CEntRE: A paragraph-level Chinese dataset for Relation Extraction among Enterprises

Peipei Liu1,2, Hong Li1,2*, Zhiyu Wang2,3, Yimo Ren1,2, Jie Liu1,2, Fei Ly2, Hongsong Zhu1,2, Limin Sun1,2
1School of Cyber Security, University of Chinese Academy of Sciences, Beijing, China
2Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
3Henan University, Kaifeng, China
{liupeipei, lihong, zhuhongsong}@iie.ac.cn

Abstract—Enterprise relation extraction aims to detect pairs of enterprise entities and identify the business relations between them from unstructured or semi-structured text data, and it is crucial for several real-world applications such as risk analysis, rating research and supply chain security. However, previous work mainly focuses on getting attribute information about enterprises like personnel and corporate business, and pays little attention to enterprise relation extraction. To encourage further progress in the research, we introduce the CEntRE, a new dataset constructed from publicly available business news data with careful human annotation and intelligent data processing. Moreover, we propose a joint entity and relation extraction network, which is capable of discovering enterprise entities and extracting business relations between them accurately. The network firstly encodes input sequences with strong semantic augmentation to learn contextual representation for each token, then a conditional random field (CRF) module is used for entity extraction. Subsequently, entity pairs are built and a new encoder based on the entity pairs is applied to get global information for relation extraction. Finally, a biaffine classifier is deployed to classify the relations. Extensive experiments on CEntRE demonstrate the effectiveness of our proposed method compared with other six excellent models, and thus our model can be considered as one strong baseline. The data and code are available at: https://github.com/LiuPeiP-CS/Mining_Entity_Relations_Among_Enterprises

I. INTRODUCTION

Enterprise relation extraction aims to automatically detect pairs of enterprise entities and identify the business relations between them from unstructured or semi-structured text data without human intervention [1], and it is a sub-task in the natural language processing (NLP). In the real world, enterprise relation extraction plays an important role in several fields such as economics and finance, information security and supply chain security, etc. On one hand, invest institutions can get the relations of equity distribution and debtor-creditor among different enterprises based on the relation extraction, and these relations are then used for risk analysis and rating research. On the other hand, as an important part of enterprise operation, the supply chain relations can help enterprises understand and analyze industries, make management decisions and select business partners, as well as enhancing the competitiveness and improving profit margin. More than that, the analysis of enterprise relations referring to social engineering is beneficial to effective surveillance and precautions for network penetration since the low security protection of subsidiary enterprises may threaten the superior enterprises and the partnership is easy to be exploited by phishing attacks, etc. Therefore, relation extraction among enterprises is crucial to the security of development and reduction of economic losses.

Previous work mainly focuses on getting attribute information about enterprises like personnel, corporate business, product or certain relations (e.g., chip supply chain) instead of general relations among enterprises from unsupervised data by using rules, dictionary, template and other hand-craft features [2]–[5]. Despite achieving desirable results, these content and methods are trapped by some inevitable restrictions in practice: they are applied to domain-specific, and are quite difficult to be used for conducting business analysis comprehensively enough; with the explosion growth and the increasing diversity of new text data, fixed patterns and manual features become more expensive and unapplicable.

Recent developments in deep learning have promoted the research of neural relation extraction (NRE), which attempts to use neural networks to automatically learn high-level semantic features and then extracts the entity relations [1], [6]–[12]. The models trained with neural networks are conveniently fine-tuned for various types of content and knowledge, which is conducive to the migration and application of extraction methods. Although NRE brings ideas to the further research about enterprise relations, deep learning is data-consuming and there is still a lack of unified gold-standard supervised dataset for enterprise relation extraction which hinders the future research in this area.

According to the above analysis, we take an initial step towards studying relation extraction (RE) among enterprises based on deep learning in this paper.

Firstly, a Chinese relation extraction dataset called CEntRE is constructed. Specifically, the dataset contains 11018 relational triples and 10894 enterprise entities on 3018 paragraphs, and 19 relations that exist between pairs of enterprise entities are covered. The CEntRE has following four characteristics: (I) Overlapping entity: The relational facts in paragraphs of CEntRE are often complicated, and different relational triplets may have overlaps in the same entities. (II) Flexible name structure: Unlike the registration information in the State Administration for Market Regulation, the enterprise names in CEntRE are arbitrary and easy to be confused with other common words, such as “我爱网络公司” (I Love Network Company) and “交个朋友” (Make Friends). (III)
Cross-sentence and Synonyms reasoning: As a substantial portion of the relational facts, some relations in CEntRE can only be extracted from multiple sentences. It means that CEntRE requires reading multiple sentences in a paragraph to recognize entity pairs and infer their relations by synthesizing all information of the paragraph. In addition, Chinese synonyms should not be ignored, such as “销售(sell to)” “供给(supply to)” and “供应(provide to)”.

IV. Wide coverage: The CEntRE originates from more than 20,000 news media reports among 7 websites of technology and finance in 6 months, and those abstract enterprise relations can cover most of the application scenarios in real world.

Taking the characteristics of CEntRE into consideration, we then propose a joint neural networks model for enterprise entity recognition and relation extraction. Four steps are taken for our proposed model: (1) Firstly, we follow [13], [14] to use word-character lattice method to initialize each character in an input paragraph. Not only that, a semantic augmentation module is created for enhancing the contextual representation of each character. (2) After that, a encoder with a sequential conditional random layer (CRF) [15] following it is designed to classify the characters and identify enterprise entity spans. (3) Subsequently, we employ another encoder to get the context-sensitive representation of each character with considering its corresponding type and position. Furthermore, we dynamically integrate character-level representation output from both encoders into the span-specific feature under the supervision of detected entity spans. (4) Finally, the span representations are utilized to perform relation extraction via a biaffine classifier. Since each possible entity pair is separately considered, selecting overlapping entity mentions is possible.

In summary, this paper makes following contributions:

- A labeled Chinese enterprise relation extraction dataset CEntRE with 3018 paragraphs and 19 relations is released for analyzing the interactions among the enterprises. The dataset is at paragraph-level, and it has some particular challenges.
- To evaluate the challenge of CEntRE, we present a novel approach towards span-based joint entity and relation extraction.
- We also conduct extensive evaluation on the other state-of-the-art NRE methods in various setting. Experiment results show that our approach appears to be simple but effective.
- Furthermore, detailed analyses on the results also reveal multiple promising directions worth pursuing.

II. RELATED WORKS

A. Enterprise relations

The values of enterprise relation management are receiving attention from researchers and practitioners in several sectors with the growing market world economy. [2] concerns on the supply relation between enterprises, especially the role of corporate entity. A library of relation word is built to judge the theme of the text, and the nearest syntactic dependent verbs are used to judge the semantic relation between enterprise entities. By using the maximum entropy model to analyze the unlabeled data and determine the optimal feature template, [5] constructs the enterprise knowledge graph for financial application. [3] presents a pipeline model for company name identification and relation extraction based on fusing rules, dictionary matching and machine learning algorithm. However, these methods cost a lot of manual efforts and are difficult to be generalized.

[4] introduces the triggering mechanism, and then the model with triggering words constraint is used to extract 5 types of relation from the standard annual report of Chinese Listed Firms. Although the method can obtain good results, it is limited to standard data. [16] designs automatic models to get relations and entities about enterprises, whereas they mainly focus on the information of enterprise-self such as location, personnel and industry. FinRE [13] is the closest work to ours, but it is simpler with only two enterprise entities and one relation between them in a single sentence compared to us. In addition, our joint model extracts entities and relations automatically while they only extract the relation given two entities.

B. Neural Relation Extraction

Neural relation extraction aims to automatically learn triplets (subject, relation, object) from the unstructured text without human intervention [1]. There are two kinds of approaches for the task [7]: the pipelined framework, which first uses named entity recognition (NER) models to detect entity spans, and then implements relation extraction between pairs of detected spans based on classification models; and the joint learning method, which combines the NER model and the RE model through different strategies, such as constraints or parameters sharing.

Compared with the joint learning methods, pipeline framework can be designed and applied more flexibly, and pre-trained models [17] and various attention mechanisms [18] are often leveraged to boost the performance for NER and RE in pipeline framework. However, pipeline models meet with some obvious drawbacks. Each component limits the performance because of the error cascading effect and there is no chance for the model to correct mistakes. In addition, such pipeline models cannot capture the explicit relation between the two sub-tasks [1].

To address the aforementioned disadvantages of the pipeline framework, joint methods of entity and relation extraction have been proposed. Following [10], we also group existing joint methods into two main categories: structured prediction and multi-task learning. The structured prediction has three major research lines: Table Filling, Tagging, and Sequence-to-Sequence (Seq2Seq). Among these approaches, the table filling method [19] requires the model to enumerate over all possible entity pairs, while the tagging methods [11] run tagging on a sentence for multiple turns. Seq2Seq model [1], however, receives the unstructured text as input and directly decodes the entity-relation triples as a sequential output.
As for multi-task learning, sub-task interaction is achieved with entity and relation prediction sharing the same features and parameters, and the model optimizes both sub-models together. [8] also employs a BiLSTM to encode each word of the sentence, and adopts a CRF to extract entity boundaries and tags. Then, multiple relations are potentially identified for each entity with multi-head selection. [20] uses a BiLSTM-CRF based model for entity recognition, and adopts interaction via a bi-affine attention layer in relation classification. [9] argues that any span could constitute a potential entity and a relation could hold between any pair of spans. So, it conducts a lightweight reasoning on the BERT [17] embeddings of input sentences, which filters entities and classifies relations with a localized, marker-free context representation.

In this paper, we combine the advantages of multi-task learning and the pipeline work to build a joint model, where we not only use a shared encoder to encode the two sub-tasks but also construct a separate relation encoder to dynamically incorporate the features of detected entities to support relational classification.

III. DATASET CONSTRUCTION
A. Collection and Annotation
Our ultimate goal is to construct a dataset for paragraph-level RE from open text, which requires necessary information including enterprise named entities and relations of all entity pairs. Due to the complexity and diversity of enterprise relations, building dataset is time-consuming and difficult. Fig. 1 shows the construction process and we will introduce the processing details in the following:

**Step 1: Relation Dictionary Establishment:** The first asset of high-quality dataset should be a high-coverage of relations space. We abstract 19 types of enterprise relations by summing up the business knowledge1, referencing existing literatures and discussing with domain experts. Table I gives a detailed description of the relations. Although individual difference exists in each form of textual expression, it is found that such relations have internal connections. For example, “参股” (equity participation) covers “控股” (hold controlling stakes).

**Step 2: Dataset Collection and Machine Filtering:** To increase the diversity of data, we crawl more than 20000 news reports from 7 different websites in 6 months (2020.9-2021.3), which cover finance and technology, and these contents are written by different authors with various styles. Based on the predefined relation dictionary, we first expand the relation vocabulary by synonyms (Table I shows some extended samples). Then, these relation words are aligned to paragraph items to coarsely filter irrelevant paragraphs. Since there should

### TABLE I
THE DESCRIPTIONS OF ENTERPRISE RELATIONS

| Chinese | English | Description | Count | Symmetric |
|---------|---------|-------------|-------|-----------|
| 合作 | Cooperation | Subject A cooperates with object B. "Cooperation" relationship includes signing, win the bidding, customer, etc. | 1184 | True |
| 供应 | Supply | Subject A supplies products, business, technology, etc. to object B. "Supply" relationship includes sold to, procurement, supply, buy, etc. | 629 | False |
| 参股 | shareholding/joint-stock/ equity participation | Subject A holds shares in subject B, or purchases shares in B from a third party. | 1162 | False |
| 转让 | transfer | Subject A transfers shares, patents, technologies, etc. to subject B. | 335 | False |
| 控股 | hold controlling stakes | Subject A holding object B. Holding and shareholding are essentially the same, except for different explicit expressions and shares. | 470 | False |
| 附属 | be the subsidiary/off belon to | Subject A is owned by object B. | 2095 | False |
| 联营 | joint venture | Subject A and object B co-create C. | 241 | True |
| 投资 | investment | Subject A invests in object B, but is not sure whether to take a stake | 819 | False |
| 授权 | licensing | Subject A licenses patents, technologies, etc. to object B. | 39 | False |
| 代管 | manage on behalf of | Subject A entrusts object B to manage certain content on its behalf. | 26 | False |
| 合并 | merger | Subject A and object B are reorganized and merged into C. | 86 | True |
| 分离 | spin-off | Subject A spins off object B. | 316 | False |
| 竞争 | competition | Subject A competes with object B in certain areas. | 1142 | True |
| 代工 | OEM | Subject A performs OEM production for object B. | 63 | False |
| 委托 | commission | Subject A commissions object B to do some work. | 38 | False |
| 更名 | change name/ rename | Subject A changed its name to object B. | 390 | False |
| 共指 | abbreviation | Object B, short for Subject A, Or Chinese with English. | 1446 | True |
| 纠纷 | lawsuit | Subject A sues object B for a certain content. Or commercial contradiction. | 121 | True |
| 关联 | correlation | Subject A is related to Object B, but it is not really a specific relation. | 416 | True |

---

1This includes business materials (annual reports of listed companies from Shanghai Stock Exchange) and financial news (such as “yicai” and “hexun”).
be at least two enterprise entities (subject and object) in each paragraph text, we further discard paragraphs containing fewer than 2 entities. A named entity recognition model trained on MSRA dataset is used for detecting enterprise entities and selecting data. MSRA dataset consists of 3 types of entity tags—“LOC” “PER” “ORG”, but we only need the tag “ORG”.

**Step3-Human Annotation:** We commit multiple master students to annotate the remaining result from second step manually, and each paragraph is annotated by two individuals separately. However, there may be some difference between the two annotations, and a third annotator would like to be asked to merge these differences and make the final decision. In the worst case, we have to remove the data which all the annotators cannot reach a consensus on. To provide high-quality annotations, all annotators are well trained that a principled training procedure is adopted and every annotator is required to pass test tasks before annotating the dataset. Apart from aforementioned matters, we transform some relations to make them meet the fact while annotating due to the time-series of business activities. For example, we get “C附属(A belong to)B” instead of “A转让(transfer the possession to)C” if “A向B转让了C的100%股权”(A has transferred 100% equity of C to B) is presented because that C no longer belongs to A.

**Step4-Verification and Standardization:** To evaluate the quality of CEntRE, we randomly selected 180 pieces of data and divided them to four extra reviewers who have known our relation requirements (60 pieces each with an overlap of 20). Each reviewer has to evaluate the correctness of the annotation by classifying each piece of data into {Correct, Incorrect or Uncertain}. The inter-reviewer agreement score of Correct for the overlap data is 96.25%, which shows that our annotation is successful. At last, we convert the data from string to structured triplets for the facilitation of application and learning. Considering that one entity may have several mentions appeared in different positions while not all of them interact with others for relations, we thus adopt the shortest dependency path (SDP) to identify the best mention for building the relation. The dependency parsing tool we use is Spacy.

### B. Statistics and Analysis

This section conducts some additional statistics and analyses to gain a better understanding of the proposed dataset CEntRE. From the statistics, we have the following observations:

1. The current version of the CEntRE dataset contains 11018 labeled triplets and 10894 enterprise entities taking place

2. In fact, only about 16% inconsistent paragraph annotations from the two separate annotators need a third intervention in this stage, which indicates that incorrect labels are limited and the annotation is reliable.

3. https://spacy.io/

### IV. PROPOSED BASELINE MODEL

The framework of our proposed model is shown in Figure 2. We will introduce the details of 4 main modules through an input sequence $X = (c_1, c_2, ..., c_12) = “$ “汽解放前身为
A. Embedding Layer

We follow [13], [14] to introduce word-character lattice to initialize each character for the structural semantic representation. Not only that, a Chinese Word2Vec model pre-trained by Tencent [21] is used for enhancing the contextual semantic representation of each character to alleviate the problem of recognizing flexible entity names.

Specifically, a character $c_i$ get its contextual semantic augmentation based on its most similar $K$ neighbors $(c_{i,j})_{j=1}^K$ with the attention mechanism. Firstly, each neighbor is assigned a weight as it makes the different contribution from others to augment the anchor:

$$
\alpha_{ij} = \frac{\exp(\text{cosine}(v_i, v_{ij}))}{\sum_{k=1}^K \exp(\text{cosine}(v_i, v_{ik}))}
$$

where $v_i$ and $v_{ij}$ are the embeddings of $c_i$ and $c_{i,j}$ from the Word2Vec model, respectively. Then the semantic augmentation representation $c_i^S$ of $c_i$ can be computed by:

$$
c_i^S = \sum_{j=1}^K \alpha_{ij} v_{ij}
$$

The input representation $c_i^T$ of $c_i$ can thus be given by concatenating the representation $c_i^E$ from lattice, $c_i^S$ and the absolute position embedding $c_i^{AP}$:

$$
c_i^T = [c_i^E; c_i^S; c_i^{AP}]W_i
$$

where $W_i$ is a trainable transformation matrix.

B. Shared Encoder and NER decoder

After the input embedding, a multi-layer Transformer [22] with the multi-head attention and feed-forward neural network(TMFF) is employed as an encoder for capturing the semantic information of each character from the context interaction.

Except for the TMF, a fine-tuned BERT is added into the shared encoder. Benefiting from its effective structure and a rich supply of large-scale corpora, BERT has a strong dynamic feature extraction capability and can thus contribute to the performance improvement of NRE. In fact, the output representation $c_i^E$ of $c_i$ from the shared encoder is got by a gate, and it is expressed by:

$$
f_i = \delta([c_i^T; c_i^B]W_{\delta})
$$

$$
c_i^E = f_i \odot c_i^T + (1 - f_i) \odot c_i^B
$$

where $\delta$ is the sigmoid activation function, $W_{\delta}$ is the trainable matrix, $c_i^T$ is the output from TMF corresponding to $c_i^E$, $c_i^B$ is the output from the BERT and 1 is a vector whose elements are all 1.

NER is the first task in our joint model, and it is here formulated as a sequence labelling problem with BIO (Beginning, Inside, Outside) tagging scheme. In this task, a CRF layer after the shared encoder is introduced to learn the tag score of each character$^4$ for identifying enterprise entity spans. With the train dataset $\{(X_t, l_t)\}_{t=1}^M$, we optimize the parameters of this task by minimizing the loss during training:

$$
L_{ner} = - \sum_{t=1}^M \log(P(l_t|X_t))
$$

where $X_t$ is the $t$-th sequence in the dataset, $l_t$ is the true tag sequence of $X_t$ and $P(l_t|X_t)$ is the probability computed by CRF. For the given example $X$, we can get two named entities $e_1 = (x_1, ..., x_4) = "\text{一汽轿车}"$ and $e_2 = (x_8, ..., x_{11}) = "\text{一汽解放}"$ in the NER task.

C. Relation Encoder

Considering that one entity participating in several relations (such as the “联想” in SPO and “华为” in EPO shown in Table III) should have different contextual representations when building entity pairs for different relations, we thus design the relation encoder (RE-encoder) with BiLSTM layers to get the most effective representation for relation extraction centred on the entity pairs.

$^4$For the English named entities including spaces and English words, we regard each space or English word as a Chinese character.
where \( \phi \) corresponds to the positions relative to subject entity and object entity, respectively. So, the input of \( c_t \) can be depicted as a concatenation by \( c_t^{Rt} = [c_t^E; c_t^{RPs}; c_t^{RPo}] \), and we regard the corresponding output from BiLSTM as \( c_t^O \). For example, centered on the entity pair \((e_1, e_2)\), the tag of “\( \text{B}\)” is “I” and its position relative to the two entities is \((0,4)\) while the tag of “\( \text{O}\)” is “O” and the position relative to two entities is \((2,-2)\). After the relation encoder, we can have \( c_1^O \) for “\( \text{B}\)” and \( c_0^O \) for “\( \text{O}\)”.

Then, we would like to get the representation of each entity span for the relation extraction of an entity pair. However, a single entity detected by CRF may be composed of several characters, hence we compute span representations through character features with the attention mechanism:

\[
\gamma_i = MLP(m_i)
\]

\[
\beta_{t,j} = \frac{\exp(\gamma_i)}{\sum_{j=\text{start}(t)}^{\text{end}(t)} \exp(\gamma_j)}
\]

\[
\hat{m}_t = \sum_{j=\text{start}(t)}^{\text{end}(t)} \beta_{t,j} m_j
\]

Here, \( m_i = [c_t^R; c_t^E] \) corresponds to the \( c_t \) which is a character in the entity span. \( MLP \) is a MultiLayer Perceptron, \( \text{start}(t) \) and \( \text{end}(t) \) denote the start and end character of \( t \)-th \((t = 1,2)\) span respectively, and \( \hat{m}_t \) is a weighted sum over the characters involved in \( t \)-th entity span. Finally, for \( t \)-th entity span, its representation \( g_t \) could be defined as:

\[
g_t = [m_{\text{start}(t)}; m_{\text{end}(t)}; \hat{m}_t; \varphi(t)]
\]

where \( \varphi(t) \) is the embedding vector for the width of \( t \)-th entity span.

D. Relation Classification

Since some relations are asymmetric and an entity may play different roles as the subject or object in a relation(i.e., \((e_1,e_2)\) is different from \((e_2,e_1)\)), we thus apply two versions of feed-forward neural network (FFNN) to distinguish the role. In particular, each \( g_t \) is projected into two separate FFNNs to generate the subject representation \( g_t^s \) and object representation \( g_t^o \):

\[
g_t^s = FFNN_s(g_t; W_s)
\]

\[
g_t^o = FFNN_o(g_t; W_o)
\]

where \( W_s \) and \( W_o \) are neural network parameters to be learned respectively.

We consider each ordered entity pair, and compute its relation probability by the biaffine classifier. For the entity pair \((g_t^s,g_t^o)\), its probability is computed for each relation type as:

\[
P(r_{i,j} | (g_t^s,g_t^o)) = \sigma(g_t^sRg_t^o + [g_t^s; g_t^o]W_r + b_r)
\]

where \( \sigma \) is the softmax function, \( R \) is a learnable \( d \times c \times d \) bi-affine tensor with the number of relation categories \( c \) and the output dimension \( d \) of FFNN, \( W_r \) is a \( 2d \times c \) matrix, and \( b_r \) is the bias.

We train our RE task with a cross entropy loss over a set of gold standard data:

\[
L_{rec} = - \sum_{i=1}^{N} \sum_{j=1, i \neq j}^{N} \log P(\hat{r}_{i,j} | (g_t^s,g_t^o))
\]

where \( N \) is the number of entity spans, \( \hat{r}_{i,j} \) is the actual relation label of \((g_t^s,g_t^o)\).

As a result, the loss of our joint model is defined as the weighted sum of both NER task and RE task:

\[
Loss = L_{ner} + \lambda L_{rec} + \frac{\lambda}{2} \|\theta\|^2_2
\]

where \( \theta \) depicts all model parameters, and \( \lambda \) is the weight parameter.

V. EXPERIMENTS AND ANALYSIS

In this section, we conduct experiments on the CEntRE dataset by our proposed model and other six excellent NRE models covering the pipeline work and the joint learning.

A. Experimental Setting and Metrics

**Setting:** We use the ratio of 8:1:1 to divide the entire dataset into training set, test set and validation set. Each model is trained for 10 times with 40 epochs each time, and we report the average result on the test set. For our proposed model, our training uses 512 hidden states in the Transformer and BiLSTM with batch size of 1. The output dimension of FFNN is 128 and the position embeddings are randomly initialized. The model hyperparameters are updated using back-propagation by the Adam optimizer. The learning rate is 5e-4, and the weight parameter \( \lambda \) of joint loss is 0.5. The \( K \) in the input semantic augmentation is set as 4. As for the hyperparameters of the compared models, we keep all them unchanged.

**Metrics:** We evaluate NER and RE with Precision (P), Recall (R), and micro F1 scores, while a predicted entity is correct if its boundaries are correct and a predicted relation is true if the triplet (subject, relation, object) is infallible.

B. Baseline Models

For the pipeline work, we first train the NER task(BERT+CRF or BiLSTM+CRF) and RE task(MLP) separately, and then pass the predicted entities from BERT+ or BiLSTM+CRF into MLP to get relation classification results for any pairs of entities during test.

As for the joint learning, we have grouped them into two main types in Section II-B. Table-Sequence [12] is an excellent method of Table Filling, while CasRel [27] and TP-Linker...
are the representative works of the novel tagging strategy. Both the Table Filling and the novel tagging strategy belong to the structured prediction. Spert [9] is the most similar work to ours, but it costs much time to enumerate all possible spans for predicting entities and it treats an entity equally no matter how it is building entity pairs for different relations. Moreover, Spert [9] uses an empirical threshold to filter the predefined relation while we adopt the proven effective bioaffine with softmax. Both the Spert and ours are the multi-task learning.

C. Main Results

### TABLE IV
**Performance Comparison on CENTRE.**

| Models    | NER   | RE   |
|-----------|-------|------|
| LSTM+MLP  | 89.52 | 81.01 |
| BERT+MLP  | 93.33 | 85.61 |
| CasRel    | 96.09 | 88.31 |
| TP-Linker | 94.97 | 88.73 |
| Table-Sequence | 91.57 | 89.02 |
| Ours      | 96.71 | 93.42 |

### Comparisons with Baselines: All experimental results are summarized in Table IV, and there are several important observations from the table. (1) Comparing BiLSTM+MLP and BERT+MLP, we find that the performances of BERT+MLP are better than BiLSTM+MLP on both tasks. This indicates that the pre-trained model is quite effective due to its large external knowledge support. (2) It is clear that joint learning methods perform better compared with all of the pipeline works since such pipeline models cannot capture the explicit interaction between the two sub-tasks and limit the performance of each component. (3) The Spert and our model can get the higher RE recall among all the models, because both of them build all valid span-pairs for predicting the relations. (4) The CasRel and Table-Sequence can get the better performance on the RE precision thanks to their advantages in dealing with the overlapping entity of relation extraction. (5) Our model can achieve the best performance on both NER and RE task with the specific designs for dataset characteristics including the semantic enhancement of flexible sparse names and the different context representations of overlapping entities. From the compared results, we approximately think that the model can be regarded as a baseline for the further research. (6) Although the NER has achieved good results, the RE effects of all models are far from the human annotation, highlighting the necessity to identify effective mentions and capture the inter-sentence context. This reflects the challenges of CentRE and demonstrates that these current works are not sufficient to solve the complexity of this dataset, presenting the opportunities for the future work.

### Ablation Study: We evaluate the effect of each individual component in our model including semantic augmentation in IV-A, the gate control in IV-B, the RE-encoder in IV-C, the attention for entity span embedding in IV-C. Table V shows that: (1) Any of these components in absence can lead to the extraction performance descent. (2) SemAug has a significant impact on entity recognition as it is designed for alleviating the data sparsity caused by flexible entity names. (3) The lack of the gate will cause a large decline in NER performance since the model cannot measure the contributions of input contents resulting in the loss of their advantages for the features with greater contributions. (4) RE-encoder has relatively little influence on entity extraction but it is very important for the effect of relation extraction. The reason is that it can produce the specific contextual representation for predicting the relation of entity pairs but obtain the results from NER. (5) We replace Span-attention with max-pooling to evaluate the role of the Span-attention. From the results, we can see that it slightly affects the overall performance compared with other modules, which may due to the ambiguous contributions of all words in the span. (6) When evaluating the effect of joint loss and separate loss, the experimental results show that we obtain the best model performance when the joint loss is adopted. These results validate that the two tasks can interact, influence and work together to improve each other.

### D. Experiments on Other Benchmarks:

We further conduct extensive evaluation on the other two popular benchmark datasets ACE05 [29] and CoNLL04 [30] to verify the effectiveness of our model and also assess the generalization ability of the model to other datasets. These two datasets are usually used for overlapping entity relation extraction, so they are the very appropriate choices.

As Table VI shows (F1-score), although the published best models have advantages in dealing with overlapping entities specially, our model can still achieve competitive results on both two datasets compared with them. For the NER in both ACE05 and CoNLL04, our model can exceed many baseline methods except HRL [33] in ACE05. The results prove the entity extraction capacity of our model by multi-task learning and improving the context representation of words or chars. As for RE task, we can see that the performance from our model is also better than some baselines, but slightly inferior to the few ones which have the complex graph structure. All these results demonstrate the generalization and effectiveness.
of our model and provide the new support for our model as a baseline on the CEntRE dataset.

VI. CONCLUSION

In this paper, we present a new Chinese dataset CEntRE with large data volume and high coverage. Our dataset focuses on enterprise relation extraction, and we hope it can prompt the research in related areas. We then propose a joint model for enterprise entity detection and relation extraction. The extensive evaluation on CEntRE has verified its effectiveness, and thus can be considered as one strong baseline for the future research.

VII. ACKNOWLEDGMENTS

Thank you to all the data annotators for their contributions to our work. This work was supported by the National Key R&D Program of China (No.2020YFB2103803) and the National Natural Science Foundation of China (No.61931019, No.U1766215).

REFERENCES

[1] D. Zeng, H. Zhang, and Q. Liu, “Copynet: Copy mechanism for joint extraction of entities and relations with multi-task learning,” Proceedings of the AAAI Conference on Artificial Intelligence, pp. 9507–9514, Apr. 2020.
[2] C. Yang and J. Wang, “Research on automatic extraction of enterprise supply relationship based on nlp,” Computer Science and Application, pp. 1823–1832, 2018.
[3] L. Meng, Z.-H. Wei, Y.-T. Hu, and et al., “An improved method for chinese name and abbreviation recognition,” in International Conference on Knowledge Management in Organizations, Cham, 2017, pp. 435–447.
[4] J. Dai, J. Mao, X. Liu, and et al., “Enterprise ecological relationship extraction based on feature vector and svo extension,” Computer Technology and Development, pp. 146–151, 2018.
[5] C. Sun, Y. Fu, W. Cheng, W.-n. Qian, and et al., “Chinese named entity relation extraction for enterprise knowledge graph construction,” Journal of East China Normal University (Natural Science), pp. 60–71, 2018.
[6] W. Huang, X. Cheng, T. Wang, and et al., “Bert-based multi-head selection for joint entity-relation extraction,” in Natural Language Processing and Chinese Computing, Cham, 2019, pp. 713–723.
[7] K. Xue, Y. Zhou, Z. Ma, and et al., “Fine-tuning bert for joint entity and relation extraction in chinese medical text,” in 2019 IEEE International Conference on Bioinformatics and Biomedicine, 2019, pp. 892–897.
[8] G. Bekoulis, J. Deleu, T. Demeester, and et al., “Joint entity recognition and relation extraction as a multi-head selection problem,” Expert Systems with Applications, pp. 34–45, 2018.
[9] M. Eberts and A. Ulges, “Span-based joint entity and relation extraction with transformer pre-training,” arXiv, 2019.
[10] Z. Zhong and D. Chen, “A frustratingly easy approach for entity and relation extraction,” in Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online, 2021, pp. 50–61.
[11] D. Dai, X. Xiao, Y. Lyu, and et al., “Joint extraction of entities and overlapping relations using position-attentive sequence labeling,” Proceedings of the AAAI Conference on Artificial Intelligence, pp. 6300–6308, 2019.
[12] J. Wang and W. Lu, “Two are better than one: Joint entity and relation extraction with table-sequence encoders,” in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 2020, pp. 1706–1721.
[13] Z. Li, N. Ding, Z. Liu, and et al., “Chinese relation extraction with multi-grained information and external linguistic knowledge,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, Jul. 2019, pp. 4377–4386.
[14] Y. Zhang and J. Yang, “Chinese NER using lattice LSTM,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, Jul. 2018, pp. 1554–1564.
[15] J. Lafferty, A. McCallum, and F. Pereira, “Conditional random fields: Probabilistic models for segmenting and labeling sequence data,” in Proc. 18th International Conf. on Machine Learning, 2001.
[16] T. Ruan, L. Xue, H. Wang, and et al., “Building and exploring an enterprise knowledge graph for investment analysis,” in The Semantic Web – ISWC 2016, Cham, 2016, pp. 418–436.
[17] J. Devlin, M.-W. Chang, K. Lee, and et al., “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, Minnesota, 2019, pp. 4171–4186.
[18] W. Wu, Y. Chen, J. Xu, and et al., “Attention-based convolutional neural networks for chinese relation extraction,” in Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data, Cham, 2018, pp. 147–158.
[19] H. Adel and H. Schütze, “Global normalization of convolutional neural networks for joint entity and relation classification,” in Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, Copenhagen, Denmark, 2017, pp. 1723–1729.
[20] D. Q. Nguyen and K. Verspoor, “End-to-end neural relation extraction using deep bijaftine attention,” in Advances in Information Retrieval, Cham, 2019, pp. 729–738.
[21] Y. Song, S. Shi, J. Li, and et al., “Directional skip-gram: Explicitly distinguishing left and right context for word embeddings,” in In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, New Orleans, Louisiana, USA, 2018, pp. 175–180.
[22] A. Vaswani, N. Shazeer, N. Parmar, and et al., “Attention is all you need,” in Proceedings of the 31st International Conference on Neural Information Processing Systems, ser. NIPS’17, Red Hook, NY, USA, 2017, p. 6000–6010.
[23] J. Wang, X. Chen, Y. Zhang, and et al., “Document-level biomedical relation extraction using graph convolutional network and multi-head attention (preprint),” JIMR Medical Informatics, 2019.
[24] J. Lee, S. Seo, and Y. S. Choi, “Semantic relation classification via bidirectional lstm networks with entity-aware attention using latent entity typing,” Symmetry, 2019.
[25] F. Christopoulos, M. Miwa, and S. Ananiadou, “A walk-based model on entity graphs for relation extraction,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, 2018, pp. 81–88.
[26] H. Zhu, Y. Lin, Z. Liu, and et al., “Graph neural networks with generated parameters for relation extraction,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, pp. 1331–1339.
[27] T. Wei, J. Su, Y. Wang, and et al., “A novel cascade binary tagging framework for relational triple extraction,” in Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 2020, pp. 1476–1488.
[28] Y. Wang, B. Yu, Y. Zhang, and et al., “Tplinker: Single-stage joint extraction of entities and relations through token pair linking,” 2020, pp. 1572–1582.
[29] W. Christopher, S. Stephanie, M. Julie, and M. Kazuaki, “Ace 2005 multilingual training corpus,” 2006.
[30] D. Roth and W.-t. Yih, “A linear programming formulation for global inference in natural language tasks,” in Proceedings of the Eighth Conference on Computational Natural Language Learning (CoNLL-2004) at HLT-NAACL 2004, Boston, Massachusetts, USA, 2004, pp. 1–8.
[31] S. Zheng, F. Wang, H. Bao, and et al., “Joint extraction of entities and relations based on a novel tagging scheme,” in Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 1227–1236.
[32] X. Zeng, D. Zeng, S. He, and et al., “Extracting relational facts by an end-to-end neural model with copy mechanism,” in Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Melbourne, Australia, Jul. 2018, pp. 506–514.
[33] R. Takanobu, T. Zhang, J. Liu, and M. Huang, “A hierarchical framework for relation extraction with reinforcement learning,” in Proceedings of the AAAI Conference on Artificial Intelligence, pp. 7072–7079, 2019.
[34] T.-J. Fu, P.-H. Li, and W.-Y. Ma, “GraphRel: Modeling text as relational graphs for joint entity and relation extraction,” in Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 2019, pp. 1409–1418.