CJE-PCHF: Chinese Joint Entity and Relation Extraction Model based on Progressive Contrastive Learning and Heterogeneous Feature Fusion

Meng He , Yunli Bai * , Dongye Wei

Posted Date: 30 July 2024

doi: 10.20944/preprints202407.2384.v1

Keywords: joint extraction; contrastive learning; heterogeneous features; interactive fusion; semantic association

 doi:10.20944/preprints202407.2384.v1

Article

CJE-PCHF: Chinese Joint Entity and Relation Extraction Model based on Progressive Contrastive Learning and Heterogeneous Feature Fusion

Meng He [1], Yunli Bai [2,*] and Dongye Wei [3]

College of Computer and Information Engineering, Inner Mongolia Agricultural University, Hohhot 010018, China; heme996@emails.imau.edu.cn (M.H.); baiyl@imau.edu.cn (Y.B.); weidy@emails.imau.edu.cn (D.W.)
* Correspondence: baiyl@imau.edu.cn

Abstract: The joint extraction of entities and relations is a critical task in information extraction, and its performance directly affects the performance of downstream tasks. However, existing joint extraction models based on deep learning exhibit weak processing capabilities for the phenomenon of multiple pronunciations of one character and multiple characters of one pronunciation when processing Chinese texts, resulting in performance loss. To address these issues, this paper introduces part-of-speech (POS) and pinyin features to aid the model in learning semantic features that are more contextually appropriate. A Chinese Joint Entity and Relation Extraction Model based on Progressive Contrastive Learning and Heterogeneous Feature Fusion is proposed (CJE-PCHF). During model training, an interactive fusion network based on progressive contrastive learning is employed to learn the dependencies between pinyin, POS, and semantic features. This guides the model in heterogeneous feature fusion, capturing higher-order semantic associations between heterogeneous features. On the commonly used DuIE evaluation dataset for joint extraction, this model achieved a significant improvement, with the F1 score increasing by 5.4% compared to the benchmark model CasRel.

Keywords: joint extraction; contrastive learning; heterogeneous features; interactive fusion; semantic association

1. Introduction

In the field of natural language processing, entity relation extraction, as a key subtask, is crucial for applications such as knowledge graph construction[1], information retrieval, and semantic understanding. In order to solve the problem of entity relation extraction, researchers have proposed pipeline models and joint models[2]. The pipeline model is divided into two stages: first extracting entities and then extracting relations. However, this method ignores the interaction between the two models, and the noise output of the entity model will directly affect the quality of the relation extraction stage and thus limit the accuracy of information extraction. The joint extraction model unifies the entity recognition task and the relation extraction task for joint modeling and training, guides the model to mine the interactive information between the two tasks, improves the model effect, and also alleviates the impact of the secondary transmission of errors on the model.

The joint extraction model has achieved good performance on English corpus, but the model that relies only on semantic features faces challenges in processing the complex semantic information of Chinese text context[3], which limits the improvement of the model in extracting triples. To address this problem, this paper introduces pinyin and POS features to enable the model to learn the multi-dimensional features of text sequences and better understand the text semantics. The introduction of pinyin features helps to alleviate the model performance loss caused by homonymy; POS features help the model to mine coarse-grained semantic features in text sequences[4]. To this end, this paper proposes an interactive fusion network based on progressive contrastive learning, which aims to fuse pinyin features and POS features with semantic features respectively, and optimize the distribution of multiple features in the feature space. Contrastive learning achieves the fusion of heterogeneous

 doi:10.20944/preprints202407.2384.v1

features by strengthening the similarity between similar samples[5]. The contrastive learning module mines and strengthens the intrinsic correlation and distinction between these features, aiming to assist the model in learning semantic features that are more in line with the context.

In summary, the work of this paper is mainly reflected in the following three aspects:

(1) To address the problem that the current joint extraction model does not adequately model the prior information of text sequences, this paper introduces POS and pinyin features during the model training process to assist the model in more accurately learning semantic features that are more consistent with contextual associations.

(2) To address the problem of integrating phonetic, POS features with semantic features, this paper proposes an interactive fusion network based on progressive contrastive learning. This network uses progressive contrastive learning to constrain the heterogeneous alignment of phonetic-semantic and POS-semantic fusion features. In the optimization stage of the model, the contrast loss is incorporated into the model optimization, and a hyperparameter is introduced to balance the contrast loss and the model self-training loss.

(3) Experimental results show that on the commonly used joint extraction evaluation dataset DuIE, compared with the baseline model CasRel,the proposed model has achieved an improvement in F1 score of 5.4%. The effectiveness of the proposed method is verified through ablation experiments.

2. Related Work

In the field of information extraction, merging entity recognition and relation classification tasks into one step can effectively reduce the negative impact of error propagation, strengthen the correlation between the two tasks, and thus improve the overall task extraction efficiency. At present, there are two main strategies for merging joint extraction tasks: parameter sharing and sequence labeling[6]. Miwa et al.[7]adopted a parameter sharing approach and proposed an end-to-end model based on long short-term memory networks (LSTM). This model combines the bidirectional LSTM of word sequences with the bidirectional LSTM of dependency syntax trees to capture the dependency information between word sequences and dependency syntax trees, jointly model entities and relations in a single model, and realize the joint representation of entities and relations. Katiyar et al.[8]proposed an RNN model based on attention mechanism. This model does not rely on dependency syntax trees, but uses LSTM and attention mechanisms to directly extract entities and relations from text. Zheng et al.[9]proposed a new joint extraction strategy that transforms entity extraction and relation classification into sequence labeling tasks, providing new ideas for subsequent joint extraction research. Zeng et al.[10]proposed an extraction model based on the copy mechanism. The encoder first encodes the text sequence, and a unified decoder or multiple independent decoders are used in the decoding process to extract entities and their relations in the text sequence. Fu et al.[11]designed an extraction model called GraphRel through a graph convolutional network. The model uses linear sequences and dependency structures to extract text features, and uses a complete word graph to extract implicit features between word pairs. Wei et al.[12]proposed the CasRel model, which uses a pointer annotation method. First, two binary classifiers are used to identify all subject entities, and then the corresponding object entities and their existing relations are predicted. However, its modeling of prior information in the text is not sufficient, which to a certain extent loses the prediction accuracy of the subject entity, thereby affecting the quality of joint extraction. Based on this consideration, this paper introduces prior information to enhance the modeling of text features to improve the extraction accuracy of the model.

As a pictographic language, Chinese contains a large amount of feature information[13], such as Wubi features, glyph features, word boundary features, POS features, and pinyin features.These feature information together constitute the rich semantic information of Chinese characters. In recent years, the field of Chinese feature fusion has attracted the attention of researchers. Zhang et al.[14]proposed the SSP2Vec method, which generates feature substrings by integrating the strokes, structural features, and pinyin information of Chinese characters, effectively improving the semantic capture ability of word embedding. Wang et al.[15]proposed a named entity recognition model that

 doi:10.20944/preprints202407.2384.v1

integrates POS information based on the characteristics of microblog text. The POS information is combined with the word embedding vector and then input into a bidirectional long short-term memory neural network, and the conditional random field is used to decode the output to achieve POS feature-assisted named entity recognition. Zhu et al.[16]introduced the pinyin features of Chinese characters and verified through experiments that the introduction of pinyin features can effectively improve the performance of existing models in text recognition. Wu et al.[17]MECT multivariate metadata embedding cross transformer to address the problem of insufficient utilization of Chinese character structural information. The MECT combines Chinese character features with radical-level embedding in a two-stream[4]proposed a named entity recognition method that combines pinyin and POS features to address the phenomenon of multiple pronunciations of one character and multiple characters of one pronunciation in Chinese medical records. The model performance was improved by effectively integrating the glyph, POS and pinyin features through the scaled dot product attention module. Studies have shown that the introduction of prior features in the Chinese named entity recognition task has a positive effect on the performance improvement of existing models. Named entity recognition, as an important part of the joint extraction task, provides a new idea for the joint extraction model when processing Chinese text. Yu et al.[18]proposed the AMFRel joint extraction model, which generates an encoding vector containing POS information by extracting electronic medical record text and POS features. After subject extraction, the model uses the information fusion module to enhance the text structure features and extracts the object according to specific relations, finally realizing the extraction of medical text triples. Li et al.[19]on the SpERT framework, combining word embedding with POS features. The feature fusion strategy was used to solve the redundancy problem in the entity extraction process based on the span method, while improving the accuracy of joint extraction. Ge et al.[20]combined word vectors with character vectors, incorporated position information to ensure the correct word order, and used a hierarchical labeling strategy to effectively mark multiple relations and object entities based on the main entity information, thereby achieving effective extraction of overlapping relations. Based on the above analysis, this paper considers introducing POS features and pinyin features in the joint extraction task, and designs a corresponding feature fusion mechanism to integrate it into the model training process, thereby improving the model's text representation ability and enhancing model performance.

Contrastive learning is a machine learning technique that strengthens the similarity between similar samples and expands the difference between different samples[21], which helps to improve the joint extraction performance of the model. Its core concept is to organize the embedding representation of samples in the latent space to ensure that the embeddings of the same or semantically related samples (positive sample pairs) are closely adjacent, while the embeddings of different or unrelated samples (negative sample pairs) are as far apart as possible. In recent years, contrastive learning has shown good performance in natural language processing tasks. Mikolov et al.[22]first introduced the contrastive learning framework into the field of natural language processing, using co-occurring words as the basis for measuring semantic similarity and effectively learning the embedding representation of words through negative sampling technology. This method improves the representation quality of words and phrases in high-dimensional space, enabling them to more accurately reflect their semantic features. Later, Fang et al.[23]proposed the CERT model, a pre-trained language representation model based on contrastive learning. The model enhances language understanding ability through sentence-level contrastive learning and provides support for natural language processing tasks. In response to the problem that existing joint extraction models do not make sufficient use of context information, Lei et al.[21]designed the JERCE model. The model extracts semantic features between sentences and entities through contrastive learning, strengthens the representation of entities and relations, and uses a weighted loss function to optimize the balance between entity recognition and relation extraction tasks. Based on the above analysis, this paper considers introducing a contrastive learning strategy in the joint extraction task, which can effectively improve the model performance by optimizing the embedding representation of instances.

 doi:10.20944/preprints202407.2384.v1

In summary, this paper proposes an interactive fusion network based on progressive contrastive learning. To address the problem of insufficient prior information modeling, the model uses POS features and pinyin features to assist the model in learning prior features, and designs a corresponding feature fusion mechanism during the introduction process to integrate it into the model training process, and uses progressive contrastive learning constraint optimization to further enhance the quality of feature fusion, so that the fused features are distributed and aligned in the feature space. Hyperparameters are introduced for loss adaptation to balance the model self-training loss and contrastive learning constraint loss, thereby improving the overall performance of the model.

3. Method Architecture

3.1. Model Overview

In view of the fact that the existing joint extraction model based on deep learning has a weak ability to handle the phenomenon of multiple pronunciations of one character and multiple characters of one pronunciation when processing Chinese texts, which leads to the performance loss of the model, this paper proposes an interactive fusion network based on progressive contrastive learning, and uses this network in the joint extraction task, and proposes a joint extraction model based on progressive contrastive interactive fusion, as shown in the Figure 1 shows:

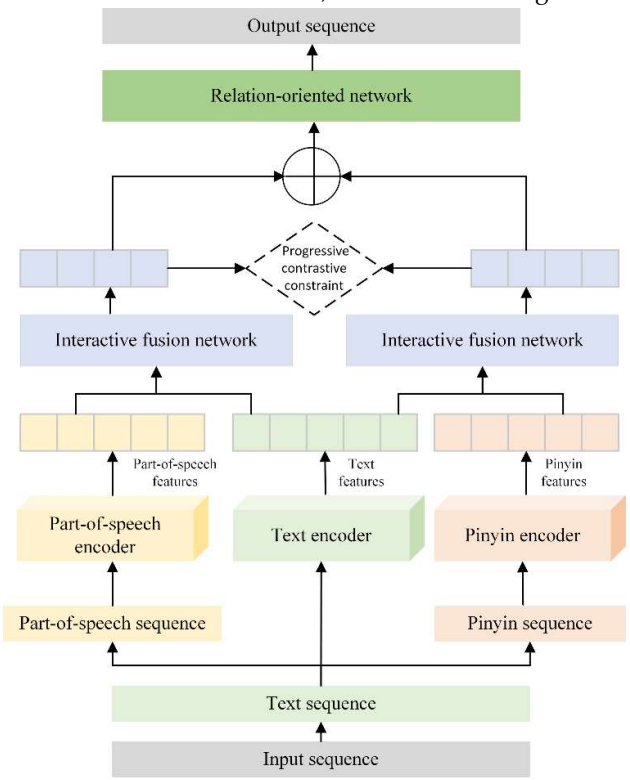

Figure 1. The overall structure of CJE-PCHF.

As can be seen from the figure1, the network is mainly divided into the encoding stage, feature fusion stage and decoding stage. In the encoding stage, the text sequence, the corresponding POS sequence and the pinyin sequence are feature embedded to obtain the corresponding feature vector. In the feature fusion stage, the interactive fusion network based on progressive contrastive learning proposed in this paper is used for feature fusion to obtain higher quality text features. In the decoding stage, the relation-oriented network of the CasRel model is used for decoding, which will not be described in detail in this section. The specific modeling process is shown in other summaries of this chapter.

 doi:10.20944/preprints202407.2384.v1

3.2. Semantic Encoding

The feature encoding involved in this article is divided into three parts, namely text semantic encoding, POS feature encoding, and pinyin feature encoding. In the semantic encoding stage, given a text sequence: $X=[x_1,x_2,...x_i,...x_n]$, x_i represents the character with index i in the text sequence, and n represents the length of the text sequence, as shown in the formula(1):

$$h_x^i = BERT(x_i) \quad (1)$$

Among them, $BERT()$ is the encoding function of text semantics. The BERT pre-trained model is used in the embedding layer of the model to generate dynamic word vectors for text sequences. This paper chooses BERT as the semantic encoder of text sequences, mainly based on the following reasons: First, compared with traditional unidirectional language models, BERT can capture the contextual semantic information of the text and show better semantic understanding ability. Secondly, although the BERT model has fewer parameters, it can still generate high-quality semantic representations and quickly adapt to downstream tasks through fine-tuning after pre-training. Based on BERT's bidirectional encoding mechanism, Transformer structure, and pre-training results on large-scale corpus, this paper can achieve high-quality text semantic encoding while maintaining a low computational cost. Therefore, this paper chooses it as the semantic encoder of text.

3.3. Pinyin Encoding

Each Chinese character has its corresponding pinyin and specific meaning. Pinyin, as a kind of phonological information, can provide additional feature dimensions for the model, thereby enriching the input feature space of the model. By incorporating the pinyin information of Chinese characters into the embedding layer, the model can learn more dimensional prior features. Before pinyin encoding the text sequence, you first need to build a pinyin word list. This word list contains all possible pinyins and assigns a unique index number to each pinyin. This process is actually to build a traversal dictionary and index it. The purpose of building a pinyin word list is to map the pinyin sequence to the corresponding index number in subsequent processing. After obtaining the pinyin word list, you can use the pypinyin tool to obtain the pinyin sequence corresponding to the text sequence X. $X_{\text{py}}=[m_1,m_2,...m_i,...m_n]$, m_i Represents the pinyin ID of the i-th token in the text sequence, that is, its index in the pinyin word list. The specific encoding formula is shown in the formula(2):

$$h_{\text{py}}^i = e_{py}^x(m_i) \quad (2)$$

Here, $e_{py}^x(\ )$ denotes the pinyin encoding function, essentially an initialized vector lookup table. Through this lookup table, pinyin indices are mapped to a continuous vector space. For instance, the pinyin of the i-th character can be converted into a 768-dimensional pinyin vector h_{py}^i. These vectors are randomly initialized at the onset of training to form the corresponding pinyin encoding space, which is then progressively fine-tuned during model training, ultimately resulting in a refined pinyin encoding space, the corresponding pinyin weights.

3.4. POS Encoding

In the training of NLP models, POS information constitutes a vital grammatical feature, aiding models in comprehending and generating text that adheres to grammatical norms. To fully exploit POS information during model training, it is imperative to first acquire POS tags for text sequences. Stanford University's NLP tool provides robust POS tagging functionality, accurately assigning tags to text. Given a text sequence X, this tool can generate its corresponding POS sequence X_{pos}. $X_{pos}=[p_1,p_2,...p_i,...p_n]$. In this process, each word is assigned a POS tag, such as noun, verb, or adjective. To ensure that POS information aligns with the text granularity in the semantic encoding process of the BERT model, this paper aligns POS tags at the character level, ensuring that both the POS and semantic information of each character are precisely represented within the model. This alignment is particularly crucial for Chinese, a language characterized by character-level

 doi:10.20944/preprints202407.2384.v1

independence and rich semantic content. Specifically, each character of a word is assigned the corresponding POS tag, ensuring consistency with BERT encoding. Here, p_i denotes the POS tag of the i-th character of text X, achieving alignment along the sequence dimension, the encoding process as shown in the formula(3):

$$h_{pos}^{i} = e_{pos}^{x}(p_i) \quad (3)$$

Here, $e_{\text{pos}}^{x}(\ )$ represents the encoding function for POS information, essentially an initialized vector lookup table used to convert discrete POS tags into vector representations in a continuous vector space. For instance, the POS tag of the i-th character is transformed into a 768-dimensional POS vector h_{pos}^{i}. This table is randomly initialized at the commencement of training and fine-tuned during task-specific training to better capture the relations between POS tags and task-relevant features, ultimately yielding an optimized POS encoding space, the corresponding POS weights.

3.5. Feature Fusion Coding

Semantic coding, POS coding and pinyin coding features are three important features, which provide feature information of different dimensions respectively. Since these three types of features are heterogeneous features, direct feature fusion is likely to cause feature distortion and affect the performance of the model. As for the fusion method of multiple features, the current main method is direct splicing, which leads to a lack of focus in representation learning. Therefore, this paper proposes an interactive fusion network based on progressive contrastive learning, which can effectively avoid feature distortion through layer-by-layer optimization. In the optimization process of each layer, the mechanism of contrastive learning is used to minimize the similarity loss and maximize the difference loss, so that the fused features are more discriminative in high-dimensional space, thereby improving the overall performance of the model. Specifically, the interactive fusion mechanism is divided into two parts: the pinyin -semantic fusion part and the POS -semantic fusion part. The interactive fusion network diagram is shown in Figure 2:

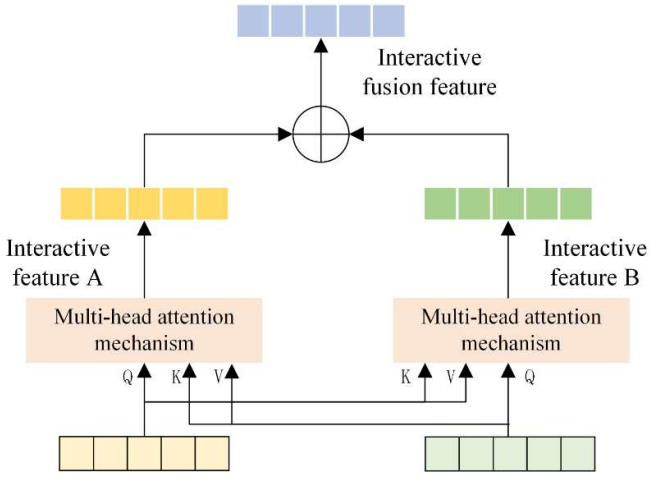

Figure 2. The interactive fusion network diagram.

3.5.1. Pinyin-Semantic Fusion

The interactive fusion network diagram of this paper is shown in Figure 2, where K is the key vector matrix, V is the corresponding value vector matrix, and Q is corresponding the query vector matrix. First, the multi-head attention mechanism is used to mine features with interactive information from pinyin features and semantic features, recorded as interactive features A and interactive features B, and then the correlation vector between pinyin features and semantic features is obtained. The obtained correlation information vector is used as the weight matrix, the pinyin feature itself is weighted, and the corresponding semantic information of the POS encoding is given;

 doi:10.20944/preprints202407.2384.v1

at the same time, the correlation between the semantic feature and the pinyin feature is obtained, and the obtained correlation information vector is used as the weight matrix, the semantic feature itself is weighted, and the pinyin encoding is given the corresponding POS information; then the two weighted vectors are concatenated and fused, and finally the concatenated vector is reduced in dimension to obtain contextual features containing high-order semantic associations between semantic features and pinyin features. Finally, in order to reduce the feature dimension, dimensionality reduction processing is performed to obtain the final fused feature vector. Among them, $H_{py}=[h_{py}^1,h_{py}^2,...h_{py}^i,...h_{py}^n]$ represents the sequence pinyin embedding feature, $H=[h_x^1,h_x^2,...h_x^i,...h_x^n]$ represents the sequence semantic embedding feature, and the pinyin-semantic fusion is shown in the formula(4):

$$H_{py}^h=Cross-ATT\left(H,H_{py}\right) \quad (4)$$

This method effectively integrates semantic features and phonetic features, and can better capture the relation between phonetics and semantics, thereby improving the model's ability to understand the semantics of text context.

3.5.2. POS-Semantic Fusion

The design of the POS-semantic fusion part is similar to that of the pinyin-semantic fusion part, but it focuses on the correlation between POS features and semantic features. Through the multi-head attention mechanism, features with interactive information are mined from the pinyin features and semantic features, recorded as interactive features A and interactive features B. Then, the correlation vector between the POS features and the semantic features is obtained, and the obtained correlation information vector is used as the weight matrix to weight the POS features themselves, and the corresponding semantic information is given to the POS encoding; at the same time, the correlation vector between the semantic features and the POS features is obtained, and the obtained correlation information vector is used as the weight matrix to weight the semantic features themselves, and the corresponding POS information is given to the semantic encoding; then the two weighted vectors are concatenated and fused to form a context feature that contains high-order semantic associations between semantic features and POS features. Finally, in order to reduce the feature dimension, dimensionality reduction processing is performed to obtain the final fused feature vector. Among them, $H_{\text{pos}}=[h_{pos}^1,h_{pos}^2,...h_{pos}^i,...h_{pos}^n]$ represents the sequence POS embedding feature, $H=[h_x^1,h_x^2,...h_x^i,...h_x^n]$ represents the sequence semantic embedding feature, and the POS-semantic fusion is shown in the formula(5):

$$H_{\text{pos}}^h=Cross-ATT\left(H,H_{pos}\right) \quad (5)$$

By interactively fusing POS features and semantic features, the high-order correlation between semantic information and POS information can be retained to a large extent, thereby improving the model's ability to understand text.

3.6. Progressive Contrast Constraint

After obtaining the feature vectors of the interactive fusion of pinyin and semantics, and the feature vectors of the interactive fusion of part of speech and semantics, in order to effectively constrain the model to achieve feature alignment for these three types of heterogeneous features, this paper proposes a progressive contrastive learning constraint. Its learning framework is shown in Figure 3.

 doi:10.20944/preprints202407.2384.v1

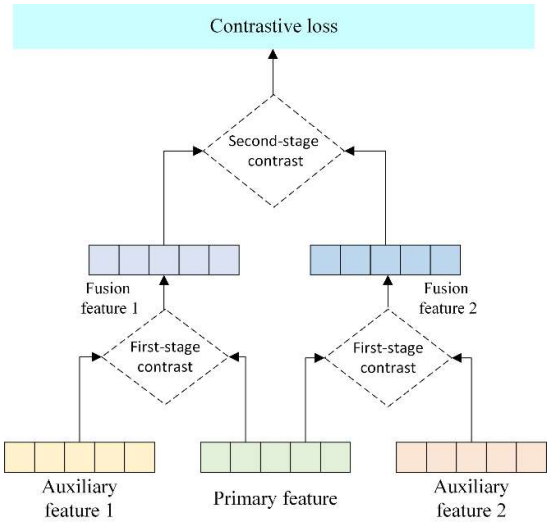

Figure 3. The framework of Progressive contrastive learning constraint.

This paper proposes a two-stage contrastive learning framework to address the heterogeneity problem between the main features and auxiliary features (including auxiliary feature 1—pinyin features, auxiliary feature 2—POS features) in text processing, aiming to effectively integrate these features and improve the learning efficiency and stability of the model.

Phase 1 comparison: interactive feature fusion. In this phase, this paper first focuses on how to promote effective interaction between features while maintaining feature discriminability. By designing a multi-head attention mechanism, the main feature (the most discriminative semantic feature) and the auxiliary feature generate interactive feature A and interactive feature B respectively. Specifically, this mechanism promotes the interaction between the pinyin feature and the semantic feature to form the pinyin-semantic interactive fusion feature, and the POS feature and the semantic feature to interact to form the POS-semantic interactive fusion feature. As described in the summary of 3.5, this process enhances the information expression within each feature and lays the foundation for subsequent deep integration. Phase 2 comparison: contrast constraint optimization. The interactive fusion features generated in the first phase are further optimized through the contrast constraint mechanism. Specifically, the pinyin-semantic interactive fusion feature (fusion feature 1) and the POS-semantic interactive fusion feature (fusion feature 2) are placed in the same contrast learning framework to enhance the model's ability to recognize the differences and commonalities of heterogeneous features. This comparison not only consolidates the mutual understanding and complementation between features, but also promotes the model's comprehensive grasp of complex contextual information, ensuring excellent generalization performance in a diverse feature space.

This progressive contrastive learning strategy improves the stability of the model when processing heterogeneous features by gradually deepening the interaction and fusion of features in stages, while overcoming the training oscillation phenomenon caused by feature heterogeneity when directly minimizing the cosine distance. In addition, by gradually introducing contrast constraints, the model is guided to learn a heterogeneous feature fusion mode that is more in line with the text context, thereby demonstrating the ability to distinguish and generalize in the feature space, and improving the overall performance of the model. This method provides new ideas and practical paths for processing the fusion of heterogeneous features.

3.7. Loss Function

In order to optimize the model's fusion of heterogeneous features, this paper designs a multi-task loss function. This loss function is optimized by the self-training loss and contrastive learning constraint loss of the linear combination model, and the sum of the weights of the two is always kept as 1. Among them, the contrastive learning constraint loss function is shown in the formula(6):

 doi:10.20944/preprints202407.2384.v1

$$loss_{contrast} = F_{contrast}\left(H^{h}_{\text{pos}}, H^{h}_{py}\right) \quad (6)$$

The cosine distance similarity is used to calculate the cosine value of the angle between the POS feature vector and the pinyin feature vector. The closer the value is to 1, the more similar the two vectors are, thereby guiding the model to learn more consistent feature representations.

4. Experiment

4.1. Dataset

This paper uses the DuIE dataset[24]for experimental evaluation. The DuIE dataset is a large-scale information extraction dataset in the Chinese domain, covering more than 210,000 real-world Chinese sentences and their contained relational triples. The dataset predefines 49 common relation types, such as "author", "graduation school", and "starring". All data are sourced from Baidu Encyclopedia and Baidu News Abstracts. Each data consists of a sentence and the relational triples contained in the sentence. The wide coverage and diversity of the DuIE dataset make it an ideal dataset for evaluating the performance of information extraction models. DuIE dataset is shown in Table1:

Table 1. DuIE dataset.

Data	Training	Training Triplets	Validation	Validation Triplets
DuIE	173108	349266	21639	43739

4.2. Experimental Environment and Parameter Settings

Table 2. The experimental environment settings.

Experimental Environment	Configuration
Operating system	Windows 10
Programming language	Python 3.8.19
Deep learning frameworks	Pytorch 1.13.0
Graphics	Tesla V100 GPU
Memory	128GB

Table 3. The experimental parameter settings.

Parameter	Parameter Value
Bert dimension	768
Learning rate	1e-5
Batch size	32
Maximum text length	300
Pinyin embedding dimension	768
POS embedding dimension	768
Dropout parameters	0.1
Transformer layers	12
Loss balance coefficient	0.7
Epoch	100

4.3. Evaluation Indicators

In order to accurately measure and effectively evaluate the experimental results, this study selected precision (P), recall (R) and F1 value as the core evaluation indicators, and loaded the optimal training model to predict the entity relation triples in the test set. The higher the value of the evaluation indicator, the better the performance of the corresponding method. The definitions of each evaluation indicator are given below: TP, TN, FP, FN represent true positive examples, true negative

 doi:10.20944/preprints202407.2384.v1

examples, false positive examples and false negative examples respectively. The specific calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (9)$$

4.4. Baseline Model

In this section, the proposed model will be experimentally compared with the following seven models on the DuIE dataset.

(1) MultiR[25]is an entity relation extraction framework based on multi-instance joint learning. It implements weakly supervised learning with the help of corpus resources and is good at processing and extracting triples with overlapping relations.

(2) The CoType model[26]achieves the joint extraction of entity relations in text by unifying entity recognition and relation classification tasks into a global sequence labeling framework.

(3) The pointer annotation model[27]innovatively uses the pointer mechanism to directly identify the constituent elements of the relation triple, and integrates the co-entity features and attention mechanism to enhance the extraction performance of the model.

(4) The FETI method[28]optimizes the prediction process of relation triples by integrating the category information of the subject and the object. At the same time, it implements constraints on entity categories in the prediction stage, thereby improving the accuracy and specificity of the prediction.

(5) CasRel[12]uses the head and tail pointer strategy to simultaneously predict entity pairs and their relations, effectively dealing with the complex situation of overlapping relation triples extraction and improving the extraction accuracy.

(6) The word-character hybrid model[20]combines character-level and word-level representations to reduce the impact of Chinese word segmentation boundary errors on model performance, and adopts a hierarchical labeling strategy to effectively solve the problem of relation overlap extraction, thereby improving the robustness of the model.

(7) BSCRE[29]captures bidirectional semantic relations in text by defining positive and negative relations, avoiding extraction errors caused by unclear unidirectional semantic features and achieving end-to-end entity relation joint extraction of Chinese text.

4.5. Experimental Results Analysis

4.5.1. Comparative Test

On the DuIE dataset, the experimental results of entity-relation joint extraction for the proposed model and seven other benchmark models are presented in Table 4. Among the seven compared joint extraction methods, MultiR is a weakly supervised entity-relation extraction method, CoType employs sequence annotation for joint extraction, and the pointer annotation model, FETI, CasRel, word mixture model, and BSCRE are joint extraction methods based on parameter sharing. From the data in the table, it is evident that the CJE-PCHF model proposed in this paper achieved accuracy, recall, and F1 scores of 82.1%, 82.2%, and 82.2%, respectively. All metrics are significantly higher than those of the compared benchmark models, with average improvements of 10.8%, 22.1%, and 17.1% in accuracy, recall, and F1 score, respectively. This confirms the effectiveness of the proposed model in the task of entity-relation extraction.

Table 4. The comparative experimental results.

Model	Accuracy	Recall	F1

 doi:10.20944/preprints202407.2384.v1

MultiR	0.577	0.356	0.440
CoType	0.661	0.605	0.632
Pointer Annotation Model	0.694	0.639	0.665
FETI	0.757	0.356	0.440
CasRel	0.772	0.764	0.768
Word Mixture Model	0.813	0.781	0.797
BSCRE	0.816	0.795	0.805
CJE-PCHF(ours)	**0.821**	**0.822**	**0.822**

Compared with other joint extraction models, this model demonstrates superiority for two reasons: (1) The introduction of part-of-speech (POS) and pinyin information assists the model in learning prior features. This paper proposes an interactive fusion network that effectively merges semantic features with POS features and pinyin features, thereby enhancing the model's feature representation capabilities. (2) The proposal of a progressive contrastive learning constraint model. This model optimizes the distribution of the fused features in the feature space through progressive contrastive learning constraints, ensuring alignment of the distributions of the fused features. By introducing a hyperparameter for loss adaptation, the balance between contrastive constraint loss and the model's self-training loss is maintained, thereby improving the overall performance of the model.

In summary, the proposed model surpasses existing baseline models across multiple metrics, verifying its effectiveness and advantages in the task of entity-relation extraction.

4.5.2. Ablation Experiment

This work can be delineated into two primary contributions:

First, to address the insufficiency of prior information consideration in joint extraction tasks, POS information and pinyin information are introduced to enhance model representation. These elements bolster the contextual modeling capability of the model, thereby augmenting its text representation and improving its performance in entity-relation extraction tasks.

Second, to improve the feature integration quality of heterogeneous information, a progressive contrastive learning fusion network is proposed. This network aligns heterogeneous features through constraint-based feature alignment, optimizing their distribution in the feature space, and thus significantly enhancing model performance.

The model is constructed using the parameters that yielded the highest F1 score on the DuIE dataset, and an ablation study is designed to evaluate the impact of prior information and the progressive contrastive learning fusion network on model performance. Three models are involved: the base model (baseline model CasRel); base +a model (CasRel model with the addition of POS and pinyin information); and our proposed model CJE-PCHF, which incorporates POS information, pinyin information, and the progressive contrastive learning fusion network.

Table 5. The ablation experiment results.

Model	Accuracy	Recall	F1
base	0.772	0.764	0.768
base + a	0.816	0.817	0.817
CJE-PCHF	**0.821**	**0.822**	**0.822**

The comparison between the base +a model and the base model reveals that the incorporation of POS and pinyin information for enhanced modeling significantly improves model performance, thereby validating the effectiveness of the first contribution. Furthermore, the comparison between our proposed model and the base + a model demonstrates that the introduction of the progressive contrastive learning fusion network, which aligns heterogeneous features, effectively enhances model performance, confirming the efficacy of the second contribution.

 doi:10.20944/preprints202407.2384.v1

In summary, by incorporating POS and pinyin information for prior feature learning and proposing a progressive contrastive learning fusion network, this paper improves the performance of models in entity relation extraction tasks. The experimental results indicate that these contributions not only enhance the contextual modeling capability of the model but also optimize the feature integration quality of heterogeneous information, thereby boosting the overall performance of the model.

4.5.3. Loss Balance Coefficient Experiment

This study introduces contrastive learning loss into the original loss function of the CasRel model. The loss function employed by our model consists of two parts: the CasRel model's loss and the contrastive learning loss, as elaborated in Section 3.7. The complete loss function is formulated in formula(10):

$$\text{loss} = \lambda loss1 + (1-\lambda)\text{loss}_{\text{contrast}} \quad (10)$$

In this context, loss1 represents the CasRel model's loss function, which measures the discrepancy between the model's predicted outcomes and the actual labels. $\text{loss}_{contrast}$ denotes the contrastive learning loss proposed in this paper, which provides additional constraints and adjustments to enhance the model's training process. The parameter λ signifies the loss balance coefficient. To validate the effectiveness of the proposed loss function, this section designs ablation experiments to explore the impact of varying the loss balance coefficient from 0 to 1 on the model's performance. The experimental results are presented in Table 6 and Figure 4.

Table 6. The loss balance coefficient experiment results.

λ	**0.0**	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**	**0.9**	**1.0**
F1	0.793	0.802	0.812	0.821	0.819	0.819	0.820	0.822	0.820	0.819	0.809

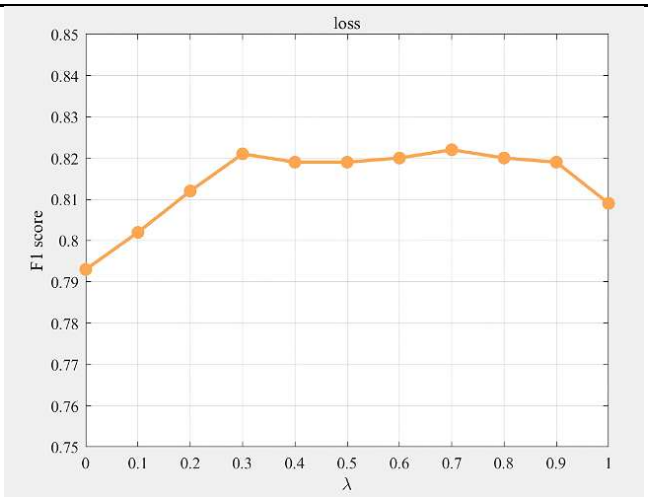

Figure 4. The loss balance coefficient experiment results.

As shown in Table 6 and Figure 4, when λ=0, it indicates the use of only the CasRel model's training loss, and when λ=1, it signifies the exclusive use of the contrastive learning loss function proposed in this paper. According to the trend in Figure 4, the F1 score increases with λ in the range of (0, 0.3). When λ is between (0.3, 0.9), the model's performance maintains a high and stable F1 score, peaking at 82.2% when λ=0.7. This suggests that, within this interval, the weight configuration of the model's self-training loss and the contrastive learning loss achieves a relatively balanced state for the current task and dataset. Subsequently, as λ increases, the F1 score begins to decline. Specifically, when λ is between (0.2, 0.9), the F1 score is higher than when λ=0 or λ=1, indicating that a linear combination of the two losses is more effective in enhancing model performance compared to using a single loss function.

 doi:10.20944/preprints202407.2384.v1

In summary, the experimental results strongly support the effectiveness of the proposed contrastive learning loss function. By adjusting the value of λ, an appropriate balance point can be found, and linearly combining the two losses can effectively integrate their advantages, further improving the model's performance in complex relation extraction tasks.

5. Conclusion

This paper proposes an enhanced method for entity relation extraction in the Chinese domain, building upon the CasRel model. By introducing part-of-speech (POS) and pinyin features, the method effectively strengthens the model's ability to learn prior knowledge from Chinese texts, thereby enhancing its representation capabilities. Subsequently, the interactive fusion network and progressive contrastive learning strategy further optimize the integration of heterogeneous features, achieving a deep fusion of semantic, POS, and pinyin attributes. This leads to improved performance in complex Chinese text extraction scenarios. On the commonly used DuIE evaluation dataset for joint extraction, the proposed model achieved accuracy, recall, and F1 scores of 82.1%, 82.2%, and 82.2%, respectively. Compared to the baseline CasRel model, these represent improvements of 4.9%, 5.8%, and 5.4%, respectively. These results convincingly demonstrate the effectiveness of the proposed model in Chinese entity relation extraction tasks. Additionally, ablation experiments validate the critical role of POS and pinyin information, as well as the progressive contrastive learning mechanism, in boosting model performance.

Author Contributions: Conceptualization, Y.B. and M.H.; Methodology, M.H.; Software, M.H.; Writing—original draft, M.H.; Writing—review & editing, Y.B.; Supervision, D.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Inner Mongolia Autonomous Region Science and Technology Plan Project under Grant No. 2022YFHH0102 and No. 2021GG0090, the Basic Business Funding Project for Universities Directly under Inner Mongolia Autonomous Region under Grant No. BR220145.

Data Availability Statement: No new data were created.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Liu, Q.; Li, Y.; Duan, H.; Liu, Y.; Qin, Z. A survey of knowledge mapping construction techniques. *J. Comput. Res. Dev.* **2016**, *53*, 582-600.
2. Sun, C.; Gong, Y.; Wu, Y.; Gong, M.; Jiang, D.; Lan, M.; Sun, S.; Duan, N. Joint type inference on entities and relations via graph convolutional networks. In Proceedings of the Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 2019; pp. 1361-1370.
3. Mao, T.; Li, X.; LIU, J.; Zhang, D.; Yan, W. Chinese entity and relation extraction model based on parallel heterogeneous graph and sequential attention mechanism. *J. Comput. Appl.* **2023**, 0, doi:10.11772/j.issn.1001-9081.2023071051.
4. LU, X.; SUN, L.; LING, C.; TONG, Z.; LIU, J.; TANG, Q. Named Entity Recognition of Chinese Electronic Health Records Incorporating Phonetic and Part-of-speech Features. *J. Chin. Comput. Syst.* **2024**, 1-12.
5. Saunshi, N.; Plevrakis, O.; Arora, S.; Khodak, M.; Khandeparkar, H. A theoretical analysis of contrastive unsupervised representation learning. In Proceedings of the International Conference on Machine Learning, 2019; pp. 5628-5637.
6. Shao, Y. Research on Chinese entity relationship extraction based on deep learning. 2020.
7. Miwa, M.; Bansal, M. End-to-end relation extraction using lstms on sequences and tree structures. *arXiv preprint arXiv:1601.00770* **2016**.
8. Katiyar, A.; Cardie, C. Going out on a limb: Joint extraction of entity mentions and relations without dependency trees. In Proceedings of the Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2017; pp. 917-928.
9. Zheng, S.; Wang, F.; Bao, H.; Hao, Y.; Zhou, P.; Xu, B. Joint extraction of entities and relations based on a novel tagging scheme. *arXiv preprint arXiv:1706.05075* **2017**.
10. Zeng, X.; Zeng, D.; He, S.; Liu, K.; Zhao, J. Extracting relational facts by an end-to-end neural model with copy mechanism. In Proceedings of the Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2018; pp. 506-514.

11. Fu, T.; Li, P.; Ma, W. Graphrel: Modeling text as relational graphs for joint entity and relation extraction. In Proceedings of the Proceedings of the 57th annual meeting of the association for computational linguistics, 2019; pp. 1409-1418.
12. Wei, Z.; Su, J.; Wang, Y.; Tian, Y.; Chang, Y. A novel cascade binary tagging framework for relational triple extraction. *arXiv preprint arXiv:1909.03227* **2019**.
13. Wang, Y.; Zhang, C.; Bai, F.; Wang, Z.; Ji, C. Review of Chinese named entity recognition research. *J. Front. Comput. Sci. Technol.* **2023**, *17*, 324, doi:10.3778/j.issn.1673-9418.2208028.
14. Zhang, Y.; Liu, Y.; Zhu, J.; Zheng, Z.; Liu, X.; Wang, W.; Chen, Z.; Zhai, S. Learning Chinese word embeddings from stroke, structure and pinyin of characters. In Proceedings of the Proceedings of the 28th ACM International Conference on Information and Knowledge Management, 2019; pp. 1011-1020.
15. Wang, H.; Shi, Y.; Liu, G.; Duan, J. Named Entity Recognition of Microblog Texts Based on POS Features. *J. North China Univ. Technol.* **2019**, *31*.
16. Zhu, W.; Jin, X.; Ni, J.; Wei, B.; Lu, Z. Improve word embedding using both writing and pronunciation. *PLoS ONE* **2018**, *13*, e0208785.
17. Wu, S.; Song, X.; Feng, Z. MECT: Multi-metadata embedding based cross-transformer for Chinese named entity recognition. *arXiv preprint arXiv:2107.05418* **2021**.
18. Yu, X.; Li, L.; ZhouJianlun; Ma, H.; Chen, P. AMFRel: A method for joint extraction of enity relations in Chinese electronic medical records. *J. Chongqing Univ. Technol. (Nat. Sci.)* **2024**, *38*, 189-197, doi:10.3969/j.issn.1674-8425(z).2024.02.021.
19. Li, S.; Chang, Z.; Liu, Y. Joint Extraction of Entities and Relations Based on Multi-feature Fusion. *Int. J. Emerg. Technol. Adv. Appl.* **2024**, *1*.
20. Ge, J.; Li, S.; Fang, Y. Joint extraction method of Chinese entity relationship based on mixture of characters and words. *Appl. Res. Comput.* **2021**, *38*, doi:10.19734 /j.issn.1001-3695.2021.01.0006.
21. Lei, J.; Lai, K.; Yang, S.; Wu, Y. Joint entity and relation extraction based on contextual semantic enhancement. *J. Comput. Appl.* **2023**, *43*, 1438-1444, doi:DOI : 10.11772/j.issn.1001-9081.2022040625.
22. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2013**, *26*.
23. Fang, H.; Wang, S.; Zhou, M.; Ding, J.; Xie, P. CERT: Contrastive Self-supervised Learning for Language Understanding. *arXiv preprint arXiv:2005.12766* **2020**.
24. Li, S.; He, W.; Shi, Y.; Jiang, W.; Liang, H.; Jiang, Y.; Zhang, Y.; Lyu, Y.; Zhu, Y. Duie: A large-scale chinese dataset for information extraction. In Proceedings of the Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, October 9–14, 2019, Proceedings, Part II 8, 2019; pp. 791-800.
25. Hoffmann, R.; Zhang, C.; Ling, X.; Zettlemoyer, L.; Weld, D.S. Knowledge-based weak supervision for information extraction of overlapping relations. In Proceedings of the Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies, 2011; pp. 541-550.
26. Ren, X.; Wu, Z.; He, W.; Qu, M.; Voss, C.R.; Ji, H.; Abdelzaher, T.F.; Han, J. Cotype: Joint extraction of typed entities and relations with knowledge bases. In Proceedings of the Proceedings of the 26th international conference on world wide web, 2017; pp. 1015-1024.
27. Wang, Y.; Mu, H.; Zhou, L.; Xing, W. Joint extraction method of entity and relationship based on pointer network. *Appl. Res. Comput.* **2021**, *38*, doi:10.19734/j.issn.1001-3695.2020.04.0113.
28. Chen, R.; Zheng, X.; Zhu, Y. Joint entity and relation extraction fusing entity type information. *Comput. Eng.* **2022**, *48*, 46-53, doi:10.19678/j.issn.1000-3428.0060745.
29. Yu, K. Joint Extraction of Chinese Entity Relationship Based on Bidirectional Semantic Learning Model. 2022.