MusREL:
A Utility-Weighted Multi-Strategy Relation Extraction Model-Based Intelligent System for Online Education

Zhen Zhu, Keyi College, Zhejiang Sci-Tech University, China & AI Lab, NetEase Inc., China
Huaiyuan Lin, Wenzhou Vocational College of Science and Technology, China
Dongmei Gu, Hangzhou City University, China
Liting Wang, Keyi College of Zhejiang Sci-tech University, China*
Hong Wu, Keyi College of Zhejiang Sci-tech University, China
Yun Fang, AI Lab, NetEase Inc., China

ABSTRACT

In order to enhance the utility of online educational digital resources, the authors propose a practical and efficient multi-strategy relation extraction (RE) model in online education scenarios. First, the effective relation discrimination model is used to make relation predictions for non-structured teaching resources and eliminate the noise data. Then, they extract relations from different path strategies using multiple low-computational resources and efficient relation extraction strategies and use their proposed multi-strategy weighting calculator to weigh the relation extraction strategies to derive the final target relations. To cope with the low-resource relation extraction scenario, the relation extraction results are complemented by using prompt learning with a big model paradigm. They also consider the model to serve the commercial scenario of online education, and they propose a global rate controller to adjust and adapt the rate and throughput requirements in different scenarios, so as to achieve the best balance of system stability, computation speed, and extraction performance.

KEYWORDS

Artificial Intelligence, Low-Cost Cloud Computing, Multiply Strategies, Online Learning System, Relation Extraction

With the rapid development of a knowledge-based economy, educational resources have become one of the important factors of national and regional competitiveness. However, the uneven distribution of educational resources is considered to be an important cause of developmental differences in different regions of the country (Juárez-Varón et al., 2023; Li et al., 2022). Educators have been calling for educational reforms that combine machine learning and intelligent technologies to build online

DOI: 10.4018/IJSWIS.329965
*Corresponding Author

education systems that add new vitality to education and promote the balance of regional educational resources (Gaurav et al., 2022).

After the outbreak of COVID-19, the online education industry has grown rapidly, such as with NetEase Open Course, Youdao Excellent Course, and K12 online courses. At present, online education is not only popular in the field of adult education (Al-Qerem et al., 2020; George & Lal, 2021), but is also popular in the field of primary and secondary school subject education (Gupta et al., 2023; Gupta & Quamara, 2020). In the years of the COVID-19 epidemic, numerous offline educational institutions began to turn online. Change in the teaching paradigm has further increased the market share of online education (Elgendy et al., 2021; Singla et al., 2022).

With the development of online teaching content, a large number of teaching plan materials, teaching videos, and learning records are stored digitally. Unstructured raw data is often difficult to utilize. However, turning to structured data with an automatic system makes tasks possible in many teaching and learning scenarios; this was impossible in the past, and it plays an important role in the education field (Kaur et al., 2021). For example, educational knowledge graph construction and structured lesson plans generated based on structured data play an important role in teaching management, educational asset inheritance, and learning practices. In addition, structured data also plays an important role in teaching knowledge answering. Relation extraction technology is the main means to solve the problem of transforming educational resources from unstructured data to structured data (Kar, 2022). Unfortunately, most of the existing approaches are expensive, demand high computational resources, take a long time to compute, and do not support dynamic adjustment; Most of the existing approaches are difficult to be applied to today's complex multi-scenario online education systems.

This paper proposes a multi-strategy relation extraction model named MusREL for structured data generation in online education scenarios that achieves the best balance of result stability, computational speed, and extraction performance. The model has been widely applied in teaching plan materials structuring, multi-resource educational knowledge graph construction, and educational Q&A by the NetEase education group, and it has achieved good practical value.

RELATED WORK

The technique of relation extraction was first formally proposed in the 1990s, and after more than 30 years of development, the paradigm of relation extraction has experienced great changes. The earliest approach used rule-based template matching, in which manual rule templates played a crucial role, and linguists were the key to the task of relation extraction at that stage (Huffman, 1995; Kim & Moldovan, 1995). One of these approaches used trigger words (Califf & Mooney, 1997; Nakashole et al., 2012); another approach was based on dependent syntactic analysis (Fundel et al., 2006) and template matching based on lexical information and positional relations (Nédellec, 2005; Nebhi, 2013). The approaches used rule-based template matching are simple and efficient but can match only very limited relations. By the 2000s, traditional machine learning approaches had become popular, and feature engineering approaches such as SVM and maximum entropy were used extensively (Lafferty et al., 2001; Och et al., 2004). Researchers and domain experts used domain knowledge to extract features from the original corpus and then used traditional machine learning methods of classification and clustering to do relation prediction. Among them, supervised approaches became mainstream and achieved the best performance at that time. Some scholars have used syntactic tree kernel functions using syntactic dependencies instead of shallow string information and added lexical, syntactic semantic labels and dependencies as supplementary features of the kernel functions (Culotta & Sorensen, 2004; Zhou et al., 2007). As a result, the approach of tree kernels trained on Penn Tree Bank achieved more promising results at that time (Zhang et al., 2008). Distanted-supervision is another approach to relation extraction. This approach is based on a hypothesis that if two entities have some relations in the knowledge base, unstructured sentences containing those two entities will contain these relations (Mintz et al., 2009); this approach can complement the existing structured data, but the noise problem has been an obstacle to the application of this approach (Riedel et al., 2010). The unsupervised machine learning approach is another way of completing the RE task, and a K-means clustering statistical machine learning approach has been proposed, which provides a new idea for relation extraction task (Chen et al., 2005).

Later, neural network models appeared (Bengio et al., 2013; Chung et al., 2014), such as CNN and RNN, which led to a dramatic change in the paradigm of relation extraction. Around 2012, neural network models were introduced to the work of relation extraction. After that, the focus of extraction work shifted to the model architecture. Designing well-defined network architectures that facilitates learning to extract relational features improves the performance of relation extraction. Some scholars proposed the Piece-wise CNN model, which slices the original sentence based on the position of the occurrence of entity words, and each part is subjected to separate convolution and pooling operations and is used as input to the final Softmax, which was evaluated on the New York Times corpus with an experimental result and reached to 78.3% of f-score (Zeng et al., 2015). Inspired by this approach, some scholars proposed a CNN-based multilayer attention mechanism model in 2016, which combines the lexical-semantic features with the relative distance of word positions, and the f-score reached 88% of f-score on SemEval 2010 Task 8 (Wang et al., 2016). However, the CNN model is not suitable for dealing with long sentences, so some scholars proposed the Bi-LSTM overlaying attention mechanism to do relation extraction. Firstly, the sentence is used as a sequence layer in the LSTM layer, and the attention layer is overlaid to capture the word-word interrelationship to learn the representation of the whole sentence and then complete the prediction of relation. This approach achieved good performance (Zhou et al., 2016). However, by 2018, big models started to prevail, and the research paradigm of relation extraction changed again. Many researchers shifted their research focus to a combination of pre-trained models and fine-tuning approaches (Dong et al., 2019; Lewis et al., 2020). In this paradigm, a general language model with a fixed architecture and large size is first trained (Peters et al., 2018). Then, additional training data is introduced for fine-tuning the specific task of relation extraction. Thus, the model can do relation extraction tasks with excellent performance.

In recent years, a large number of new relation extraction approaches have been proposed, one of which is the joint extraction approach named CasRel (Wei et al., 2020), which uses the new label to tag the position of the subject, and the object, and perform joint relation extraction; this can extract multiple relations in a sentence and achieves good performance. This paper is inspired by the extraction idea of CasRel and extends multiple extraction paths to implement the multi-strategy relation model named MusREL. In addition, TPLink (Wang et al., 2020), StereoRel (Tian et al., 2021), and RIN (Sun et al., 2020) also achieved good results in RE tasks using the new architecture. Unfortunately, rather less research has been done in the area of Chinese relation extraction, even though those works are very useful for Chinese scenarios. According related references, some scholars proposed a BLSTM model for Chinese relation extraction (Zhang & Wang, 2015; Zhou et al., 2016), which achieved good results at that time. In addition, some scholars were inspired to propose a CNN model for multi-instance learning based on the attention mechanism, which further improved the performance (Lee et al., 2019; Lin et al., 2016). For the scenario of joint extraction, some scholars added the label of RE to the NER model and proposed a relation extraction model for multilingual information. Recently, other scholars have proposed ExSoftwords+BERT (Kong et al., 2021) and BERT+MuitiView (Yang et al., 2023) by taking advantage of the powerful model generalization capability of BERT and it achieves the best results. The MusREL method proposed in this paper, in order to solve the relation extraction problem in commercial scenarios in education, uses a multi-strategy approach and experiments on Chinese and English corpora separately, and the performance is significantly improved relative to the baseline. Moreover, due to the low resource requirement of the model and the fast computation rate, and high stability, it has higher practical value in commercial scenarios.

THE MUSREL MODEL

MusREL Framework

The MusREL framework proposed in this paper is shown in Figure 1. The authors divide the model framework into six parts, each of which plays an important role in its respective objectives:

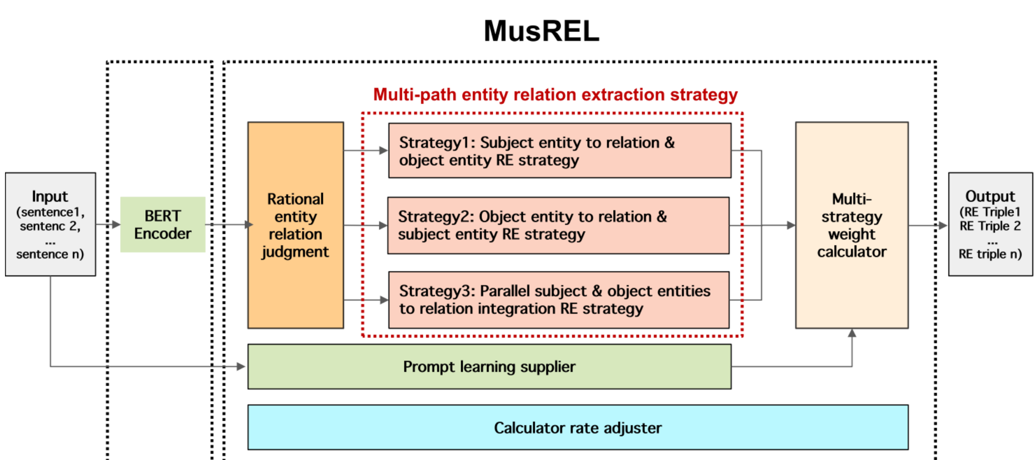
1. BERT encoder: Based on the BERT pre-trained language model, the input sentences are encoded so as to obtain the vectorized representation of the tokens in the sentences.
2. Rational entity relation judgment: In practical scenarios, a large number of sentences do not have entity relations, so it is necessary to do relation existence judgment on the sentences to effectively reduce the loss of computational resources as well as to a certain extent to improve the prediction of the recall rate.
3. Multi-path entity relation extraction strategy: This paper uses three parallel paths of relation extraction, the first two using joint extraction, the latter using the pipeline approach. The authors' multi-strategy relation extraction approach can effectively reduce the occurrence of prediction shortfalls for a single relation category.
4. Multi-strategy weight calculator: Multi-strategy weight calculator assigns different weights to multiple relation extraction methods based on categories and obtains the comprehensive optimal results through weighted summation.
5. Prompt learning supplier: Prompt learning supplier uses the large language model in the general scenario to do the relation prediction supplement to assist the multi-strategy module decision.
6. Calculator rate adjuster: This module is mainly used for different scenarios to do strategy use state adjustment. Thus, in the application of an online education system, performance, rate, and stability are balanced.

MusREL Encoder

First of all, for encoder selection, the authors used the pre-trained model BERT to encode contextual features (Vaswani et al., 2017). In fact, the authors believe that the use of other pre-trained models, such as T5 (Raffel et al., 2020) and GPT (Katz et al., 2023), is still valid here as well.

Here we briefly review BERT, a large language model based on a multilayer bi-directional Transformer. It has proven to be very effective in many downstream areas. Specifically, it consists of a stack of N identical Transformer blocks. The authors encode each input sentence by BERT and

Figure 1. MusREL framework



transform it into embedding vectors that aggregate semantic and positional information. BERT-based encoders are selected by language; in other words, the authors use different languages trained models for different languages, such as English and Chinese.

Rational Entity Relation Judgment

Because there are a large number of unrelated sentences in the actual corpus, in the online education scenario, the authors find that more than 90% of the corpus is unrelated. To improve efficiency, it is necessary to first train a rational relation judgment to do the first filtering of sentences. In a real business scenario, this module is significant for efficiency improvement. Specifically, the authors first predict a subset of potential relations that may exist in the sentence and then predict and finally decide the highest-scoring relation in the subsequent multi-strategy prediction session. The rational entity relation judgment uses the vector output from the MusREL encoder as the input and then outputs a probability value. The probability value is used to indicate the likelihood that an entity relationship exists, as shown in Eqs. (1-2).

$$h^{maxpool} = maxpool\left(h_i\right) \tag{1}$$

$$\mathrm{P}^{\mathrm{relation}} = \tilde{\mathrm{A}}\left(\mathrm{W}_{\mathrm{relation}}\mathrm{h}^{\mathrm{max}} + \mathrm{b}_{\mathrm{relation}}\right) \tag{2}$$

$\mathrm{P}^{\mathrm{relation}}$ is the relation probability value, where $maxpool\left(h_i\right)$ is the maximization pooling operation, because the authors believe that $maxpool$ is more effective in characterizing the features used for relation classification in the sentence. h_i is the vectorized representation of the i-th token in the sentence. $\mathrm{W}_{\mathrm{relation}}$ is the trainable matrix, and A is the sigmoid activation function.

At this point, the authors use potentially valid relations for binary prediction. When the probability exceeds a threshold, the relation is given the label 1, which means it is considered to have a valid relation; otherwise, it is given the label 0, which is considered to be no valid relation. Later, the authors only need to perform relation extraction for the sentences with potential relations. The authors use loss calculation of cross entropy to train the judge for each relation and do training with Eq. (3) based on the training data.

$$\mathrm{Loss}_{\mathrm{relation}} = -\frac{1}{\mathrm{count}\left(\mathrm{token}\right)}\sum_{\mathrm{i}=1}^{\mathrm{count}\left(\mathrm{token}\right)}\left(\mathrm{t}_{\mathrm{i}}\log\mathrm{P}^{\mathrm{relation}} + \left(1-\mathrm{t}_{\mathrm{i}}\right)\log\left(1-\mathrm{P}^{\mathrm{relation}}\right)\right) \tag{3}$$

where $\mathrm{Loss}_{\mathrm{relation}}$ is the cumulative loss value, t_{i} is the real label at position i, $\mathrm{count}\left(\mathrm{token}\right)$ is the number of tokens in the sentence, and $\mathrm{P}^{\mathrm{relation}}$ is the probability value of the current entity relation.

Multi-Path Entity Relation Extraction Strategy

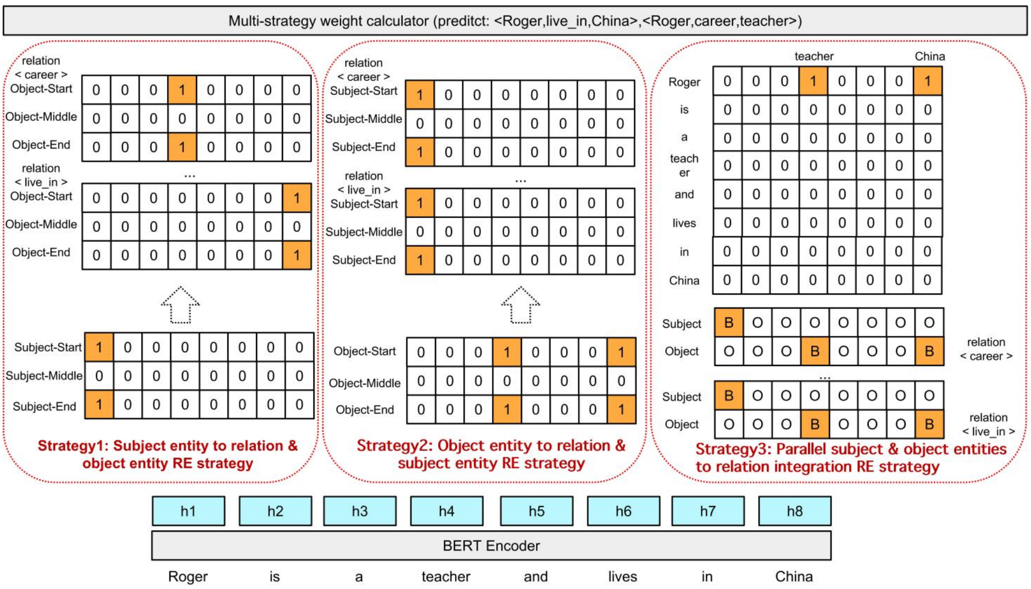
The authors use three different strategies to extract relations from different paths, as shown in Figure 2. The input is the encoded vectors by BERT, and the output is the relation predicted probability for each strategy.

Strategy 1: Subject Entity to Relation and Object Entity RE Strategy

The first relation extraction strategy the authors adopt is the forward relation extraction approach as shown in Figure 2. Specifically, using the encoding vector, the authors first extract the position of the subject entity; and then, they extract the object entity with the position and relation matrix.

Identify the Position of the Subject Entity. The subject position is represented by three position vectors, which denote the start position, middle position, and end position of the subject in the

Figure 2. Multi-path entity relation extraction strategy



sentence, respectively. This way is distinguished from the tagging method of CasRel (Wei et al., 2020), TPLinker (Wang et al., 2020), and PMEI (Sun et al., 2021); this paper argues that the identification of middle position features can improve the performance of the model for relation judgment while the correct identification of subject features will directly affect the subsequent object entity recognition and relation classification, and this approach allows the model to recognize more features for more precise relation prediction. This type of approach can support scenarios where multiple subjects appear in a sentence. The three positional vectors are equal to the length of the sentence, and each element represents the probability of the position token as the start position, middle position, and end of the subject entity, as shown in Eqs. (4-6).

$$P_{subject}^{i,start} = \sigma\left(W_{subject}^{start} h_i + b_{subject}^{start}\right) \quad (4)$$

$$P_{subject}^{i,middle} = \sigma\left(W_{subject}^{middle} h_i + b_{subject}^{middle}\right) \quad (5)$$

$$P_{subject}^{i,end} = \sigma\left(W_{subject}^{end} h_i + b_{subject}^{end}\right) \quad (6)$$

Among them, $P_{subject}^{i,start}$, $P_{subject}^{i,middle}$, and $P_{subject}^{i,end}$ represent the probabilities of the i-th token as the start position, middle position, and end position of the subject, respectively. $W_{subject}^{*}$ is a trained matrix, and $b_{subject}^{*}$ is a vector of bias. σ is a sigmoid activation function, which is used to output probability values from 0 to 1.

After the final activation function outputs the probability value, the authors set the threshold value and transform the result into the 0 or 1 label by threshold. When the result is less than the threshold value, it is set to 0, which means that this position has not any relation exists, and if the result is larger than the threshold value, it is set to 1, which means that this position has a relation exists. After calculating the result, the authors restrict the prediction result with the rule. Only when value of $P_{subject}^{i,start}$ and the value of $P_{subject}^{j,end}$ are exists and j must be greater than or equal to i. Take the nearest $P_{subject}^{i,start}$ ... $P_{subject}^{j,end}$ as the entity of the subject. The $P_{subject}^{k,middle}$ can exist only between the start position and the end position. All cases other than rules are discarded.

Identify Entity Relation and Object Entity Position. In this part, the positions of the relations and the object are identified mainly by the positions of the previously identified subjects. Specifically, given a selected subject, the token in each sentence is judged to be the object in each relation classification and the probability values of three types of object positions, where the three position vectors are start position, middle position, and end position, as shown in Eqs. (7-10).

$$h_{k-subject}^{weightpool} = weightpool\left(h_{k-subject}^{start}, h_{k-subject}^{middle}, \ldots, h_{k-subject}^{end}\right) \tag{7}$$

$$P_{k-object}^{i,start} = \sigma\left(W_{k-object}^{start}\left(h_i \times h_{k-subject}^{weightpool}\right) + b_{object}^{start}\right) \tag{8}$$

$$P_{k-object}^{i,middle} = \sigma\left(W_{k-object}^{middle}\left(h_i \times h_{k-subject}^{weightpool}\right) + b_{object}^{middle}\right) \tag{9}$$

$$P_{k-object}^{i,end} = \sigma\left(W_{k-object}^{end}\left(h_i \times h_{k-subject}^{weightpool}\right) + b_{object}^{end}\right) \tag{10}$$

$h_{k-subject}^{start}, h_{k-subject}^{middle}, \ldots, h_{k-subject}^{end}$ denotes the k-th subject and those vectors are used to characterize the semantic information for the token. $P_{k-object}^{i,start}$, $P_{k-object}^{i,middle}$, and $P_{k-object}^{i,end}$ are the probabilities that the i-th token becomes the start position, middle position, and end position of the object related to the k-th subject. And $\times$ is the dot product operation of the matrix. $W_{k-object}^{*}$ is the matrix being trained, and b_{object}^{*} is the bias. $weightpool\left(*\right)$ is the weight pooling operation. The authors consider that the token at the start and end position has more important features than the token at the middle position; therefore, the authors use the weight-based pooling operation. The weights of the start and end positions are 1, and the weight of the middle position is 0.5.

Here, based on each relation classification, each token of the same length as the sentence is calculated $P_{k-object}^{i,start}$, $P_{k-object}^{i,middle}$, and $P_{k-object}^{i,end}$ values, and when there is no legal object position in a relation, the authors consider that there is no relation of that class in the sentence. Eventually, the authors find the positions of all objects that meet the restraints under all relations.

Joint Cross-Entropy Loss Calculation. After the authors finish the prediction, the authors have to calculate the loss function based on the subject and object positions of start, middle, and end in the predicted relations with that in the real labeled data. The authors use Joint entropy-based loss calculation, as shown in Eqs. (11-15).

$$\mathrm{ce}\left(\mathrm{P},\mathrm{t}\right) = -\left[\mathrm{tlogP} + \left(1-\mathrm{t}\right)\log\left(1-\mathrm{P}\right)\right] \tag{11}$$

$$\mathrm{Loss}_{\mathrm{subject}} = -\frac{1}{2\times\mathrm{count}\left(\mathrm{token}\right)}\sum_{\mathrm{j}\epsilon\left\{\mathrm{start,middle,end}\right\}}\sum_{\mathrm{i}=1}^{\mathrm{count}\left(\mathrm{token}\right)}\mathrm{ce}\left(\mathrm{P}_{\mathrm{subject}}^{\mathrm{i,j}}, \mathrm{t}_{\mathrm{subject}}^{\mathrm{i,j}}\right) \tag{12}$$

$$\mathrm{Loss}_{\mathrm{k-object}} = -\frac{1}{2\times\mathrm{count}\left(\mathrm{token}\right)}\sum_{\mathrm{j}\epsilon\left\{\mathrm{start,middle,end}\right\}}\sum_{\mathrm{i}=1}^{\mathrm{count}\left(\mathrm{token}\right)}\mathrm{ce}\left(\mathrm{P}_{\mathrm{k-object}}^{\mathrm{i,j}}, \mathrm{t}_{\mathrm{k-object}}^{\mathrm{i,j}}\right) \tag{13}$$

$$\mathrm{Loss}_{\mathrm{object}} = \sum_{\mathrm{k}=1}^{\mathrm{count}\left(\mathrm{subject}\right)}\mathrm{Loss}_{\mathrm{k-object}} \tag{14}$$

$$\mathrm{Loss}_{\mathrm{total}} = \sum_{\mathrm{m}=1}^{\mathrm{count}\left(\mathrm{relation}\right)}\left(\mathrm{Loss}_{\mathrm{subject}} + \mathrm{Loss}_{\mathrm{object}}\right) \tag{15}$$

$\mathrm{ce}\left(\mathrm{P},\mathrm{t}\right)$ is the cross-entropy loss function, where t is the labeled true value and P is the predicted value. In addition, the $\mathrm{Loss}_{\mathrm{subject}}$ and $\mathrm{Loss}_{\mathrm{k-object}}$ are the loss functions of subject position and object position with the k-th subject in the sentence. The $\mathrm{count}\left(\mathrm{token}\right)$ is the total number of tokens in the sentence, and $\mathrm{count}\left(\mathrm{subject}\right)$ is the number of subjects in all relations in the sentence. Each token can be predicted as a subject or an object at position labels ranging from start, middle, and end. $\mathrm{ce}\left(\mathrm{P},\mathrm{t}\right)$ is the cross-entropy loss function, and the loss values of the two parts are optimized together. The authors use the Adam (Diederik & Ba, 2015) optimizer for training. In particular, note that k is the number of subject entities for all relations present in a single sentence.

Strategy 2: Object Entity to Relation and Subject Entity RE Strategy

The second strategy of the authors' relation extraction is similar to the previous strategy, but the extraction path is different. This strategy goes to extract the position of object entities first and calculates the start position, middle position, end position, and entity relation of the subject based on the position of each object entity. The loss calculation also uses joint cross-entropy loss function. The authors believe that this strategy is effective and some relations and object entities are closer and easier to be identified correctly in this strategy. this strategy can complement each other with the Strategy 1. The calculation method is similar to the Strategy 1 and will not be repeated here.

Strategy 3: Parallel Subject and Object Entities to Relation Integration RE Strategy

For example, there may be an entity that is both a subject and an object. For such cases, the authors need to use two parallel paths of relation extraction starting from the subject and starting from the object. The authors divide the relation extraction into two stages as shown in Figure 2.

Parallel Prediction of Subject Position and Object Position Under Each Relation. In the third strategy of relation extraction, again using the vector representation after the BERT encoded as input, the authors perform independent prediction of subject entity position and object entity position on each relation as shown in Eqs. (16-21).

$$Pk_{subject}^{i,start} = Softmax\left(W_{subject}^{start}\left(h_i + r_k\right) + b_{subject}^{start}\right) \tag{16}$$

$$Pk_{subject}^{i,middle} = Softmax\left(W_{subject}^{middle}\left(h_i + r_k\right) + b_{subject}^{middle}\right) \tag{17}$$

$$Pk_{subject}^{i,end} = Softmax\left(W_{subject}^{end}\left(h_i + r_k\right) + b_{subject}^{end}\right) \tag{18}$$

$$Pk_{object}^{i,start} = Softmax\left(W_{object}^{start}\left(h_i + r_k\right) + b_{object}^{start}\right) \tag{19}$$

$$Pk_{object}^{i,middle} = Softmax\left(W_{object}^{middle}\left(h_i + r_k\right) + b_{object}^{middle}\right) \tag{20}$$

$$Pk_{object}^{i,end} = Softmax\left(W_{object}^{end}\left(h_i + r_k\right) + b_{object}^{end}\right) \tag{21}$$

where Pk represents the probability of the start, middle, and end positions of the i-th token as subject or object. r_k represents the trainable vector of the k-th relation, and h_i is the hidden vector representation after the encoder of the i-th element. $W_{subject}^{*}$ and W_{object}^{*} are trainable matrices for subject and object, respectively, and the authors train the matrices for start, middle, and end positions, respectively. b^{*} is the corresponding bias.

Global Entity Relation Judgment. After the start, middle, and end positions of all the subject and object entities are recognized, the authors need to pick the positions of the relation and subject and object entity pairs that satisfy the constraints of these relations and keep them when the current relation and entity predictions reach a certain threshold. The vector representation of the relation and the calculation of the relation probability are shown in Eqs. (22-23).

$$h_{rel} = maxpool\left(h_{subject}^{start},\ldots,h_{subject}^{end}\right) + maxpool\left(h_{object}^{start},\ldots,h_{object}^{end}\right) \tag{22}$$

$$P_{rel}^{m} = \sigma\left(W_{rel}h_{rel}\right) + b_{rel}) \tag{23}$$

Suppose the sentence has m pairs of subject and object entities and their relations satisfy the threshold. $h_{subject}^{*}$, h_{object}^{*} is the subject hidden vector representation and object hidden vector representation in the target sentence under this relation, rel is this target relation classification, and W_{rel} is the trainable matrix of the relation rel. b_{rel} is the bias vector of the rel relation, and σ is the sigmoid activation function. P_{rel}^{m} is the probability value of the m-th entity relation pair satisfying the constraints.

Global Loss Calculation. The entity position prediction and global entity relation judgment are calculated using two loss functions. The authors splice the two loss functions together for calculation as shown in Eqs. (24-26).

$$\mathrm{Loss}_{\mathrm{position}} = -\frac{1}{\mathrm{count}\left(\mathrm{token}\right)\times \mathrm{n}_{\mathrm{rel}}}\sum_{\mathrm{k=1}}^{\mathrm{n_{rel}}}\sum_{\mathrm{i=1}}^{\mathrm{count}(\mathrm{token})}\log\left(t_{position}^{i,k}Pk_{position}^{i}\right) \tag{24}$$

$$\mathrm{Loss}_{\mathrm{rel}} = \frac{1}{count\left(m\right)}\sum_{m=1}^{count(m)}\mathrm{ce}\left(P_{rel},t_{rel}\right) \tag{25}$$

$$\mathrm{Loss}_{\mathrm{total}} = \frac{1}{count\left(position\right)}\sum_{\mathrm{position}}^{\mathrm{count(position)}}\mathrm{Loss}_{\mathrm{position}} + \mathrm{Loss}_{\mathrm{rel}} \tag{26}$$

$\mathrm{n}_{\mathrm{rel}}$ is the total count of valid relations, and the position label includes six types of positions with labels of $\mathrm{start,middle,end}$ in subject and object. $\mathrm{Loss}_{\mathrm{position}}$ is the sum of the loss of each token in each valid relation under a particular position, and $\mathrm{Loss}_{\mathrm{rel}}$ is the sum of cross entropy for each valid relation. Finally $\mathrm{Loss}_{\mathrm{total}}$ is the sum of joint $\mathrm{Loss}_{\mathrm{position}}$ and $\mathrm{Loss}_{\mathrm{rel}}$ for the uniform optimization of the model parameters.

Multi-Strategy Weight Calculator

Above, the authors trained three relation extraction strategies independently. The three strategies have different prediction performances in various classifications because of structural differences, so it is necessary to integrate the advantages of the three strategies to output the best extraction results overall. Therefore, the authors train the weights of the three relation extraction strategies for each classification on the training data. They apply this multi-strategy model to the final relation extraction task. Specifically, they take 80% of the training data set for training and 20% for weight calculation. They do the training on 80% of the data using the three strategies, and after the training is completed, they do the result prediction in the remaining 20% of the data and assign the weight of the strategy in the class based on the f-score of the three classifiers in predicting the relation as shown in Eqs. (27-28).

$$\text{»}_{\mathrm{i}} = \frac{\mathrm{f-score_i}}{\sum_{\mathrm{j=1}}^{\mathrm{count(strategy)}} \mathrm{f-score}} \tag{27}$$

$$P_{target} = \sum_{\mathrm{i=1}}^{\mathrm{count(strategy)}} \text{»}_{\mathrm{i}} P_{strategy\_i} \tag{28}$$

$\mathrm{f-score}$ is used for the weight calculation, which is calculated in each relation extraction strategy in the specific class. The authors use »_{i} to present the result. For prediction, the $\mathrm{P_{strategy\_i}}$ and the weights »_{i} are multiplied and accumulated to get the prediction score of the model in $\mathtt{target}$ relation. In outcome prediction, subject and object positions are determined and the relation is predicted to exist when a specific threshold is met.

Prompt Learning Supplier With Big Model

To the above multi-strategy model, the authors additionally complement the paradigm of utilizing cue learning and large language models, and the authors utilize the large language models trained by NetEase on a generalized corpus to provide universal knowledge for relation extraction.

The authors do not apply the strategy in every relation because, at present, their language model still performs much worse than the above relation extraction strategy in high-precision relation extraction tasks. However, the above relation extraction strategy is not capable of generalizing to unseen domains (i.e., if the relation classification does not occur or rarely occurs in the training corpus, the above strategy is not capable enough); therefore, the authors utilize the prompt learning paradigm to mainly address scenarios where the relations are sparse or there is no training corpus in the first place. Specifically, the authors adopt the manual prompting template approach and the relational result bag-of-words mapping approach as complementary strategies for such scenarios.

Calculator Rate Adjuster

In actual commercial scenarios, a calculator rate adjuster is necessary, and for different scenarios, modular switches will be configured for the scenario based on response time requirements and machine computing cost requirements. Based on the scenario, the appropriate shutdown is carried out so as to achieve the timeliness and stability goals. For example, in the scenario of real-time entity relation extraction in online live teaching, because the authors are concerned with entity relation within the course, the authors use the calculator rate adjuster to turn off the prompt learning supplier module, which reduces the unseen generalization ability of the scenario but improves the throughput and rate so as to achieve the online real-time requirements of the scenario.

EXPERIMENTS

Experiment Settings

Dataset

For a fair comparison, the authors evaluate their model on two publicly available datasets, NYT (Riedel et al., 2010) and WebNLG (Gardent et al., 2017), which have two different versions of NYT*, WebNLG*, NYT, and WebNLG, respectively. The former annotates the last token of the entity and the latter annotates the whole entity span (Zeng et al., 2018). In this paper, the authors argue that the complete annotated data can better show the performance of the model.

Because a large number of Chinese languages are involved in the domestic education field, the authors especially selected Chinese datasets for their experiments. Since there are few publicly available Chinese RE corpora, this paper only supplements two Chinese corpora. Firstly, Considering that the online teaching scenario involves more literary texts, the authors use the Chinese SanWen corpus (Xu et al., n.d.), which contains 837 Chinese documents, of which 17,227 sentences are used for training, 2,220 sentences are used for testing. The second one is the SciRE (Zhao et al., 2019) corpus, which is a corpus of 3,500 Chinese scientific papers containing four types of relationships.

Evaluation

In this paper, the results are evaluated using the indicators of accuracy, recall, and F1 score. The relation extraction task is divided into two matching standards: one is the lenient matching standard (i.e., the extracted relation is considered correct if the predicted relation as well as the subject entity and object are correct). The second is the strict matching standard, where the entity location relations (beginning, middle, and end) are considered to be correctly matched only if they are exactly matched on the correct ones. The authors follow the baseline practice for NYT* and WebNLG* by lenient matching standard. And the strict matching standard is used in NYT and WebNLG.

Baseline

In this paper, for relation extraction, the authors find representative methods in Chinese and English scenarios as the baseline of the paper, and most of the results are directly adopted from the original paper, but there are some representative methods that do not provide experimental results on Chinese datasets, such as CasRel (Wei et al., 2020). In this paper, the authors use the source code provided in the paper, add data preprocessing logic and replace the tokenizer, and conduct several experiments on the Chinese dataset to get the best results possible. The reference baselines in this paper include ETL-Span (Yu et al., 2019), WDec (Nayak & Ng, 2020), RSAN (Yuan et al., 2020), CasRel (Wei et al., 2020), TPLinker (Wang et al., 2020), Chinese aspect BLSTM, Att-PCNN (Lee et al., 2019), MG-Lattice (Li et al., 2019), MultiView+biword (Yang et al., 2023), ExSoftwords(Lewis et al., 2020) and BERT+MultiView (Yang et al., 2023).

Experimental Details

We use AdamW (Diederik & Ba, 2015) for model optimization; the position judgment threshold for subject and object is set to 0.5, the learning rate is 1.5*e-5. The authors set the batch size to 12, the model training is performed 10 times, and the average result is taken as the reported result. Epoch is set to 150, and for the weight parameter the authors use random initialization.

Specifically, when the authors train the strategy weights for each relation, they split the training set into 80% and 20% randomly, where 80% is used for independent training of each strategy. First, the authors train a certain number of times to reach the basic stability of the model (e.g., 100 times). Then the three strategies are used to make predictions for each of the remaining 20% of the training data, and the weight of that strategy on the output results is assigned based on the predicted f-score. Then the authors put the remaining 20% back into the training set and train the original strategy for the remaining 50 epochs to allow the model effect to be further improved.

Experimental Results

Overall Experimental Results

Our main experimental results are shown in Table 1. The authors compare the experimental results on datasets of NYT*, WebNLG*, NYT, and WebNLG with those of representative methods, and the MusREL achieves the best results in the f1-score. The authors observe that MusREL has a significant improvement in both recall and F1-score, which proves that the multi-strategy approach is able to extract relations from different perspectives and can obtain more comprehensive entity relations than a single strategy can, and therefore it has a significant improvement in recall, which confirms the authors' hypothesis.

Table 1. Comparison of experimental results of representative approaches

Approach	NYT*			WebNLG*			NYT			WebNLG		
	P	R	F	P	R	F	P	R	F	P	R	F
RSAN (Yuan et al., 2020)	-	-	-	-	-	-	85.7	83.6	84.6	80.5	83.8	82.1
WDec (Nayak & Ng, 2020)	-	-	-	-	-	-	88.1	76.1	81.7	-	-	-
ETL-Span (Yu et al., 2019)	84.9	72.3	78.1	84.0	91.5	87.6	85.5	71.7	78.0	84.3	82.0	83.1
RIN (Sun et al., 2020)	87.2	87.3	87.3	87.6	87.0	87.3	83.9	85.5	84.7	77.3	76.8	77.0
TPLinker (Wang et al., 2020)	91.3	92.5	91.9	91.8	92.0	91.9	91.4	92.6	92.0	88.9	84.5	86.7
CasRel (Wei et al., 2020)	89.7	89.5	89.6	93.4	90.1	91.8	89.8	88.2	89.0	88.3	84.6	86.4
MusREL	91.1	94.1	92.6	93.1	94.7	93.9	91.2	93.9	92.5	88.5	89.7	89.1

Experimental Results of Chinese Dataset

For the actual scenario of domestic online education, the authors especially supplement two Chinese datasets and compare representative baselines in the Chinese domain; while CasRel does not provide experimental results for Chinese datasets, this paper makes secondary modifications to the source code so that the source code can handle Chinese datasets and switch to support Chinese BERT tokenizer. The experimental results are shown in Table 2. MusREL also achieves the best results in the SanWen dataset. However, it does not perform as well as the BERT+MultiView method in the SciRE corpus. The authors consider that MuREL's relation extraction ability in complex corpus still needs to be improved. After analysis, the authors' current strategies are under-performing for sparse relational data and open relational data. In the future, the authors will try to add more relation extraction strategies for complex scenarios under the multi-strategy framework so as to further improve the comprehensive processing capability of the model.

Comparison of Experimental Results of Different Relation Types

There are differences between the extraction effects of different path extraction strategies on different relationship types. Because the words and location features and associated entities that are of interest to a particular relation classification are closely related, the authors do experimental comparisons on

Table 2. Comparison of experimental results of Chinese representative approaches

Approach	SanWen	SciRE
	F-Score	F-Score
BLSTM (Zhou et al., 2016)	61.0	87.4
Att-PCNN (Lee et al., 2019)	60.5	88.5
MG-Lattice (Li et al., 2019)	65.5	89.8
MultiView+biword (Yang et al., 2023)	68.5	91.3
ExSoftword+BERT (Kong et al., 2021)	67.0	-
BERT+MultiView (Yang et al., 2023)	73.0	92.2
CasRel (Wei et al., 2020)	71.7	89.9
MusREL	73.4	91.2

each type based on the three extraction path strategies. The authors compare the experimental results with the NYT dataset, after training the model for 100 epochs, as shown in Table 3.

The authors find that the performances of using different path strategies for different types of relation extraction are different; for example, in the location relation type, Strategy 1 works significantly better than other strategies, while in the capital relation type, Strategy 2 also significantly outperforms other strategies. The authors find that the correct identification of the subject or the object and the acquisition of semantic features play a key role in the correct judgment of relations. There are also some relations where Strategy 3 works better, especially when the structure of the relations in the type is more complex. And MusREL, because of combining three path strategy features, outperforms the prediction performance of a single strategy under most types in terms of overall prediction effect. This phenomenon verifies the effectiveness of the multi-strategy relationship extraction approach.

Ablation Study

In this part, the authors used ablation experiments on NYT data to verify the validity of each module of MusREL, as shown in Table 4.

The authors found that the accuracy of prediction has a significant decrease when there is no rational entity relation judgment module, because there are a large number of sentences without relation in the sentences, and the lack of this module will make the relation extraction in a large number of unrelated classifications, and such a situation will introduce a large amount of noise, which will affect the prediction results. On the other hand, the problem is more apparent in real commercial scenarios;

Table 3. Representative entity relation classifications

Relation Type	Strategy 1: Subject Entity to Relation and Object Entity RE Strategy	Strategy 2: Object Entity to Relation and Subject Entity RE Strategy	Strategy 3: Parallel Subject and Object Entities to Relation Integration RE Strategy	MusREL
nationality	91.1	92.7	91.4	93.4
location	83.2	93.0	85.8	92.3
place_of_death	95.1	95.1	89.3	95.3
company	89.7	89.1	90.1	90.7
country	92.7	90.1	93.2	93.3
capital	98.7	93.6	94.7	97.1
children	83.3	82.9	84.2	84.7
place_of_birth	94.0	94.1	95.3	94.3

Table 4. Ablation experiments

Approach	NYT		
	P	R	F
MusREL	91.1	94.1	92.6
- Rational entity relation judgment	87.7	93.9	90.7
- Strategy 1: Subject entity to relation & object entity RE strategy	90.3	92.1	91.2
-Strategy 2: Object entity to relation & subject entity RE strategy	90.5	93.1	91.8
-Strategy 3: Parallel subject & object entities to relation integration RE strategy	90.7	91.9	91.3

when the requests increase, RE strategies consume a lot of computational resources on invalid relation prediction. Experimentally, the authors find that the overall throughput of the online system drops to 27%, leaving the system TPS (number of transactions per second) extremely low. In addition, the authors find that with the addition of the multi-strategy weighting module, different tasks have some improvement in the prediction results. Without considering the computational resources, it is still possible to continuously improve the performance of the model by continuously adding strategies. The effectiveness and scalability of this multi-strategy extraction relation architecture are verified.

Comparison of Computational Rates and Computational Resource Requirements

The authors deployed TPLinker, CasRel, and MusREL in commercial scenarios, also using a single CPU (16core8G) + GPU (RTX3090-24G), and found that the number of sentences per second processed by MusREL is much higher than that of the other two methods (i.e., the rate is nearly five times higher than that of the baseline method CasRel). In addition, because MusREL utilizes the effective relation discriminator and rate regulator, it greatly reduces the dependence on GPU matrix operations, whereas the other two methods would not be able to meet the commercial requirements if only CPUs were employed as shown in Table 5.

The MusREL model was originally designed for the issue of real-time relation extraction for commercial scenarios, especially for online education scenarios, providing relation extraction capabilities for multiple upstream applications. Therefore, the authors' model has the advantages of low computational resource consumption, fast response time, and high throughput. The authors' model can run entirely on CPU servers during actual training and deployment. The reliance on expensive GPU resources is reduced. And with the computation rate regulator, some modules of the model can be turned off or downgraded according to the business scenario to further improve the computation rate. Because none of the multiple extraction strategies the authors' chose is weak in individual prediction, the model is still able to maintain good performance. The low-cost and high-performance features are particularly important for the current commercial scenario. In addition, the authors provide a prompt learning method supplement that can provide more support for the relation extraction task in sparse scenarios, using the model's generic knowledge, prompt templates, and answer mapping bag of words to provide supplementary capability support for low-resource relation extraction scenarios.

Other Notes on the Experimental Results

Our MusREL model has a significant improvement in the recall of the experimental results and the speed rate of relation extraction is faster than the mainstream methods in the baseline. From the observation of the result data, it is clear that the multiple strategies the authors use are extracting relations from multiple perspectives, because the paths of extraction and the features to focus on are different, resulting in a model that can find more features overall and can identify relations in the corpus more effectively. The accuracy of the authors' method is improved over the CasREL approach. However, it is not the best among all baselines, partly due to the introduction of noise caused by the increased strategy. However, the authors use the features of each strategy trained using a type-based weight assignment strategy, together with a rational entity relation judgment, to control the effect of

Table 5. Comparison of computational rates and computational resource requirements

Approach	Sentences per Second	Minimum Commercial Computing Resource Requirements
TPLinker	1,113	GPU+CPU
CasRel	617	GPU+CPU
MusREL	3,259	only CPU

this type of noise as much as possible. As a result, the authors' accuracy still remains high among all baselines.

CONCLUSION

In this paper, the authors propose a multi-strategy relation extraction model MusREL, which can provide reliable stability and throughput guarantee for online relation extraction. Using this model, it can better support the relation extraction tasks in teaching plan material structuring, multi-resource educational knowledge graph construction, and educational Q&A scenarios, and it has achieved good practical value. MusREL can provide high-quality structured resources to existing online education systems and promote the balance of regional educational resources to some extent.

The authors designed and optimized three efficient relation extraction strategies and used the weight calculator of different relation types to make the overall f-value of relation extraction reach a high level; by overlaying the relation prediction module, the authors further improved the recall rate of relation extraction and the effectiveness of relation extraction calculation. The relation extraction system architecture designed in this paper is mainly applied to the commercial scenario of online education; it can provide high performance and a highly reliable computing rate for online relation extraction, and the computational model is simple and can even be distributed and deployed with cheap CPU resources, which largely reduces the commercial cost. In future work, in order to further enhance the application scenarios of MusREL, the authors will further explore efficient relation extraction strategies under the open domain to cope with the diversity of entity relations in educational resources so as to obtain more unknown relations in educational resources.

CONFLICTS OF INTEREST

The author declares that there is no conflict of interest regarding the publication of this paper.

FUNDING STATEMENT

This research was funded by the Talent Fund of Keyi College of Zhejiang Sci-Tech University and by the NetEase Frontier Technology Research Project. This work was also supported by Youdao Inc., Zhejiang Sci-Tech University, and Zhejiang University City College.

REFERENCES

Al-Qerem, A., Alauthman, M., Almomani, A., & Gupta, B. B. (2020). IoT transaction processing through cooperative concurrency control on fog–cloud computing environment. *Soft Computing*, *24*(8), 5695–5711. doi:10.1007/s00500-019-04220-y

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. doi:10.1109/TPAMI.2013.50 PMID:23787338

Califf, M. E., & Mooney, R. J. (1997). Relational learning of pattern-match rules for information extraction. *Proceedings of the sixteenth national conference on artificial intelligence*, *328*, 334-341.

Chen, J., Ji, D., Tan, C. L., & Niu, Z. (2005). Unsupervised feature selection for relation extraction. *Companion Volume to the Proceedings of Second International Joint Conference on Natural Language Processing*, 262-267.

Chung, J., Caglar, G., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *NIPS 2014 Workshop on Deep Learning*, 1-9.

Culotta, A., & Sorensen, J. (2004). Dependency tree kernels for relation extraction. *Proceedings of the 42nd Annual Meeting of the Association for Computational Linguistics*, 423-429.

Diederik, P. K., & Ba, J. (2015). Adam: A method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations, 5*, 6.

Dong, L., Yang, N., Wang, W., Wei, F., Liu, X., Wang, Y., Gao, J., Zhou, M., & Hon, H. (2019). Unified language model pre training for natural language understanding and generation. In H. M. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. B. Fox, & R. Garnett (Eds.), Advances in neural information processing systems 32*: Proceedings of the annual conference on neural information processing systems* (pp. 13042-13054). Academic Press.

Elgendy, I. A., Zhang, W. Z., He, H., Gupta, B. B., & Abd El-Latif, A. A. (2021). Joint computation offloading and task caching for multi-user and multi-task MEC systems: Reinforcement learning-based algorithms. *Wireless Networks*, *27*(3), 2023–2038. doi:10.1007/s11276-021-02554-w

Fundel, K., Küffner, R., & Zimmer, R. (2006). RelEx-relation extraction using dependency parse trees. *Bioinformatics, 23*(3), 365-371.

Gardent, C., Shimorina, A., Narayan, S., & Perez-Beltrachini, L. (2017). Creating training corpora for NLG micro-planners. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, 1*, 179-188. doi:10.18653/v1/P17-1017

Gaurav, A., Hu, B., Choi, C., & Almomani, A. (2022). Evaluation and comparative analysis of semantic web-based strategies for enhancing educational system development. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–14.

George, G., & Lal, A. M. (2021). A personalized approach to course recommendation in higher education. *International Journal on Semantic Web and Information Systems*, *17*(2), 100–114. doi:10.4018/IJSWIS.2021040106

Gupta, B. B., Gaurav, A., & Panigrahi, P. K. (2023). Analysis of the development of sustainable entrepreneurship practices through knowledge and smart innovative based education system. *The International Entrepreneurship and Management Journal*, *19*(2), 923–940. doi:10.1007/s11365-023-00853-6

Gupta, B. B., & Quamara, M. (2020). An overview of Internet of Things (IoT): Architectural aspects, challenges, and protocols. *Concurrency and Computation*, *32*(21), e4946. doi:10.1002/cpe.4946

Huffman, S. B. (1995). Learning information extraction patterns from examples. *Proceedings of IJCAI*, 246-260.

Juárez-Varón, D., Bellido-García, I., & Gupta, B. B. (2023). Analysis of stress, attention, interest, and engagement in onsite and online higher education: A neurotechnological study. *Comunicar*, *31*(76), 21–33. doi:10.3916/C76-2023-02

Kar, R. (2022). To study the impact of social network analysis on social media marketing using graph theory. *International Journal of Software Science and Computational Intelligence*, *14*(1), 1–20. doi:10.4018/IJSSCI.304437

KatzD. M.BommaritoM. J.GaoS.ArredondoP. D. (2023). Gpt-4 passes the bar exam. *Available at* SSRN 4389233.

Kaur, M., Singh, D., Kumar, V., Gupta, B. B., & Abd El-Latif, A. A. (2021). Secure and energy efficient-based E-health care framework for green internet of things. *IEEE Transactions on Green Communications and Networking*, *5*(3), 1223–1231. doi:10.1109/TGCN.2021.3081616

Kim, J. T., & Moldovan, D. I. (1995). Acquisition of linguistic patterns for knowledge-based information extraction. *TKDE*, *7*(5), 713–724.

Kong, B., Liu, S., Wei, F., Jia, L., & Wang, G. (2021). Chinese relation extraction using extend Softword. *IEEE Access : Practical Innovations, Open Solutions*, *9*, 110299–110308. doi:10.1109/ACCESS.2021.3102225

Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data. *Proceedings of the International Conference on Machine Learning (ICML'01)*, 282-289.

Lee, J., Seo, S., & Choi, Y. S. (2019). Semantic relation classification via bidirectional LSTM networks with entity-aware attention using latent entity typing. *Symmetry*, *11*(6), 785. doi:10.3390/sym11060785

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2020). Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *Proceedings of the 58th Annual Association for Computational Linguistics*, 7871-7880. doi:10.18653/v1/2020.acl-main.703

Li, P., Zhong, H., & Zhang, J. Z. (2022). Spatial patterns and development characteristics of China's postgraduate education: A geographic information system approach. *International Journal on Semantic Web and Information Systems*, *18*(1), 1–21. doi:10.4018/IJSWIS.313190

Li, Z., Ding, N., Liu, Z., Zheng, H. T., & Shen, Y. (2019). Chinese relation extraction with multi-grained information and external linguistic knowledge. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 4377-4386. doi:10.18653/v1/P19-1430

Lin, Y., Shen, S., Liu, Z., Luan, H., & Sun, M. (2016). Neural relation extraction with selective attention over instances. *Proc. ACL,* 2124-2133. doi:10.18653/v1/P16-1200

Mintz, M., Bills, S., Snow, R., & Jurafsky, D. (2009) Distant supervision for relation extraction without labeled data. *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP,* 1003-1011. doi:10.3115/1690219.1690287

Nakashole, N., Weikum, G., & Suchanek, F. (2012). PATTY: A taxonomy of relational patterns with semantic types. *Proceedings of EMNLP-CoNLL*, 1135-1145.

Nayak, T., & Ng, H. T. (2020). Effective modeling of encoder-decoder architecture for joint entity and relation extraction. *The Thirty- Fourth AAAI Conference on Artificial Intelligence,* 8528-8535.

Nebhi, K. (2013). A rule-based relation extraction system using DBpedia and syntactic parsing. *Proceedings of the 2013th International Conference on NLP & DBpedia, 1064*, 7479.

Nédellec, C. (2005). Learning language in logic-genic interaction extraction challenge. *Proceedings of the 4th Learning Language in Logic Workshop*, 31-37.

Och, F. J., Gildea, D., Khudanpur, S., Sarkar, A., Yamada, K., Fraser, A., Kumar, S., Shen, L., Smith, D., Eng, K., Jain, V., Jin, Z., & Radev, D. (2004). A smorgasbord of features for statistical machine translation. *Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics*, 161-168.

Peters, M., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K., & Zettlemoyer, L. (2018). Deep contextualized word representations. *Proceedings of the Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1*, 2227-2237.

Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., Zhou, Y., Li, W., & Liu, P. (2020). Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of Machine Learning Research*, *21*(1), 5485–5551.

Riedel, S., Yao, L., & McCallum, A. (2010). Modeling relations and their mentions without labeled text. *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 148-163. doi:10.1007/978-3-642-15939-8_10

Singla, A., Gupta, N., Aeron, P., Jain, A., Garg, R., Sharma, D., Gupta, B. B., & Arya, V. (2022). Building the metaverse: Design considerations, socio-technical elements, and future research directions of metaverse. *Journal of Global Information Management*, *31*(2), 1–28. doi:10.4018/JGIM.315283

Sun, K., Zhang, R., Mensah, S., Mao, Y., & Liu, X. (2020). Recurrent interaction network for jointly extracting entities and classifying relations. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, (EMNLP)*, 3722-3732. doi:10.18653/v1/2020.emnlp-main.304

Sun, K., Zhang, R., Mensah, S., Mao, Y., & Liu, X. (2021). Progressive multi-task learning with controlled information flow for joint entity and relation extraction. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(15), 13851–13859. doi:10.1609/aaai.v35i15.17632

Tian, X., Jing, L., He, L., & Liu, F. (2021). StereoRel: Relational triple extraction from a stereoscopic perspective. *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, 1*, 4851-4861. doi:10.18653/v1/2021.acl-long.375

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Proceedings of NIPS*, 5998-6008.

Wang, L., Cao, Z., Melo, G., & Liu, Z. (2016). Relation classification via multi-level attention CNNs. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 1*, 1298-1307. doi:10.18653/v1/P16-1123

Wang, Y., Yu, B., Zhang, Y., Liu, T., Zhu, H., & Sun, L. (2020). TPLinker: Single-stage joint extraction of entities and relations through token pair linking. *Proceedings of the 28th International Conference on Computational Linguistics*, 1572-1582. doi:10.18653/v1/2020.coling-main.138

Wei, Z., Su, J., Wang, Y., Tian, Y., & Chang, Y. (2020). A novel cascade binary tagging framework for relational triple extraction. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 1476-1488. doi:10.18653/v1/2020.acl-main.136

Xu, J., Wen, J., Sun, X., & Su, Q. (n.d.). A discourse-level named entity recognition and relation extraction dataset for Chinese literature text. *Training (New York, N.Y.)*, 604.

Yang, J., Ji, B., Li, S., Ma, J., Peng, L., & Yu, J. (2023). Dynamic multi-view fusion mechanism for Chinese relation extraction. In H. Kashima, T. Ide, & W. C. Peng (Eds.), Lecture notes in computer science: Vol. 13935. *Advances in knowledge discovery and data mining. PAKDD 2023*. doi:10.1007/978-3-031-33374-3_32

Yu, B., Zhang, Z., Shu, X., Liu, T., Wang, Y., Wang, B., & Li, S. (2019). Joint extraction of entities and relations based on a novel decomposition strategy. *ECAI*, 2282-2289.

Yuan, Y., Zhou, X., Pan, S., Zhu, Q., Song, Z., & Guo, L. (2020). A relation-specific attention network for joint entity and relation extraction. *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 4*, 4054-4060. doi:10.24963/ijcai.2020/561

Zeng, D., Liu, K., Chen, Y., & Zhao, J. (2015) Distant supervision for relation extraction via piecewise convolutional neural networks. *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing,* 1753-1762. doi:10.18653/v1/D15-1203

Zeng, X., Zeng, D., He, S., Liu, K., & Zhao, J. (2018). Extracting relational facts by an end-to-end neural model with copy mechanism. *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 1*, 506-514. doi:10.18653/v1/P18-1047

Zhang, D., & Wang, D. (2015). Relation classification via recurrent neural network. *Big Data Mining and Analytics*, *1*(3), 234–244.

Zhang, M., Zhou, G., & Aw, A. (2008). Exploring syntactic structured features over parse trees for relation extraction using kernel methods. *Information Processing & Management*, *44*(2), 687–701. doi:10.1016/j.ipm.2007.07.013

Zhao, H., Luo, Z., Feng, C., Zheng, A., & Liu, X. (2019) A context-based framework for modeling the role and function of online-resource citations in scientific literature. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 5206-5215.

Zhou, G., Zhang, M., Ji, D., & Zhu, Q. (2007). Tree kernel-based relation extraction with context-sensitive structured parse tree information. *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 728-736.

Zhou, P., Shi, W., Tian, J., Qi, Z., Li, B., Hao, H., & Xu, B. (2016) Attention-based bidirectional long short-term memory networks for relation classification. *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, 2*, 207-212. doi:10.18653/v1/P16-2034