT model. The aggregation module then merges the legal facts and event information, representing the encoded relationship between the case pair. Specifically, for the query case $\mathcal{S} = \{s_1, s_2, \cdots, s_m\}$ and the candidate case $\mathcal{C} = \{c_1, c_2, \cdots, c_n\}$:

$$\mathcal{S}^* = \mathcal{S} + E^q, \tag{6}$$

$$\mathcal{C}^* = \mathcal{C} + E^d. \tag{7}$$

Among them, $\mathcal{S}^*$ and $\mathcal{C}^*$ represent the query case and candidate cases aggregated with legal event information, respectively. $E^q = \{e_{q1}, e_{q2}, \cdots, e_{qm}\}$ and $E^d = \{e_{d1}, e_{d2}, \cdots, e_{dn}\}$ represent the encoded case information with event information indexed from the enhanced case database. Therefore, the input to the model can be expressed as:

$$input = [CLS]\mathcal{S}^*[SEP]\mathcal{C}^*[SEP]. \tag{8}$$

After constructing matching case pairs that combine the query and candidate cases with legal event information, we further enhance the model's performance by employing supervised contrastive learning (Khosla et al., 2020). This method trains the retrieval model using a contrastive loss function, $\mathcal{L}_{cl}$, designed to bring similar case pairs closer together within the embedding space. In this framework, positive and negative samples are differentiated by their labels: samples sharing the same label are treated as positive pairs, while those with different labels are viewed as negative pairs. The supervised contrastive loss operates on low-dimensional embeddings, working to pull positive samples together and push negative samples apart. It is formally expressed as follows:

$$\mathcal{L}_{cl} = -\sum_{n \in N} \frac{1}{|B(n)|} \sum_{b \in B(n)} \log \frac{\exp(x_n \cdot x_b / \tau)}{\sum_{a \in A(n)} \exp(x_n \cdot x_a / \tau)}, \tag{9}$$

Where x_l denotes the low-dimensional embedding, and τ is the temperature parameter that controls the smoothing of similarity scores. The set $A(n) \equiv I \backslash \{n\}$ includes all samples except the anchor sample n, ensuring that self-similarity is not considered in the loss calculation. The set $B(n)$ consists of all samples in $A(i)$ that share the same label $\overline{y}$ as the anchor sample, and these samples are treated as positive pairs.

The purpose of $\mathcal{L}_{cl}$ is to learn embeddings that position positive sample pairs (i.e., the query case and its matching candidate) near each other in the embedding space, while separating negative pairs (query and non-matching candidate cases). During training, we apply the standard cross-entropy loss, $\mathcal{L}_{ce}$. Hence, the final loss $\mathcal{L}$ is represented as:

$$\mathcal{L} = (1 - \lambda) * \mathcal{L}_{ce} + \lambda * \mathcal{L}_{cl}, \tag{10}$$

Where λ is a loss weight parameter.

The legal events in the candidate cases support the reasoning and judgment process of the query case. Based on our experimental findings, we hypothesize that this legal event information aids in summarizing the case documents by focusing on the case's core content (such as fact-based information), filtering out irrelevant text and enabling the model to prioritize the essential details.

4. Experiment

4.1. Experiment Setup

4.1.1. Datasets

To assess the effectiveness of RAEvent in the SCR task, we utilize the LeCaRD (Ma et al., 2021) dataset. For the auxiliary task, we leverage the LEVEN (Yao et al., 2022) dataset to train the legal event extraction model, which enriches RAEvent's legal event knowledge base and enhances its performance in subsequent retrieval tasks. Table 1 provides an overview of the statistics for both datasets, with further details outlined below:

As a Chinese legal case retrieval dataset, LeCaRD is built on official documents[1]. It includes 107 query cases and 10,700 candidate cases, primarily drawn from criminal law. Each query has approximately 100 candidate cases, and the model ranks them according to their similarity to the query text, with higher similarity resulting in a better rank.

The LEVEN dataset is the largest legal text event extraction dataset in terms of event types and data volume. It covers 108 types of legal events and annotated by seasoned legal professionals, spanning common categories such as fraud, violence, and accidents. While the LeCaRD dataset mainly consists of criminal cases, LEVEN covers high-frequency legal events in the criminal domain well.

4.1.2. Implementation Details

All experiments in this study were conducted on an RTX 8000 GPU with 48 GB of memory using the PyTorch[2] framework, built on Transformers[3]. We employed five-fold cross-validation to evaluate our approach, with results measured using normalized discounted cumulative gain (NDCG), precision (P), Macro F1 (F_{macro}), and mean average precision (MAP). P@K assesses the precision of the system's top-K search results, while NDCG@K measures the relevance and ranking quality, normalized against the ideal ranking order. In all these metrics, higher values indicate superior performance.

For consistent comparison, we used the Chinese criminal pre-trained model (Crime-BERT[4]) for methods requiring pre-trained model encoding, considering that the dataset is

1. https://wenshu.court.gov.cn/
2. https://pytorch.org/
3. https://github.com/huggingface/transformers
4. https://thunlp.oss-cn-qingdao.aliyuncs.com/bert/xs.zip

Table 1: **Statistics of datasets**

Dataset	Statistic	Number
LeCaRD	Query	107
	Candidate document	10700
	Average length of query	445
	Average relevant case per query	10.33
	Average length per case document	8275
LEVEN	Event	150997
	Event type	108
	Average length	496
	Total document	8116

from the criminal domain. In the baseline experiments, for studies with available open-source code, the replicated workflows were conducted in line with the methodologies described in the original papers, and a uniform evaluation was performed on the files containing the retrieval results. For the reproduction of studies associated with different pre-trained models, a uniform set of hyperparameters was employed. The experiments were carried out using the transformer library within the PyTorch framework, with the majority of hyperparameters set to the default values of the framework. All experiments were conducted with identical data input preprocessing and data distribution. The model's output scores were those produced by the final fully connected layer. The training objectives were optimized using cross-entropy loss. Except for Lawformer (Xiao et al., 2021), which has an input length of 1800, all other models have an input length of 512, with a learning rate set to 1e-5.

4.2. Baseline Methods

We selected the following baseline models for comparison with our proposed model: traditional bag-of-words retrieval methods, such as BM25 (Robertson and Walker, 1994); legal field-based pre-trained models, including StructBERT-law[5] and Lawformer (Xiao et al., 2021); and neural network-based retrieval models, including RetroMAE (Xiao et al., 2022), NS-LCR (Sun et al., 2024), SAILER (Li et al., 2023b), and PromptCase (Tang et al., 2023). The following is a detailed introduction of each model:

BM25: BM25 (Robertson and Walker, 1994) is a well-established information retrieval algorithm. It factors in inverse document frequency, average document length, and the term frequency, maintaining its status as a robust baseline in the field of retrieval.

StructBERT-law: StructBERT-law is a Chinese legal pre-trained model based on the StructBERT pre-trained model, using 400GB of judicial corpus for pre-training. The training corpus includes texts from judgments, laws and regulations, court transcripts, legal Q&A, and legal encyclopedias.

Lawformer: Lawformer (Xiao et al., 2021) is based on the Longformer architecture and is designed to handle large-scale legal case documents. Due to its ability to process extended

5. https://modelscope.cn/models/iic/nlp_structbert_backbone_base_law/summary

Table 2: The overall performance of the experimental results, where P, F_{macro}, MAP and NDCG represent precision, Macro F1, mean average precision, and normalized discounted cumulative gain respectively. P@K measures the precision of a system's top-K search results, while NDCG@K evaluates the relevance and ranking quality of those results, normalizing for ideal ordering.

Methods	**P@5**	$\mathbf{F_{macro}}$	**MAP**	**NDCG@5**	**NDCG@30**
BM25	38.87	21.89	46.40	69.17	83.76
StructBERT-law	36.36	21.66	42.59	75.45	89.68
Lawformer	40.26	24.23	45.81	80.11	91.31
RetroMAE	46.36	24.77	56.33	77.39	89.35
NS-LCR	47.03	27.08	50.14	79.01	91.17
SAILER	47.63	28.40	53.10	85.18	93.44
PromptCase	51.59	30.75	59.14	86.24	93.13
RAEvent	**53.88**	**32.35**	**60.37**	**87.21**	**94.34**

texts, we set the model's maximum input length to 1800 tokens to optimize its performance on long documents.

RetroMAE: RetroMAE (Xiao et al., 2022) introduces a masked autoencoder framework for pre-training language models specifically for retrieval tasks. The authors proposed enhanced decoding to fully utilize pre-training data, improving the model's performance in zero-shot dense retrieval and supervised dense retrieval.

NS-LCR: NS-LCR (Sun et al., 2024) incorporates explicit legal case matching through the learning of case-level and law-level logical rules, which are then integrated into the retrieval process using a neural symbolic approach.

SAILER: SAILER (Li et al., 2023b) is a structure-aware pre-trained model focused on LCR. It captures long-range dependencies across various structures and is designed to detect critical legal elements within cases.

PromptCase: PromptCase (Tang et al., 2023) is a novel legal case retrieval framework based on prompt learning. It identifies key features in cases that facilitate legal case retrieval and designs a prompt-based encoding scheme. Finally, it uses a language model for effective encoding to complete the case retrieval task.

4.3. Main Results

In this subsection, we demonstrate the performance of RAEvent. As shown in Table 2, our method performs better across various evaluation metrics. With rich legal event information, RAEvent improves the extraction of semantically similar features between case pairs through the use of supervised contrastive learning, leading to a significant enhancement in the retrieval performance for query cases.

Regarding traditional retrieval models like BM25 (Robertson and Walker, 1994), although they lack the semantic understanding capabilities of neural network-based models, they still perform remarkably well, particularly in the MAP metric, achieving a score of 46.40. Legal domain pre-trained models, by leveraging their rich legal semantic knowledge,

Table 3: Ablation experiment results. These demonstrate the influence of both the supervised contrastive learning module and the legal event information component on the performance of the RAEvent framework.

Methods	P@5	$\mathbf{F_{macro}}$	MAP	NDCG@5	NDCG@30
RAEvent	**53.88**	**32.35**	**60.37**	**87.21**	**94.34**
w/o event	43.71	24.59	48.68	76.80	89.42
w/o contrastive	50.69	30.14	56.90	85.51	93.63
w/o {contrastive & event}	42.55	22.92	48.20	76.62	89.47

demonstrate improved retrieval performance. Specifically, StructBERT-law, a BERT model tailored for the Chinese legal domain, showcases the advantages of pre-trained legal knowledge over traditional methods. Lawformer (Xiao et al., 2021), by overcoming the limitations of traditional input, achieves better results through access to more comprehensive textual information, particularly in the NDCG@30 metric, where it reaches 91.31. This suggests that obtaining more information from longer texts improves the ranking quality of relevant cases in the retrieval results list. For neural network-based retrieval models, enhancing the extraction of similar case features through various methods to improve retrieval performance is a promising direction.

SAILER (Li et al., 2023b), a pre-trained model specifically designed for LCR tasks, understands long-range dependencies between different structures and is sensitive to key legal elements within cases. It exhibits higher recognition capacity for case structures, such as the case name, fact section, and judgment section, leading to strong performance in similar case retrieval. Our method focuses on extracting finer-grained event information from key parts of the cases (e.g., case name, basic case information) and using this event-centered information as auxiliary information to complete downstream retrieval tasks, ultimately achieving better performance across various metrics. PromptCase (Tang et al., 2023) excels by correctly rephrasing case inputs using prompt learning, allowing the model to better understand and represent cases, which leads to outstanding performance across all metrics, especially in MAP, indicating good overall retrieval accuracy. While these baselines perform well across various metrics, they still overlook the role of finer-grained critical legal event information in cases. Our experimental results demonstrate that legal event information is crucial in establishing the relationship between query cases and candidate cases, serving as a basis for argumentation and judgment in similar cases.

4.4. Ablation Study

Ablation studies were performed to assess the contributions of RAEvent's two core components: legal event information and contrastive learning. As shown in Table 3, combining legal event information with contrastive learning significantly improves performance. Using only legal event information also greatly enhances retrieval performance. When either event information or contrastive learning methods are removed, performance declines, indicating that both components are essential to our framework.

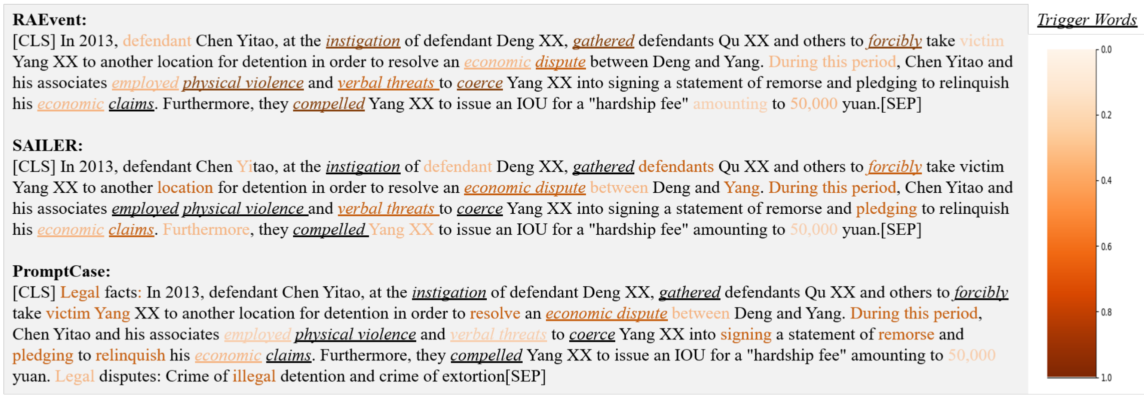

Figure 4: Attention visualization of model output. We have visualized the model's attention distribution for three methods separately, and the underlined words are the trigger words marked in the case. The top section of the visualization represents our proposed method. The case shown in the figure is a real Chinese case translated into English.

4.5. Visualisation Analysis

We provided attention visualization comparisons of case output results for several methods, as shown in Figure 4. The text was translated from real Chinese cases. Among them, PromptCase adds prompt templates as model input, and we visualized the model output according to the prompt templates in the original paper. Our method focuses more on the legal event information in the cases, such as key events like "instigate", "coerce", and "physical violence" which influence legal judgment and responsibility determination. Other methods lack this focus and are more susceptible to random noise, affecting their final performance.

5. Conclusion

We proposed RAEvent, a retrieval-augmented contrastive framework for legal case retrieval based on legal event information. Our goal is to utilize key event elements in cases to assist downstream similar case retrieval tasks. First, we selected the event extraction task to construct an enhanced case event information database and use contrastive learning techniques to enhance the extraction of similar features between case pairs. The experimental outcomes confirm the efficacy of RAEvent and offer a fresh perspective for research in this domain. Moving forward, we advocate for the continued use of case event information in downstream legal tasks and encourage further exploration of its interpretability within these tasks.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (No. 62376062), the Ministry of Education of Humanities and Social Science Project (No. 23YJAZH220, No. 24YJAZH244), the Philosophy and Social Sciences 14th Five-Year Plan Project of Guangdong Province (No. GD23CTS03), and the Guangdong Basic and Applied Basic Research Foundation of China (No. 2023A1515012718).

References

Sophia Althammer, Arian Askari, Suzan Verberne, and Allan Hanbury. Dossier@ coliee 2021: Leveraging dense retrieval and summarization-based re-ranking for case law retrieval. *arXiv preprint arXiv:2108.03937*, 2021.

Ilias Chalkidis, Manos Fergadiotis, Prodromos Malakasiotis, Nikolaos Aletras, and Ion Androutsopoulos. Legal-bert: The muppets straight out of law school. *arXiv preprint arXiv:2010.02559*, 2020.

Yanguang Chen, Yuanyuan Sun, Zhihao Yang, and Hongfei Lin. Joint entity and relation extraction for legal documents with legal feature enhancement. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 1561–1571, 2020.

Zhuyun Dai and Jamie Callan. Deeper text understanding for ir with contextual neural language modeling. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 985–988, 2019.

Yi Feng, Chuanyi Li, and Vincent Ng. Legal judgment prediction via event extraction with constraints. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 648–664, 2022.

Yu Hong, Jianfeng Zhang, Bin Ma, Jianmin Yao, Guodong Zhou, and Qiaoming Zhu. Using cross-entity inference to improve event extraction. In *Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies*, pages 1127–1136, 2011.

Heng Ji and Ralph Grishman. Refining event extraction through cross-document inference. In *Proceedings of ACL-08: Hlt*, pages 254–262, 2008.

Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

Canjia Li, Andrew Yates, Sean MacAvaney, Ben He, and Yingfei Sun. Parade: Passage representation aggregation fordocument reranking. *ACM Transactions on Information Systems*, 42(2):1–26, 2023a.

Haitao Li, Qingyao Ai, Jia Chen, Qian Dong, Yueyue Wu, Yiqun Liu, Chong Chen, and Qi Tian. 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*, pages 1035–1044, 2023b.

Qingquan Li, Qifan Zhang, Junjie Yao, and Yingjie Zhang. Event extraction for criminal legal text. In *2020 IEEE International Conference on Knowledge Graph (ICKG)*, pages 573–580. IEEE, 2020.

B Liu, S Li, L Ye, and H Zhang. Similar case recommendation algorithm based on legal elements. *Intelligent Computer and Applications*, 11(6):1–4, 2021.

Yixiao Ma, Yunqiu Shao, Yueyue Wu, Yiqun Liu, Ruizhe Zhang, Min Zhang, and Shaoping Ma. Lecard: a legal case retrieval dataset for chinese law system. In *Proceedings of the 44th international ACM SIGIR conference on research and development in information retrieval*, pages 2342–2348, 2021.

Jay M Ponte and W Bruce Croft. A language modeling approach to information retrieval. In *ACM SIGIR Forum*, volume 51, pages 202–208. ACM New York, NY, USA, 2017.

Juliano Rabelo, Mi-Young Kim, Randy Goebel, Masaharu Yoshioka, Yoshinobu Kano, and Ken Satoh. A summary of the coliee 2019 competition. In *New Frontiers in Artificial Intelligence: JSAI-isAI International Workshops, JURISIN, AI-Biz, LENLS, Kansei-AI, Yokohama, Japan, November 10–12, 2019, Revised Selected Papers 10*, pages 34–49. Springer, 2020.

Stephen E Robertson and Steve Walker. Some simple effective approximations to the 2-poisson model for probabilistic weighted retrieval. In *SIGIR'94: Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, organised by Dublin City University*, pages 232–241. Springer, 1994.

Yunqiu Shao, Jiaxin Mao, Yiqun Liu, Weizhi Ma, Ken Satoh, Min Zhang, and Shaoping Ma. Bert-pli: Modeling paragraph-level interactions for legal case retrieval. In *IJCAI*, pages 3501–3507, 2020.

Shirong Shen, Guilin Qi, Zhen Li, Sheng Bi, and Lusheng Wang. Hierarchical chinese legal event extraction via pedal attention mechanism. In *Proceedings of the 28th international conference on computational linguistics*, pages 100–113, 2020.

Karen Sparck Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of documentation*, 28(1):11–21, 1972.

Zhongxiang Sun, Kepu Zhang, Weijie Yu, Haoyu Wang, and Jun Xu. Logic rules as explanations for legal case retrieval. *arXiv preprint arXiv:2403.01457*, 2024.

Yanran Tang, Ruihong Qiu, and Xue Li. Prompt-based effective input reformulation for legal case retrieval. In *Australasian Database Conference*, pages 87–100. Springer, 2023.

Chaojun Xiao, Xueyu Hu, Zhiyuan Liu, Cunchao Tu, and Maosong Sun. Lawformer: A pre-trained language model for chinese legal long documents. *AI Open*, 2:79–84, 2021.

Shitao Xiao, Zheng Liu, Yingxia Shao, and Zhao Cao. Retromae: Pre-training retrieval-oriented language models via masked auto-encoder. *arXiv preprint arXiv:2205.12035*, 2022.

Feng Yao, Chaojun Xiao, Xiaozhi Wang, Zhiyuan Liu, Lei Hou, Cunchao Tu, Juanzi Li, Yun Liu, Weixing Shen, and Maosong Sun. Leven: A large-scale chinese legal event detection dataset. *arXiv preprint arXiv:2203.08556*, 2022.

Weijie Yu, Zhongxiang Sun, Jun Xu, Zhenhua Dong, Xu Chen, Hongteng Xu, and Ji-Rong Wen. Explainable legal case matching via inverse optimal transport-based rationale extraction. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval*, pages 657–668, 2022.

Haoxi Zhong, Chaojun Xiao, Cunchao Tu, Tianyang Zhang, Zhiyuan Liu, and Maosong Sun. How does nlp benefit legal system: A summary of legal artificial intelligence. *arXiv preprint arXiv:2004.12158*, 2020.