

PAPER

Integrating Event Elements for Chinese-Vietnamese Cross-Lingual Event Retrieval

Yuxin HUANG^{†,††}, Yuanlin YANG^{†,††}, Enchang ZHU^{†,††}, Yin LIANG^{†,††}, and Yantuan XIAN^{†a)}, *Nonmembers*

SUMMARY Chinese-Vietnamese cross-lingual event retrieval aims to retrieve the Vietnamese sentence describing the same event as a given Chinese query sentence from a set of Vietnamese sentences. Existing mainstream cross-lingual event retrieval methods rely on extracting textual representations from query texts and calculating their similarity with textual representations in other language candidate sets. However, these methods ignore the difference in event elements present during Chinese-Vietnamese cross-language retrieval. Consequently, sentences with similar meanings but different event elements may be incorrectly considered to describe the same event. To address this problem, we propose a cross-lingual retrieval method that integrates event elements. We introduce event elements as an additional supervisory signal, where we calculate the semantic similarity of event elements in two sentences using an attention mechanism to determine the attention score of the event elements. This allows us to establish a one-to-one correspondence between event elements in the text. Additionally, we leverage the multilingual pre-trained language model fine-tuned based on contrastive learning to obtain cross-language sentence representation to calculate the semantic similarity of the sentence texts. By combining these two approaches, we obtain the final text similarity score. Experimental results demonstrate that our proposed method achieves higher retrieval accuracy than the baseline model.

key words: *Chinese-Vietnamese cross-lingual event retrieval, pre-trained language model, event elements, attention mechanism*

1. Introduction

Cross-language event retrieval, a subtask of cross-language information retrieval [1], involves retrieving information about reports of the same event described in different languages. This retrieval process aids in obtaining comprehensive and holistic descriptions of events by considering multiple languages. Cross-language event retrieval helps to obtain descriptions of the same event in different languages to obtain a more comprehensive and holistic description of the event. Specifically, Chinese-Vietnamese cross-lingual event retrieval refers to inputting a Chinese query sentence and retrieving Vietnamese sentences that describe the same event as the query from a set of candidate Vietnamese sentences. As shown in Fig. 1, using query sentence “阿根廷时隔36年再次夺冠,世界杯迎来新三星球队 (Argentina wins the World Cup again after 36 years and welcomes new three-star team),” then retrieve the Vietnamese sentence de-

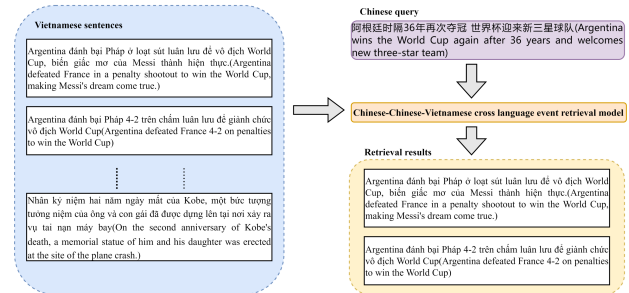


Fig. 1 Example of Chinese-Vietnamese cross-language event retrieval.

scribing the same event from the set of candidate news sentences, “Argentina đánh bại Pháp ở loạt sút luân lưu để vô địch World Cup, biến giấc mơ của Messi thành hiện thực (Argentina defeated France in a penalty shootout to win the World Cup, making Messi’s dream come true.)” and “Argentina đánh bại Pháp 4-2 trên chấm phạt đền để vô địch World Cup (Argentina defeated France 4-2 on penalties to win the World Cup).”

The prevailing research approach for cross-lingual event retrieval tasks involves transforming them into cross-lingual event matching tasks [2]. Cross-lingual event retrieval can currently be categorized into two primary approaches [3]: machine translation-based and cross-lingual representation-based. The machine translation-based approach first translates both the query statement and the text to be retrieved into the same language, then performs monolingual retrieval [4]–[6]. This approach achieves better results in resource-rich languages, however for low-resource languages, due to the scarcity of translation data, there are some entity translation errors during translation, which further leads to weak performance in Chinese-Vietnamese cross-lingual event retrieval. The approach based on cross-language representation employs a pre-trained model for text representation to extract vectors that correspond to texts written in different languages [7], [8]. These vectors are employed for similarity calculation to enable cross-lingual event retrieval, such as Devlin [9] and Chidambaram [10]. However, existing cross-lingual event retrieval methods fail to adequately enable the model to effectively compare the event elements present in the text, which comprise crucial information about the events. As a result, the model erroneously classifies a significant portion of sentences with similar sentence meanings but different event elements as describing the same event element. As shown in Fig. 2, sentence 1: “马来西 亚雪 兰莪 州的 洪灾 导致 两个 地区 的 大约 200 名 居民 被 疏散。(Flooding

Manuscript received March 6, 2024.

Manuscript revised April 21, 2024.

Manuscript publicized June 4, 2024.

[†]Yunnan Key Laboratory of Artificial Intelligence, Kunming University of Science and Technology, Kunming 650500, China.

^{††}School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China.

a) E-mail: xianyt@kust.edu.cn (Corresponding author)

DOI: 10.1587/transinf.2024EDP7055

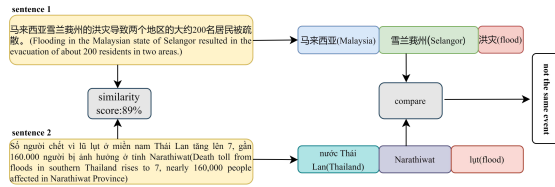


Fig. 2 An example of whether the event element can determine whether the text is similar.

in the Malaysian state of Selangor resulted in the evacuation of about 200 residents in two areas.)” and sentence 2: “Số người chết vì lũ lụt ở miền nam Thái Lan tăng lên 7, gần 160.000 người bị ảnh hưởng ở tỉnh Narathiwat (Death toll from floods in southern Thailand rises to 7, nearly 160,000 people affected in Narathiwat Province)” represent distinct events, the high similarity score of 89% between these two sentences would mistake to consider both news texts as describing the same event. However, by comparing the semantic similarity between the event elements in the two sentences, it is possible to distinguish the text of sentences with similar sentence meanings but different events. As shown in Fig. 2, sentence 1 contains event elements such as Malaysia, Selangor, and flood, and sentence 2 contains nước Thái Lan (Thailand), Narathiwat, lụt (flood). By comparing the semantics of the event elements based on the similarity, it can be judged that sentence 1 and sentence 2 belong to different events.

To address the above problem, we propose a Chinese-Vietnamese cross-lingual event retrieval method that integrates event elements to improve event retrieval accuracy. This method fully mines the event element information in the sentence text, matches the event elements by calculating attention parameters, calculates the similarity between event elements, and combines the sentence text similarity to perform cross-language event retrieval, allowing the model to compare query sentences and retrieve the similarity between event element knowledge between sentences instead of comparing the similarity between sentences.

We conduct extensive experiments on the Chinese-Vietnamese cross-lingual event retrieval data constructed in this paper. Experimental results prove that our proposed method can effectively utilize event elements in sentence text and achieve better performance compared with multiple baselines. We summarize our main work as follows:

1. We propose a novel method of integrating event elements for cross-language event retrieval between Chinese and Vietnamese.
2. We constructed a Chinese-Vietnamese cross-language event retrieval dataset.
3. We demonstrate the effectiveness of the proposed method based on experiments.

2. Related Work

Currently, cross-lingual event retrieval methods can be categorized into two types based on their training approaches: traditional text-matching and deep learning representation-

based retrieval methods.

The traditional approach to text matching retrieval relies on extracting features from texts and utilizing them to compute similarity. These features include TF-IDF, BM25 [11], and lexical information, which are extracted at various levels within the texts. Subsequently, the extracted features are employed to calculate similarity scores between the texts. For instance, Singh et al. [12] extracted key terms from the document using TF-IDF and assigned weights to these terms, which were then represented in the vector space to measure the correlation between the document and the query. Dragoni et al. [13] introduced a vector space model that represents documents and queries based on concepts instead of terms, utilizing WordNet as a lightweight ontology. This representation mitigates information overlap when compared to traditional semantic expansion techniques. However, these approaches are sensitive to the semantic nature of the text, potentially leading to suboptimal document representations.

The deep learning-based approach utilizes deep learning models to represent matching texts and query texts into a unified vector space. Text matching and retrieval are achieved through the calculation of similarity between the characterized text vector and query vector. For instance, Hu et al. [6] proposed the ARC-II model, which utilizes a convolutional neural network (CNN) to extract text features from the query and text, and subsequently generates word vectors. Subsequently, the similarity between the query and the text is computed by word vectors. Similarly, Paul Neculoiu et al. [14] employ a bidirectional LSTM-based Siamese network structure to project variable-length strings into a fixed-dimensional embedding space and text similarity computation. Wang et al. [2] introduce a multi-perspective interactive matching Siamese network model that incorporates multiple perspectives to compute text similarity, enhancing the utilization of text information. Omar Khattab et al. [15] propose the colBERT model, which employs the pre-trained language model BERT to represent the query and retrieval texts. An interaction step is utilized to model the similarity between the query and retrieval texts, followed by similarity computation. This method effectively leverages the expressive capacity of pre-trained language models. Deep learning-based text-matching models have demonstrated effectiveness in tackling the challenge of semantic understanding in textual content. However, their applicability to event retrieval tasks is limited because they fail to adequately consider the event elements embedded within the text. Furthermore, deep learning text matching models can identify semantic correlations between words in the text, and utilize structural characteristics, to enhance the text matching process. This critical process is predominantly accomplished through sentence representation methods, where the effectiveness of sentence representation directly impacts the accuracy of sentence matching. Therefore, in the context of cross-lingual event retrieval between Chinese and Vietnamese, we not only consider the representation of text sentences but also emphasize the similarity between event elements present in the text. To enhance the model’s performance in event retrieval tasks, we introduce

event elements as an additional supervisory signal and reinforce the model’s understanding of event elements through enhanced pre-training language model.

3. Model

Based on the idea of integrating event elements into cross-language event retrieval, we propose a Chinese cross-language event retrieval model that integrates event elements. The specific model shown in Fig. 3, consists of three main modules: event element matching similarity calculation module, a similarity calculation model based on sentence Representation, and a match calculation module.

The event element matching module extracts annotated event elements from input Chinese-Vietnamese sentence pairs and feeds each event element into a cross-lingual pre-trained model, generating cross-lingual word embeddings for each vocabulary item. To obtain aligned event element pairs, we introduce an event element matching layer that employs attention calculations to derive matched event element pairs from the two sentences. To capture the semantic relations between the two sentences at different levels, we extract word representations for event elements and sentence representation vectors to compute the matching degree. The sentence

representation vector extraction module finetunes a multi-lingual pre-trained language model specifically to acquire Chinese-Vietnamese sentence representation vectors. It calculates the similarity between bilingual sentence representation vectors, combines the resulting similarity score with the associated score computed from the event elements, and ranks all pre-selected sentences based on the obtained similarity score. This comprehensive process ultimately generates the final retrieval results.

3.1 Event Elements Matching Similarity Calculation Module

The module receives Chinese text sentences $P = \{p_1, p_2, \dots, p_k\}$ and Vietnamese sentences $Q = \{q_1, q_2, \dots, q_j\}$ tagged with event elements where k and j represent the number of words contained in Chinese sentences and Vietnamese sentences, respectively. The attention mechanism computes the significance of each event element word, p_k and q_j extracted from the Chinese sentence P and Q , to form pairs of event elements for matching purposes. Once the event element pairs are obtained, the prediction layer predicts the relationship between these two sentences based

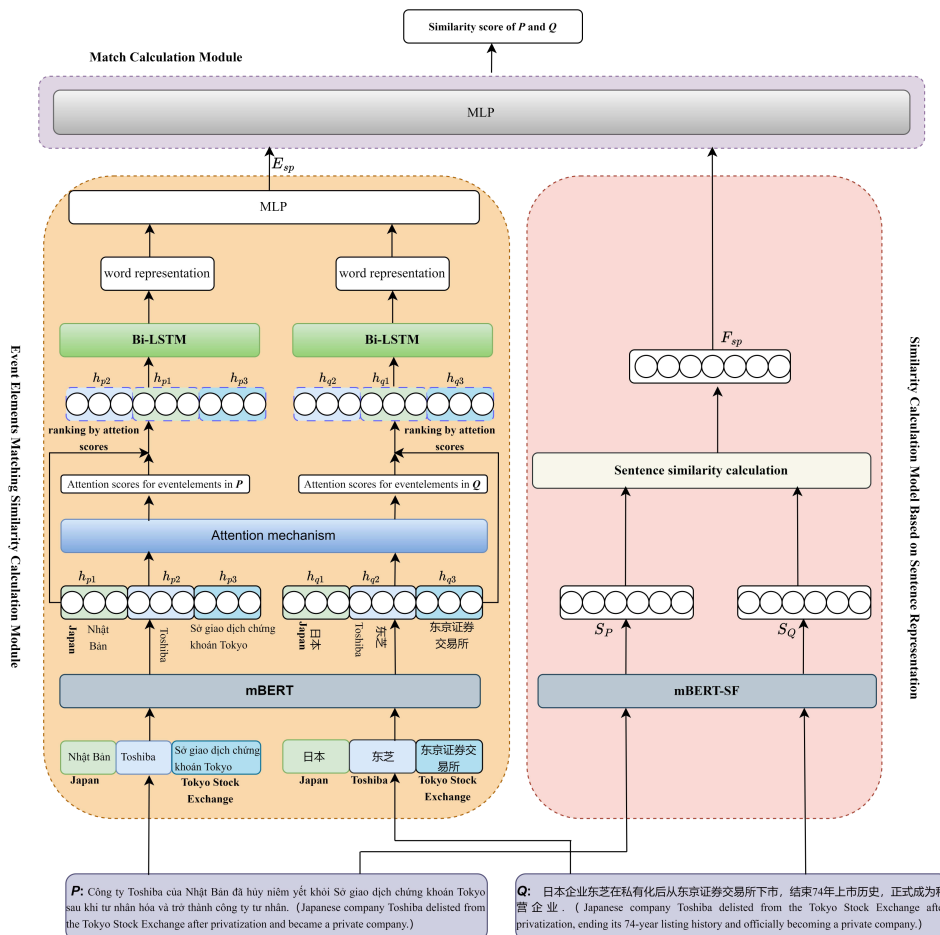


Fig. 3 A Chinese-Vietnamese cross-language event retrieval model integrating event elements.

on the similarity of the event element pairs. This module consists of a text word representation extraction layer, text event element extraction layer, and similarity prediction layer. The text word representation extraction layer mainly extracts the tagged event elements from the input sentences P and Q . Then, by inputting all the words into the multilingual word representation model mBERT [7], it obtains the word representation vectors of the event elements, denoted as $E_p = \{h_{p1}, h_{p2}, \dots, h_{pm}\}$ and $E_q = \{h_{q1}, h_{q2}, \dots, h_{qn}\}$ of the event elements where n and m denote the number of event elements contained in Chinese sentences and Vietnamese sentences, respectively.

The text word representation extraction layer mainly extracts the tagged event elements from the input sentences P and Q . Then it obtains the sum after inputting all words into the multilingual word representation model mBERT to obtain the word representation vectors $E_p = \{h_{p1}, h_{p2}, \dots, h_{pm}\}$ and $E_q = \{h_{q1}, h_{q2}, \dots, h_{qn}\}$ of the event elements where n and m are the numbers of event elements contained in Chinese sentences and Vietnamese sentences, respectively. The text word representation extraction layer extracts event elements from input sentences P and Q . Then, the sum of all words is input into the multilingual word representation model mBERT, enabling the extraction of word representation vectors $E_p = \{h_{p1}, h_{p2}, \dots, h_{pm}\}$ and $E_q = \{h_{q1}, h_{q2}, \dots, h_{qn}\}$ for their respective event elements. Here, and indicate the number of Chinese and Vietnamese event elements, respectively. In this paper, we consider that event elements contain the following characteristics when the relationships are similar: 1. have rich semantic representations; 2. are very important in both P and Q ; 3. have similar semantic representations in both P and Q . For these three characteristics this paper calculates the semantic representation of event elements, the attention score a_p for each event element in P , and the attention score a_q for each event element in Q . The specific calculation, as shown in Fig. 4.

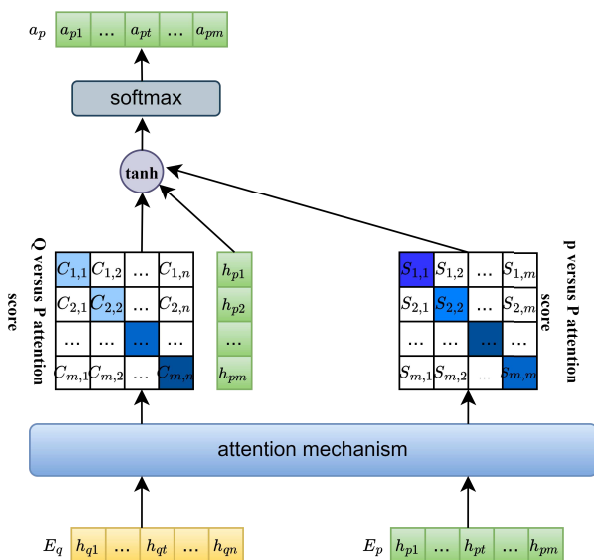


Fig. 4 Schematic diagram of text event elements matching method.

The attention score a_p of each event element in P is calculated as follows [16], by multiplying E_p and E_q matrices to get $C_{m,n}$, where $C_{m,n}$ represents the attention score between the n -th word in Q and the m -th word in P . $S_{x,y}$ is obtained by multiplying the E_p and E_p^T matrices, $S_{x,y}$ represents the attention score of the x -th word in P to the y -th word in P . Then these two values and the vocabulary representation h_p are weighted, summed and activated using the function, as shown in the formula (1).

$$m_p = \tanh(h^p h^{pt} W_{pp} + h^p h^{qt} W_{pq} + h^p W_p) W_d \quad (1)$$

$$m_q = \tanh(h^q h^{pt} W_{qq} + h^q h^{pt} W_{pq} + h^q W_q) W_d \quad (2)$$

In the formula: m_p is the attention score activated by the m words in P to the n -th word in Q through the tanh function, where W_{pp} , W_{pq} , W_{pq} and W_d are fixed parameters.

After performing a linear transformation on m_p , the softmax function is used to calculate the final attention score of each word, and the specific calculation is shown in the formula (3). In the formula, a_p represents the total attention score of each event element in P for two sentences.

$$a_p = \frac{\exp(m^p)}{\sum_{t=1}^N m_t^p} \quad (3)$$

$$a_q = \frac{\exp(m^q)}{\sum_{v=1}^N m_v^q} \quad (4)$$

The attention score for each event element in Q follows the same method as the attention score for each event element in P , as demonstrated in formulas (2) and (4). In formula (2), W_{qq} , W_{qp} and W_d are fixed parameters. Then, the sequences of event elements in sentences P and Q are reorganized based on their corresponding attention scores, resulting in the construction of Chinese-Vietnamese cross-lingual event element word pairs.

To enhance the model's ability to capture the relationship between two sentences using the proposed event elements, we adopt the sequence of event elements in each sentence to represent the sentence representation. By arranging the proposed event elements based on their attention scores, we merge them into a new sequence and encode the sequences using a Bidirectional Long Short-Term Memory (BiLSTM) network. In the final step, we utilize the vectors obtained from the last time step in the BiLSTM network to construct the crucial semantic feature vectors for sentence P and sentence Q , respectively. These vectors serve as the word-level representation of the sentences.

3.2 Similarity Calculation Model Based on Sentence Representation

In order to obtain a text representation suitable for Chinese-Vietnamese cross-lingual event retrieval, we adopt the cross-lingual sentence representation model based on contrastive learning (mBERT-SF) proposed by Liang et al. [17]. This model comprises mBERT and a Siamese network's linear fine-tuning layer, and combined with contrastive learning

for training, it can effectively solve the problem of poor semantic alignment of cross-language sentence embeddings in Chinese-Vietnamese contexts in multi-language pre-training models due to the scarcity of Chinese-Vietnamese sentence levels, and can obtain better Chinese-Vietnamese text representations. Chinese and Vietnamese sentences are independently inputted into the fine-tuned mBERT-SF model to yield cross-lingual sentence representations for the two texts. Subsequently, the resulting sentence representation vectors are utilized as the final input, denoted as S . The specific formula (5), (6) is as follows:

$$S_p = \text{mBERT-SF}(P) \quad (5)$$

$$S_q = \text{mBERT-SF}(Q) \quad (6)$$

After obtaining the cross-lingual sentence representations for the two texts, we calculate their similarity using the Euclidean distance. This computation yields the final score, denoted as F_{sp} , which serves as the output for the sentence-level similarity calculation task. The specific formula is as follows:

$$F_{sp} = \text{Euclidean}(S_p, S_q) \quad (7)$$

Where Euclidean represents the Euclidean distance calculation method, using this method, the Euclidean distance is computed between and in the semantic space to obtain the sentence-level similarity F_{sp} of the two texts.

3.3 Match Calculation Module

To predict the relationship between two sentences based on the sequence of event elements and the acquired sentence representation information, we employ a multi-layer perceptron (MLP) with separate inputs from word-level representation vectors and sentence-level representation vectors derived from the BiLSTM. The MLP comprises two fully connected hidden layers activated by ReLU, along with an output layer activated by softmax. We feed the interaction vectors of the K event element word pairs obtained by the event element matching layer through Bi-LSTM into K different MLPs for classification. The outputs of these MLPs are averaged to generate the predicted result for word-level similarity calculation. Additionally, the sentence-level semantic interaction vectors obtained from the sentence representation extraction layer are fed into an MLP to obtain the predicted result for sentence-level similarity calculation. We combine these two predicted results using a weighted summation to derive the final score. During model training, the cross-entropy loss function is typically employed as the optimization objective to minimize the loss.

$$\begin{aligned} \text{loss} = & \alpha \times \text{MLP}^{sp}(F_{sp}) \\ & + (1 - \alpha) \times \frac{1}{K} \sum_{k=1}^K \text{MLP}_k^{wp}(E_{spk}) \end{aligned} \quad (8)$$

We denote the contribution of the sentence-level similarity to the model as α , while $(1 - \alpha)$ represents the contribution of the word-level similarity. F_{sp} represents the

acquired sentence-level representation vector, E_{spk} represents the word-level representation vector of the extracted vocabulary. After calculating the similarity of all sentence texts in the candidate sentence text library by the module sorts all the sentences according to the similarity predicted by the model. It outputs all the candidate sentence texts with a similarity greater than 0.9.

4. Experiment

4.1 Dataset

Currently, there is no corresponding cross-lingual event retrieval dataset in the Chinese-Vietnamese language scenario. Therefore, we have constructed a Chinese-Vietnamese cross-lingual event retrieval dataset. In the construction process, we selected 20 hot-topic events of mutual concern between China and Vietnam, such as “South China Sea issue,” “Vietnamese Deputy Prime Minister Chen Luong leads delegation to attend the opening ceremony of UN Human Rights Council,” and “Chinese and Vietnamese militaries conduct 33rd joint patrol in the Gulf of Tonkin,” among others. We found that Vietnamese news websites, such as VietnamPlus, typically provide bilingual reports in Chinese and Vietnamese about these events. Hence, we utilized web crawlers to collect bilingual news headlines of these events in Chinese and Vietnamese as the data source for our dataset.

Based on the information of 20 hot events, we manually screened the news text titles crawled from news websites. Under each hot-topic event, we filtered 100 text data, forming pairs of “Chinese news headline - Vietnamese news headline” or “Vietnamese news headline - Chinese news headline”, and annotated 100 pairs of positive retrieval sentences under each event category, as well as annotated the event elements contained in the sentences. For Chinese and Vietnamese sentences, we used the Jieba tool and VnCoreNLP tool respectively to extract entities, and also utilized KeyBERT to extract keywords. Subsequently, we manually filtered the extracted entities and keywords. During data annotation, we used binary labels, where 1 indicates that the query sentence and retrieval sentence are for different events, and 0 indicates that the query sentence and retrieval sentence are for the same event. Through this process, we annotated 100 pairs of Chinese and Vietnamese sentences for the same event under each hot-topic event and used them as positive examples in the retrieval.

Finally, to maintain a balance between positive and negative instances in the dataset, we randomly selected news headline data from the remaining 19 hot events as negative examples. We constructed 100 pairs of negative examples under each hot event category, forming a dataset of 4000 pairs of news event retrievals. The final data example is shown in Table 1. The dataset is divided into a training set and a test set, containing a total of 9,362 events elements, where the training set size is 4000 and the test set size is 200, as shown in Table 2.

Table 1 Examples of parallel and non-parallel Chinese-Vietnamese news sentences in the dataset.

| Chinese sentence | Event elements | Vietnamese sentences | Event elements | Label |
|---|--|--|---|-------|
| 马来西亚雪兰莪州的洪灾导致两个地区的大约200名居民被疏散。(Flooding in the Malaysian state of Selangor resulted in the evacuation of about 200 residents in two areas.) | 马来西亚 (Malaysia), 雪兰莪州 (Selangor), 洪灾 (flood) | Số người chết vì lũ lụt ở miền nam Thái Lan tăng lên 7, gần 160.000 người bị ảnh hưởng ở tỉnh Narathiwat. (Death toll from floods in southern Thailand rises to 7, nearly 160,000 people affected in Narathiwat Province) | nước Thái Lan (Thailand), Narathiwat, lụt (flood) | 1 |
| 马来西亚雪兰莪州的洪灾导致两个地区的大约200名居民被疏散。(Flooding in the Malaysian state of Selangor resulted in the evacuation of about 200 residents in two areas.) | 马来西亚 (Malaysia), 雪兰莪州 (Selangor), 洪灾 (flood) | Tại Selangor, Malaysia, đường cao tốc bị cắt đứt do lũ lụt. Cảnh sát Malaysia cho biết lũ lụt do mưa liên tục nhiều ngày đã khiến 8 người thiệt mạng ở bang Selangor. (In Selangor, Malaysia, the highway was cut off due to flooding. Malaysian police said floods caused by continuous rain for many days have killed 8 people in Selangor state.) | Malaysia, lụt (flood), Xa lộ (highway), Selangor | 0 |
| ... | ... | ... | ... | ... |

Table 2 Dataset data volume.

| Dataset | Data Volume |
|----------------|-------------|
| Training Set | 4000 |
| Test Set | 200 |
| Event Elements | 9362 |

4.2 Parameter Setting and Evaluation Metrics

In this section, we set the dimension of word representations and sentence representations to 200. The Adam algorithm is employed as the optimizer [18], with a learning rate of 10^{-6} . The weight of sentence-level similarity is set to 0.7 and the batch size is set to 5. For the mBERT-SF model, we used 4032 Chinese-Vietnamese parallel sentence pairs to train the model and used the trained mBERT-SF model to obtain Chinese-Vietnamese cross-language text representation.

We use precision (P) and recall (R) as the main evaluation metrics. Precision measures the accuracy of correctly predicting true positive samples, while recall measures the proportion of positive samples that are correctly predicted. The specific formula (9), (10) is as follows:

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{K} \quad (10)$$

where TP represents the number of news pairs correctly predicted by the model as describing the same event, FP represents the number of news pairs predicted by the model as describing the same event but describing different events, and K represents the number of input news pairs describing the same event.

4.3 Baseline Model

In the task of event retrieval, the current methods mostly transform it into a text similarity calculation task to compute the similarity of the query text. The baseline model chosen for this paper primarily relies on a deep representation model for calculating similarity.

Siamese BILSTM: Paul Neculoiu et al. [14] introduced a twin network model based on a bidirectional recurrent neural network specifically designed for calculating the similarity between two input texts.

BIMPM: Wang et al. [2] proposed the BiMPPM model, a twin network model that utilizes multiple perspectives of information to calculate text similarity.

MKPM: Lu et al. [19] presented a methodology that combines event elements extraction and utilizes keyword representation vectors for sentence matching. The method is referred to as MKPM.

HASM: Li et al. [20] proposed the Hierarchical Attention Siamese Model (HASM), which incorporates a hierarchical attention mechanism for text similarity calculation. This approach leverages TextRank for summarizing and compressing lengthy documents and employs the hierarchical attention mechanism to encode and summarize the document representation at multiple levels.

4.4 Experimental Results and Analysis

To validate the effectiveness of the proposed method, this study experiments with three parts. Firstly, an experimental comparative analysis is performed to compare the proposed method with the baseline model and verify the effectiveness of the matching approach. The second part involves verifying the validity of word and sentence representation, further substantiating the effectiveness of the event elements proposed in this paper. The third part consists of comparative analysis experiments conducted in diverse language environments to confirm the effectiveness of the method proposed in this study.

4.4.1 Comparative Experiment with Baselines Model

To verify the effectiveness of our proposed method in Chinese-Vietnamese cross-language event retrieval, the method proposed in this paper is compared with the baseline model on the Chinese-Vietnamese event retrieval datasets.

From Table 3, it can be observed that the performance of

Table 3 Comparison experiment results.

| Model | P | R |
|----------------|-------|-------|
| Siamese BILSTM | 64.89 | 61.50 |
| HASM | 65.07 | 61.38 |
| BIMPM | 63.23 | 62.65 |
| MKPM | 67.32 | 62.67 |
| Ours | 69.54 | 65.86 |

Table 4 Ablation experiment results.

| Model | P | R |
|-----------------------------------|-------|-------|
| mBERT | 66.42 | 64.36 |
| W/o Word Representation Score | 68.36 | 65.35 |
| W/o Sentence Representation Score | 68.57 | 64.91 |
| Ours | 69.54 | 65.86 |

the Siamese BILSTM and HASM methods on the Chinese-Vietnamese event retrieval dataset is average. This is attributed to the information loss that occurs during the compression and dissemination process between modules at different levels in the hierarchical attention mechanism. In contrast, Wang et al. introduced an approach that incorporates interactive matching from multiple perspectives, capturing more relevant and valuable information for matching by concatenating it with the original document representation vectors. Similarly, MKPM focuses on understanding localized information within the text. As a result, these three models have shown noticeable improvements in accuracy and recall rates on the Chinese-Vietnamese event retrieval dataset. However, these models have not shown advancements in comprehending the overall context of the text, leading to lower accuracy in comparison to the approach proposed in this paper.

The method proposed in this paper achieves good results without the need for the complex operations employed in the aforementioned models. Experiments prove that the event elements in the news can capture enough key information for the news matching task, thereby assisting the news text matching model to achieve good performance. In comparison to the baseline model, there is a notable maximum improvement in accuracy of 6.3%. These results underscore the suitability of the method proposed in this paper for event retrieval tasks, particularly for news title queries.

4.4.2 Ablation Experiment

We conduct module ablation experiments to assess the effectiveness of the proposed method in enhancing retrieval performance. The results of these experiments are shown in Table 4. In this context, “W/o word representation score” refers to the score obtained by excluding the word representation task, where only the score from the sentence representation task is considered as the final score. Likewise, “W/o sentence representation score” indicates the score obtained by excluding the sentence representation task, and only the

Table 5 Experimental results based on different pre-trained models.

| Model | P | R |
|-------|-------|-------|
| mBERT | 68.93 | 64.35 |
| XLM | 68.87 | 63.92 |
| XLM-R | 68.98 | 63.25 |
| Ours | 69.54 | 65.86 |

scores from the word representation task are utilized as the final score.

The analysis of Table 4 reveals that in the conducted ablation experiment, the model incorporating event elements achieves a performance improvement of more than 3.1% compared to the method using only mBERT encoding. Furthermore, when the event element extraction module is removed, the performance of the model experiences a decrease of approximately 1%. Furthermore, from Table 3, it also can be observed that using the Sentence Representation Score and Word Representation Score alone fails to achieve optimal results. Only using the Sentence Representation Score can calculate the similarity between Chinese and Vietnamese event sentences but does not consider the correlation between events. Using the Word Representation Score alone can calculate the event similarity between sentences but ignores the relevant information between Chinese and Vietnamese sentences. Both using event elements and sentence representations in Chinese and Vietnamese sentences can enhance the effect of Chinese and Vietnamese text event representation and improve Chinese and Vietnamese events. Retrieval recall rate and accuracy rate to achieve the best results.

4.4.3 Experiments on Different Pre-Trained Models

The experiments mentioned above validate the viability of the method proposed in this research paper. However, it is essential to note that these experiments rely on fine-tuning performed on mBERT. To further validate the approach, additional tests will be conducted using other multilingual pre-trained models. The results of these specific experiments are shown in Table 5.

From the analysis of Table 5, the disparities in accuracy and recall between the XLM model [21] and the proposed method are 0.61% and 1.51%, furthermore, the variations in the two evaluation indicators between the XLM model and the proposed method are recorded at 0.56% and 2.61%. It is evident that the methods proposed in this paper exhibit their effectiveness across different pre-trained models. In this experiment, sentence representation vectors are obtained from alternative cross-language pre-trained models rather than mBERT-SF. The experimental results consistently demonstrate that the mBERT-SF model outperforms other baseline models significantly in the cross-lingual event retrieval task and using mBERT-SF can better represent Chinese-Vietnamese sentences than other pre-trained language models. The experimental results demonstrate that the mBERT-SF model can effectively improve cross-language event retrieval performance.

4.4.4 Experiments on Different Training Dataset Sizes

To explore the influence of data size on model performance, the training data was partitioned into seven groups with varying amounts. Each group was utilized for training and evaluating the model individually. The test set results, illustrating the best performance achieved in the Chinese-Vietnamese cross-lingual event retrieval task, are presented in Table 6.

Analysis of Table 6 shows that the accuracy rate of retrieval results is significantly low and unstable when the experimental data sizes fall below 1000, accompanied by a low recall rate. When the amount of experimental data is greater than 1000, the precision and recall of model retrieval will increase with the increase of experimental training data. With the increase of training data, the model can capture more event element similarity relationships and sentence representation information between Chinese and Vietnamese sentences under the same event, improving the event retrieval effect.

4.4.5 Comparative Experiments for Different Languages

To demonstrate the effectiveness of our approach beyond the Chinese-Vietnamese language context, we evaluate our method on the Chinese-English language pair, which benefits from a large-scale training corpus. We compare our method with the baseline model using a Chinese-English dataset obtained from the Internet. This dataset contains 50,000 parallel sentence pairs that serve as positive examples. We utilize a training dataset comprising 100,000 sentence pairs during the training process. The experimental findings are presented in Table 7.

The experiment demonstrates that our method achieves better results when using a large-scale corpus for training on languages with abundant resources, such as English-Chinese, compared to low-resource languages like Chinese-Vietnamese. The possible reason is that the multi-language pre-trained language model has a better representation effect on rich-resource languages than low-resource languages. A

Table 6 Experimental results based on different data sizes.

| Train data | P | R |
|------------|-------|-------|
| 100 | 33.76 | 32.32 |
| 200 | 39.55 | 38.33 |
| 500 | 41.45 | 37.69 |
| 1000 | 47.23 | 41.46 |
| 2000 | 56.45 | 53.77 |
| 3000 | 61.84 | 61.55 |
| 4000 | 69.54 | 65.86 |

Table 7 Experimental results in Chinese-English.

| Model | P | R |
|----------------|-------|-------|
| Siamese BILSTM | 81.89 | 79.22 |
| HASM | 83.12 | 79.63 |
| BIMPM | 83.23 | 82.31 |
| MKPM | 83.29 | 81.14 |
| Ours | 85.14 | 84.26 |

large amount of training data can provide the model with more event information and stronger generalization. Furthermore, the experimental results of our proposed method outperform those of the baseline model experiments.

5. Conclusion

For the Chinese-Vietnamese cross-language event retrieval task, We propose a novel Chinese-Vietnamese cross-language event retrieval method that integrates event elements to enhance the retrieval process. This approach matches the event elements present in the text individually and obtains word representation vectors. These vectors are then combined with the sentence representation vectors obtained by fine-tuning a pre-trained language model through contrastive learning. By integrating these representations, to improve the accuracy of text matching in cross-lingual event retrieval. Experimental results demonstrate the significant performance improvement achieved by the proposed method in event retrieval tasks. The incorporation of event elements allows for more precise matching of relevant information, leading to enhanced representation and improved retrieval accuracy. In future work, we will conduct an analysis of news text characteristics better to understand their influence on the model's effectiveness. Additionally, we plan to extend the application of the proposed model to other fields, exploring its potential in diverse domains.

Acknowledgments

This study was supported by the project of the National Natural Science Foundation of China (U21B2027, 62266027, 62266028, U23A20388), the Yunnan provincial major science and the technology special plan projects (202302AD080003, 202303AP140008, 202202AD080003), the Yunnan Fundamental Research Projects (202301AT070471, 202301AS070047, 202301AT070393), and the Kunming University of Science and Technology's "Double First-rate" construction joint project (202201BE070001-021).

References

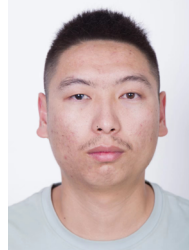
- [1] S.M. Sarwar and J. Allan, "Query by example for cross-lingual event retrieval," Proc. 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA, pp.1601–1604, 2020.
- [2] Z. Wang, W. Hamza, and R. Florian, "Bilateral multi-perspective matching for natural language sentences," Proc. 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, pp.4144–4150, 2017.
- [3] P. Sujatha and P. Dhavachelvan, "A review on the cross and multilingual information retrieval," International Journal of Web & Semantic Technology, vol.2, no.4, pp.115–124, 2011.
- [4] G. Chandra and S.K. Dwivedi, "Assessing query translation quality using back translation in Hindi-English CLIR," International Journal of Intelligent Systems and Applications, vol.9, no.3, pp.51–59, 2017.
- [5] L. Ballesteros and M. Sanderson, "Addressing the lack of direct translation resources for cross-language retrieval," Proc. 12nd ACM

International Conference on Information and Knowledge Management, New York, USA, pp.147–152, 2003.

- [6] B. Hu, Z. Lu, H. Li, and Q. Chen, “Convolutional neural network architectures for matching natural language sentences,” Proc. 28th Advances in Neural Information Processing Systems, Montreal, Canada, pp.2042–2050, 2014.
- [7] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” Proc. 57th Conference of the North American Chapter of the Association for Computational Linguistics, vol.1, pp.4171–4186, June 2019.
- [8] Y. Liu, J. Gu, N. Goyal, X. Li, S. Edunov, M. Ghazvininejad, M. Lewis, and L. Zettlemoyer, “Multilingual denoising pre-training for neural machine translation,” Transactions of the Association for Computational Linguistics, vol.8, no.1, pp.726–742, 2020.
- [9] Z. Dai and J. Callan, “Deeper text understanding for IR with contextual neural language modeling,” Proc. 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA, pp.985–998, 2019.
- [10] M. Chidambaram, Y. Yang, D. Cer, S. Yuan, Y. Sung, B. Strope, and R. Kurzweil, “Learning cross-lingual sentence representations via a multi-task dual-encoder model,” Proc. 4th Workshop on Representation Learning for NLP, Florence, Italy, pp.250–259, Aug. 2019.
- [11] S.E. Robertson and K.S. Jones, “Relevance weighting of search terms,” Journal of the American Society for Information Science, vol.27, no.3, pp.129–146, 1976.
- [12] J. Singh and S. Dwivedi, “Analysis of vector space model in information retrieval,” Proc. 22th of IJCA National Conference on Communication Technologies & Its Impact on Next Generation Computing, vol.2, pp.14–18, 2012.
- [13] M. Dragoni, C. da Costa Pereira, and A.G.B. Tettamanzi, “A conceptual representation of documents and queries for information retrieval systems by using light ontologies,” Expert Systems with Applications, vol.39, no.12, pp.10376–10388, 2012.
- [14] P. Neculoiu, M. Versteegh, and M. Rotaru, “Learning text similarity with Siamese recurrent networks,” Proc. 1st Workshop on Representation Learning for NLP, Berlin, Germany, pp.148–157, 2016.
- [15] O. Khattab and M. Zaharia, “ColBERT: Efficient and effective passage search via contextualized late interaction over BERT,” Proc. 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Virtual Event, China, pp.39–48, 2020.
- [16] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” Proc. 31st Advances in Neural Information Processing Systems, California, USA, vol.30, pp.5998–6008, Dec. 2017.
- [17] Y. Huang, Y. Liang, Z. Wu, E. Zhu, and Z. Yu, “Cross-lingual sentence embedding for low-resource Chinese-Vietnamese based on contrastive learning,” ACM Trans. Asian and Low-Resource Language Information Processing, vol.22, no.6, Article No.176, pp.1–18, 2023.
- [18] A. Barakat and P. Bianchi, “Convergence and dynamical behavior of the ADAM algorithm for nonconvex stochastic optimization,” SIAM Journal on Optimization, vol.31, no.1, pp.244–274, 2021.
- [19] X. Lu, Y. Deng, T. Sun, Y. Gao, J. Feng, X. Sun, and R. Sutcliffe, “MKPM: Multi keyword-pair matching for natural language sentences,” Applied Intelligence, vol.52, no.2, pp.1878–1892, 2022.
- [20] L. Li, J. Zhou, Y. Gu, and W. Qu, “Similar legal case retrieval based on improved Siamese network,” Acta Scientiarum Naturalium Universitatis Pekinensis, vol.52, no.2, pp.84–90, 2019.
- [21] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, and V. Stoyanov, “Un-supervised cross-lingual representation learning at scale,” Proc. 58th Annual Meeting of the Association for Computational Linguistics, Stroudsburg, USA, pp.8440–8451, 2020.



Yuxin Huang was born in 1983. He received the Ph.D. degree from Kunming University of Science and Technology in 2021. Now, he is an associate professor at Kunming University of Science and Technology. His research interests include natural language processing, text generation, etc.



Yuanlin Yang is pursuing a master’s degree at the School of Information Engineering and Automation of Kunming University of Science and Technology since September 2021. His research interests include natural language processing and information retrieval, etc.



Enchang Zhu is a Ph.D. candidate in computer science at Kunming University of Science and Technology, China. His research interests include natural language processing, information retrieval, etc.



Yin Liang was born in 1996. He received the master’s degree from Kunming University of Science and Technology in 2023. Now, he is graduated from Kunming University of Science and Technology. His research interests include cross-linguistic sentence representation etc.



Yantuan Xian is currently an associate professor at Kunming University of Science and Technology, China. He graduated from Yunnan Normal University, China, in 2003. He received the M.S. degree from Shenyang Institute of Automation (SIA), China, in 2006. His research interests include pattern recognition, machine learning, and information retrieval.