WILEY | Hindawi

Research Article

Few-Shot Relation Extraction via the Entity Feature Enhancement and Attention-Based Prototypical Network

Ren Li [ID]**, Qiao Xiao** [ID]**, Jianxi Yang** [ID]**, Hao Ren** [ID]**, and Yu Chen** [ID]

School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing 400074, China

Correspondence should be addressed to Hao Ren; haor@mails.cqjtu.edu.cn

In order to overcome the problems that the feature representation and classification effect of the existing methods need to be improved in complex contexts, this paper presents a novel few-shot relation extraction approach via the entity feature enhancement and attention-based prototypical network. The proposed model uses the pretrained RoBERTa model as the encoder while using the BiLSTM module for directional feature extraction. We further incorporate the entity feature enhancement module to improve the feature representation ability of the model. At last, the attention-based prototypical network is used to predict relations. The experimental results show that the proposed method not only outperforms the baseline models on the datasets from the bridge inspection and health domains but also achieves competitive results on the FewRel dataset in the general domain.

1. Introduction

Relation extraction (RE) is a fundamental task in NLP that has garnered significant attention from both academia and industry [1]. Recently, the development of pretrained language models (PLMs), such as BERT [2] and RoBERTa [3], has dramatically improved the performance of RE models. For instance, Su and Vijay-Shanker [4] proposed a BERT-based fine-tuning mechanism for biomedical RE, while Xue et al. [5] fine-tuned BERT for joint entity and relation extraction in Chinese medical text. Nonetheless, fine-tuning on pretrained models still relies on large-scale labeled data. In practice, annotating a large amount of training data is time-consuming and labor-intensive, particularly in specific domains where annotating RE datasets from scratch requires identifying named entities and annotating relations among them based on domain-specific knowledge. Therefore, achieving efficient RE in low-resource or few-shot scenarios is still a challenging task.

Currently, many NLP efforts based on few-shot learning have been proposed. The few-shot RE task aims to predict new entity relations by learning contextual features from a small number of labeled examples. Han et al. [6] introduced the FewRel dataset, which provides training data

and evaluation criteria for few-shot RE in general domains. In addition, Geng et al. [7] proposed the few-shot RE dataset, named TinyRel-CM, for the health domain. Based on these benchmark datasets, several few-shot RE approaches have been proposed, with metric learning-based methods being the mainstream in this field. Typical metric learning methods, such as the prototypical networks, aim to learn a low-dimensional embedding space and classify query instances directly by comparing the similarities between query instances and support instances, which achieves better performance and low time complexity [8]. Despite some progress made in few-shot RE research, the existing models are still insufficient to fully adapt to specific domains. The ability to learn complex contextual features and classify domain-specific entity relations accurately needs to be improved.

To address these issues, we propose a novel few-shot RE approach by introducing the entity-level feature enhancement and the attention-based prototypical network. Specifically, our model uses the pretrained RoBERTa model as the encoder and the bidirectional long short-term memory (BiLSTM) module for directional feature extraction. To help the model better extract the relations between entities, we further incorporate the entity feature enhancement module

to improve the feature representation ability. Finally, we employ an attention-based prototypical network as the relation prediction module. Our proposed model is better adapted to complex contexts, such as the context contains multiple relations, and one relation contains multiple instances of the same context. The contributions of this paper can be summarized as follows:

(1) We propose a new few-shot RE neural model that fuses directional features and entity features based on pretrained encoding in the context feature learning stage, thus improving the model's ability to represent domain-specific and RE task-aware features in complex contexts.

(2) Our proposed model introduces a relation prediction module via a novel attention-based prototypical network, which can improve the representation capability of different relation instances, and enhances the performance of relation prediction in the few-shot settings.

(3) Our experimental results demonstrate that our proposed approach outperforms baseline models on the domain-specific TinyRel-CM and Bridge-FewRel datasets. Furthermore, on the general domain-oriented FewRel dataset, our proposed model achieves better classification results in most few-shot settings.

The remainder of this paper is organized as follows: Section 2 provides an overview of related work, Section 3 presents the architecture of the proposed model and introduces each key component in detail, Section 4 describes the experimental setup and evaluation of our proposed approach, Section 5 conducts ablation experiments to analyse the impact of each component on the performance, and finally, Section 6 concludes the paper and presents future work.

2. Related Work

In recent years, many neural models for relation extraction (RE) have been proposed. Traditional methods mainly rely on the supervised learning or distant supervision strategies to train models [9]. The emergence of pretrained language models (PLMs) has significantly improved the performance of various NLP tasks. For instance, Guo et al. [10] developed a medical question answering system using PLMs and knowledge graphs, while Li et al. [11] proposed a semantic-enhanced multimodal fusion network for fake news detection, utilizing the pretrained BERT model as the text encoder. In this section, we briefly review the research efforts related to general RE and few-shot RE tasks, and then, we summarize their limitations.

2.1. General RE Approaches. The RE task is typically considered a multilabel classification problem where text sequences and entity pairs are input into the model, and the model predicts the classification results corresponding to predefined relation types. The current mainstream methods

can be classified into pipeline-based RE and joint learning-based RE [12].

Pipeline-based approaches treat RE as a downstream task of named entity recognition (NER). In recent years, various studies on neural RE have emerged. For example, Zeng et al. [13] used a CNN model to extract lexical and sentence-level features before classifying the predefined relations, while Zhang et al. [14] proposed a globally optimized LSTM neural model for RE. However, traditional pipelined models suffer from the problem of error propagation. Joint learning-based methods treat NER and RE as simultaneous optimization tasks. For instance, Bekoulis et al. [15] applied the adversarial training method in the joint extraction task, using Word2vec embeddings as the input. They employed BiLSTM and CRF to encode the words of sentences and decode the entity boundaries and tags, respectively. Eberts and Ulges [16] presented a span-based joint entity and relation extraction approach built on top of the pretrained BERT model. Our previous work [17] proposed a joint extraction approach via an entity-correlated attention neural model. Nevertheless, these methods require a large amount of labeled data during model training or fine-tuning, and the annotation of large-scale labeled instances is time-consuming and labor-intensive.

2.2. Few-Shot RE Approaches. To address the shortage of large-scale labeled data required for general RE research, several few-shot RE models have been proposed. Currently, metric learning is an effective approach for few-shot relation extraction. Among these metric learning algorithms, the prototypical network can integrate prior knowledge into the embedding space, and its model structure can meet the conditions of optimization in low-resource scenarios.

Gao et al. [18] proposed a hybrid attention-based prototypical network called HATT for noisy few-shot RE scenarios. The HATT model highlights key instances and features that are more relevant to the query instances to obtain a closer category prototype to each query instance, improving the model's robustness. Fan et al. [19] adopted a large-margin prototypical network with fine-grained features that can better identify long-tail relations. Ren et al. [20] proposed an area prototypical network with granularity-aware measurement, considering the different granularities of relations. However, these prototypical networks and their related variants did not consider the critical role of entity words, and not all sentences in the support set equally contributed to classifying relations.

To address these issues in few-shot RE, Li et al. [21] enhanced the prototype network with hybrid attention and confusing loss. Moreover, Lv et al. [22] proposed a domain-aware prototypical network (DPNet) to address interdisciplinary few-shot relation classification. Chen et al. [23] presented an open generalized prototypical network with task-adaptive feature fusion for open generalized few-shot relation classification. These approaches optimized the prototypical networks, but the capability of feature representation still needs improvement when dealing with complex contexts in a specific domain.

In recent years, several works have applied prototypical networks with pretrained language models (PLMs) to few-shot RE tasks. For example, Soares et al. [24] built on extensions of Harris' distributional hypothesis to relations and recent advances in learning text representations to build task-agnostic relation representations solely from entity-linked text. Yang et al. [25] enhanced the prototypical network with relation and entity descriptions, designing a collaborative attention module to extract beneficial and instructional information of sentences and entities, respectively. Han et al. [26] presented a contrastive learning-based approach that learns better representations by exploiting relation label information and allows the model to adaptively focus on hard few-shot RE tasks. Liu et al. [27] proposed a direct addition approach to introduce relation information, generating the relation representation by concatenating two views of relations and then directly adding it to the original prototype for both training and prediction. Peng et al. [28] proposed an entity-masked contrastive pretraining framework for RE to gain a deeper understanding of both textual context and type information while avoiding rote memorization of entities or using superficial cues in mentions.

To summarize, prototypical networks are the mainstream models for the few-shot relation extraction task. However, for relation extraction in the specific domains, e.g., bridge inspection and health, the context is more complex, which puts higher requirements on the feature representation ability. For example, these domain-specific sequences are longer, while the head entity and tail entity have a larger interval between each other. In addition, the positions of head entity and tail entity may be reversed, which are inconsistent with the predefined relations. So, it is necessary to capture entity features from both directions of the sequence. More challenging is that there are also multiple relations in these domain-specific contexts. The relation prediction performance of the existing few-shot methods needs to be improved. Our approach is motivated by dealing with the bottlenecks in the few-shot relation extraction task described above, that is, we fuse directional features and entity features based on the pretrained encoder. In addition, we introduce a relation prediction module via a novel attention-based prototypical network, improving the performance of few-shot relation classification in complex contexts.

3. Methodology

In this section, we will formally describe the few-shot RE task, and then present the overall architecture and key components of our proposed model.

3.1. Task Overview. RE aims to extract predefined relations between named entities from unstructured text, which is typically viewed as a classification problem based on predefined relations. Few-shot RE aims to train the model using a small amount of labeled data and address the long-tail relation problem by recognizing relations from few instances, which is also known as N-way-K-shot relation extraction. In the N-way-K-shot setting, "N-way" refers to the number of distinct relations that the model needs to extract, while "K-shot" refers to the number of annotated examples available for each relation, and the relations in the training set are not included in the test set. We formally define few-shot RE as follows.

In the few-shot RE task, a relation instance is represented as $s = (x, e^h, e^t, r)$, where x is the text sequence, and e^h and e^t denote the head and tail entities in the text sequence, respectively. r is the relation between two entities. Support sets $\widehat{S}$ and query sets $\widehat{Q}$ are used in each task of model training and testing. $\widehat{S} = \left\{ S_i = [S_{ij} = (x_{ij}, e_{ij}^h, e_{ij}^t, r_i)]_{j=1}^{K} \right\}_{i=1}^{N}$, where s_{ij} is the j-th relation instance in the relation type S_i, N is the total number of predefined relation types, and K is the number of instances related to each relation type. $\widehat{Q} = \left\{ q_j = (x_j, e_j^h, e_j^t, r_{jt})_{t=1}^{N} \right\}_{j=1}^{M}$, where q_j denotes the j-th query instance, r_{jt} is the t-th relation type to be predicted corresponding to the j-th instance, and M is the number of query instances in $\widehat{Q}$.

3.2. Model Architecture. Rather than pretraining a language model from scratch, we use the pretrained RoBERTa model as the encoder. In addition, we design a BiLSTM-based feature extraction module and an entity feature extraction model to enhance the context feature representation based on the characteristics of the domain-specific context and RE task. Figure 1 illustrates the overall architecture of our proposed model.

As shown in Figure 1, we use RoBERTa as the encoder for input sequences. Based on the definition of few-shot RE, the model input is a relation instance Ins with the form in equation (1), where t_i is the i-th token in the input sequence, $[E1_s]$ and $[E1_e]$ are the start tag and end tag of the head entity, $[E2_s]$ and $[E2_e]$ are the start tag and the end tag of tail entity, and e_x^h and e_x^t denote the x-th token of head entity and tail entity, respectively. $V_{\text{ins}} \in R^{l \times d}$ is the output matrix encoded by RoBERTa, where l is the sequence length of *Ins* and d is the embedding dimension of the RoBERTa model.

$$\text{Ins} = [CLS], t_1, \ldots, [E1_s], e_1^h, \ldots, e_x^h, [E1_e], t_m, \ldots, [E2_s], e_1^t, \ldots, e_x^t, [E2_e], t_n \ldots, [SEP], \tag{1}$$

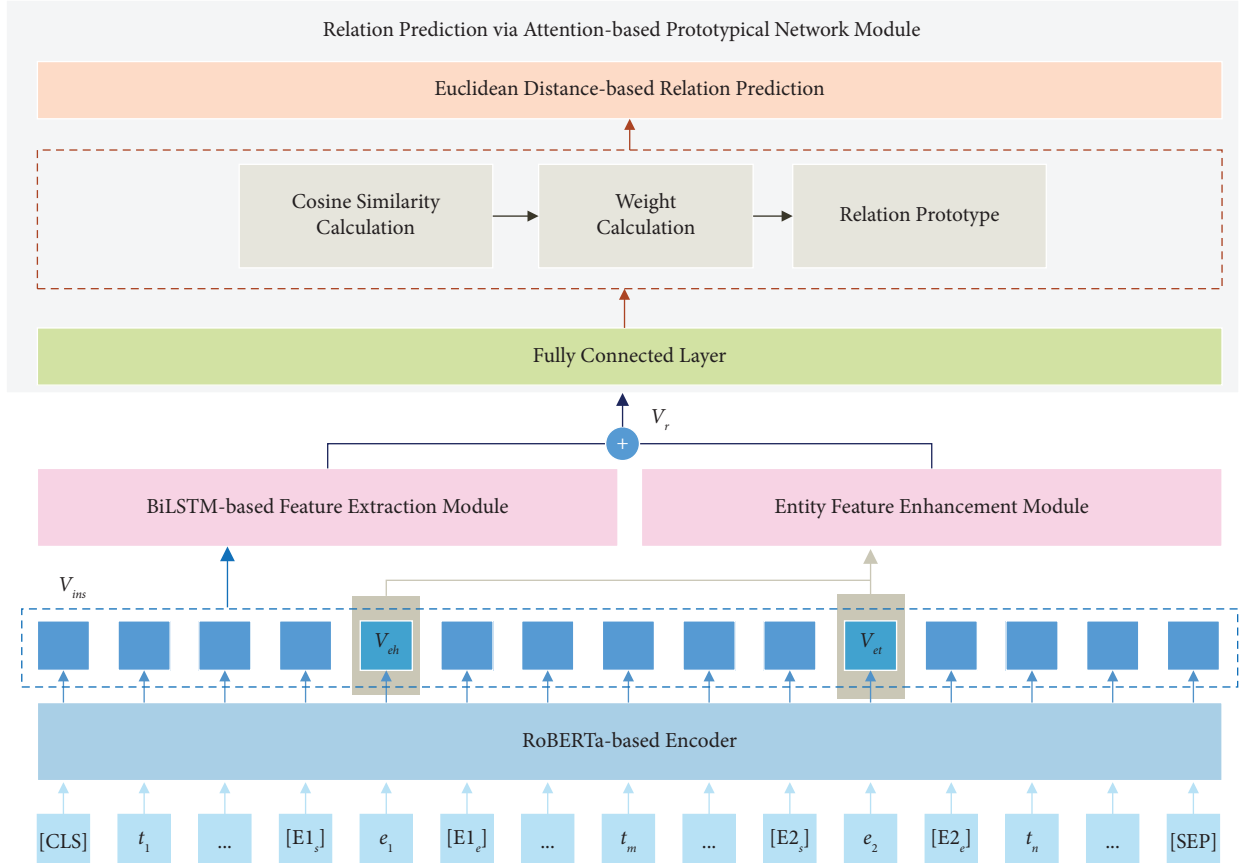$$V_{\text{ins}} = \text{RoBERTa}(\text{Ins}). \tag{2}$$

FIGURE 1: The overall architecture of the proposed model. It consists of a RoBERTa-based encoder, a BiLSTM-based directional feature extraction module, an entity feature enhancement module, and a relation prediction model via attention-based prototypical networks.

While using PLMs as encoders can effectively represent context features in most scenarios, certain domains such as healthcare and transportation have distinct context characteristics. For instance, in the TinyRel-CM [7] dataset for the health domain, more than half of the text sequences have over 50 tokens, whereas the sequence length in the FewRel dataset is always less than 40 tokens. In long text sequences, the span between head and tail entities significantly increases. In addition, in some domains, relation-type prediction relies on reverse context information. For example, in the bridge inspection report, extracting the location relation between a bridge member and structural defect may require reversing the order of the two entity types. Moreover, specific domains may involve multiple overlapping relations to be predicted, where an instance of context can contain multiple relations, or a relation can contain multiple instances of the same context. To improve the performance of few-shot RE, we further employ two feature enhancement modules with domain adaptability.

3.3. BiLSTM-Based Feature Extraction and Entity-Aware Feature Enhancement.

We use a BiLSTM neural model to extract the forward and backward features of the sequences, which is formally calculated as shown in equation (3), where

V_f and V_b are the forward and backward encodings generated by the BiLSTM.

$$V_f = \text{LSTM}_{\text{forward}}(V_{\text{ins}}) \in R^{1 \times d},$$
$$V_b = \text{LSTM}_{\text{backward}}(V_{\text{ins}}) \in R^{1 \times d}. \quad (3)$$

After that, we concatenate V_f and V_b and use the output of the fully connected layer and the tanh function as the final feature representation. The detailed calculation process is listed in equation (4), where W_1 and b_1 are trainable matrix and bias parameters.

$$V_c = [V_f; V_b] \in R^{1 \times 2d},$$
$$\tilde{V}_c = \tanh(V_c W_1 + b_1) \in R^{1 \times d}. \quad (4)$$

In addition, to better represent entity features and improve the effect of relation classification in few-shot settings, we further design an entity-aware feature enhancement module. It extracts the vector representation of the head and tail entities from the context embeddings V_{ins} via the corresponding labels. We use equation (5) to calculate the span vectors of head entity V_{eh} and tail entity V_{et}, where V_i^{eh} is the i-th token vector in the span of head entity, V_j^{et} corresponds

to the j-th token vector in the span of tail entity, and m and n are the span length of labeled head entity and tail entity, respectively.

$$V_{eh} = \frac{1}{m} \sum_{i=1}^{m} V_i^{eh},$$
$$V_{et} = \frac{1}{n} \sum_{j=1}^{n} V_j^{et}. \tag{5}$$

And then, we use a fully connected layer and a tanh function in equation (6) to calculate the encoding vectors of head and tail entities, denoted as $\widetilde{V}_{eh}$ and $\widetilde{V}_{et}$, respectively.

$$\widetilde{V}_{eh} = W_2 \left(\tanh\left(V_{eh}\right)\right) + b_2,$$
$$\widetilde{V}_{et} = W_3 \left(\tanh\left(V_{et}\right)\right) + b_3. \tag{6}$$

Finally, the final representations of a relation instance are concatenated as shown in the following equation:

$$V_r = \left[\widetilde{V}_c; \widetilde{V}_{eh}; \widetilde{V}_{et}\right]. \tag{7}$$

3.4. Relation Prediction via Attention-Based Prototypical Networks.

For relation prediction, we utilize an attention-based prototypical network. Formally, the encoded relation instances V_S in the support set and query instances V_Q in the query set are represented in equation (8), where N is the number of relation types in the support set, K is the number of instances related to each relation type, and M is the number of query instances in the query set. In addition, $V_r^{S_i^k}$ denotes the vector of k-th instance corresponding to the i-th relation type. $V_r^{q_j}$ is the vector of the j-th query instance.

$$V_S = \left\{ V_{S_i} = \left\{ V_r^{S_i^k} \right\} \mid i = 1, \ldots, N; k = 1, \ldots, K \right\},$$
$$V_Q = \left\{ V_r^{q_j} \mid j = 1, \ldots, M \right\}. \tag{8}$$

First, $V_r^{S_i^k}$ and $V_r^{q_j}$ are further encoded via a fully connected layer and a tanh function as listed in equation (9). The cosine similarity among $\widehat{V}_r^{q_j}$ and the vectors of all relation instances related to relation type S_i are calculated in equation (10).

$$\widehat{V}_r^{S_i^k} = \tanh\left(W_4 V_r^{S_i^k} + b_4\right),$$
$$\widehat{V}_r^{q_j} = \tanh\left(W_5 V_r^{q_j} + b_5\right), \tag{9}$$

$$\left\{ \mathrm{sim}_{kj} = \left(\widehat{V}_r^{S_i^k}, \widehat{V}_r^{q_j}\right) \mid i = 1, \ldots, N, k = 1, \ldots, K \right\}, \quad j = 1, \ldots, M. \tag{10}$$

To obtain the weight of relation instance in different relation types, the softmax function is used. For each relation type S_i, the weight among the relation instance and the j-th query instance is calculated in the following equation:

$$\alpha_z = \mathrm{softmax}\left(sim_{zj}\right) = \frac{\exp\left(sim_{zj}\right)}{\sum_{k=1}^{K} \exp\left(sim_{kj}\right)}. \tag{11}$$

The weight coefficient α_z represents the relevance between the z-th support instance and the query instance. All the instances of the relation type S_i are weighted and summed by the weight, which is used as the feature representation of the relation prototype $proto_i$ as listed in the following equation:

$$\mathrm{proto}_i = \sum_{k=1}^{K} \alpha_z \widehat{V}_r^{S_i^k}. \tag{12}$$

And then, we perform relation prediction by computing the Euclidean distance $D_{i,j}$ between the relation prototype and the query instance in the feature space. The max function is further used to predict the relation type, as listed in the following equation:

$$D_{i,j} = \sqrt{\sum_{i=1}^{N} \left(\mathrm{proto}_i - \widehat{V}_r^{q_j}\right)^2},$$
$$\widehat{r}_{jt} = \max\left(-1 \times D_{i,j}\right), \tag{13}$$

where $\widehat{r}_{jt}$ is the predicted relation type and t is the index of relations. Finally, the cross-entropy loss and interclass loss are used to optimize the model. The cross-entropy loss Loss_{cross} measures the discrepancy between the predicted and true relation types. Meanwhile, the interclass loss Loss_{class} quantifies the dissimilarity between representations of distinct relation types. By minimizing the cross-entropy loss, the model can achieve improved classification accuracy of relations. Similarly, by optimizing the interclass loss, the model can effectively differentiate and classify diverse input relations into distinct relation types. Equations (14)–(16) show the computational process of overall loss.

$$\text{Loss}_{\text{cross}}(r, \text{label}) = -\sum_{j=1}^{M} \log\left(\frac{\exp(r_j[\text{label}])}{\sum_i^N \exp(r_j[i])}\right), \tag{14}$$

$$\text{Loss}_{\text{class}} = \max\left(0, \ 1 - \frac{1}{2 \times N^2} \sum_i^N \sum_j^N \left(V_{S_i} - V_{S_j}\right)^2\right), \tag{15}$$

$$\text{Loss} = \text{Loss}_{\text{class}} + \text{Loss}_{\text{cross}}. \tag{16}$$

4. Experiments

In line with recent few-shot RE research efforts, such as MTB [24], HCPR [26], and HATT [18], we first evaluate our approach on the FewRel [6] and TinyRel-CM [7] datasets. Furthermore, we constructed a Chinese few-shot RE dataset for bridge inspection to further evaluate our approach's performance. In the following sections, we describe the datasets and experimental settings in detail and then present the experimental results compared to the baseline models. The source codes and data are available at https://github.com/Institute-of-BDKE-CQJTU/FRE.

4.1. Datasets. The FewRel dataset is a large-scale few-shot RE dataset constructed through distant supervision and manual labeling. It consists of 100 types of relations, each with 700 labeled instances. The dataset is split into subsets containing 64, 16, and 20 types of relations for training, validation, and testing, respectively. The TinyRel-CM dataset for the health domain contains four types of entities and 27 types of relations, with each relation corresponding to 50 instances. These two datasets are widely used benchmarks for evaluating few-shot RE models.

To further evaluate the proposed approach's performance in domain-specific scenarios, based on the previous works [17, 29], we constructed a Chinese few-shot RE dataset called Bridge-FewRel for the bridge inspection domain. The raw data for this dataset comes from over 1,300 Chinese bridge inspection reports. To meet practical engineering requirements, we define four types of named entities, namely (BRI) bridge name, ENT (bridge entity), DIS (structural defect), and ENTE (structural element), as well as 16 types of relations (as shown in Table 1) in Bridge-FewRel. Each relation type contains at least 100 relation instances.

In addition, to compare the differences between general datasets and domain-specific relation extraction datasets, we also added text length comparison information for the three datasets in Table 2. Compared to FewRel 1.0, the Bridge-FewRel and TinyRel-CM datasets have more long-text instances and a high proportion of long-text instances in TinyRel-CM, which makes the input text for domain-specific relation classification task more complex and more prone to interference.

Following the mainstream evaluation manners in research efforts [7, 18, 24, 26], the experiments are performed in the few-shot settings of 5-way-1-shot, 5-way-5-shot, 10-way-1-shot, and 10-way-5-shot on FewRel. For TinyRelA-CM, the few-shot settings of 15-way-K-shot are employed in the training stage, while *N*-way-5-shot, *N*-way-10-shot, and *N*-way-15-shot are employed in the testing stage. For the Bridge-FewRel dataset, we use 5-way-1-shot and 5-way-5-shot for evaluation.

4.2. Experimental Settings. We conduct experiments on a server with Intel i9-10900x CPU, 128 GB RAM, NVIDIA RTX 3090 GPU, Ubuntu 20.04, and CUDA 11.2. We implement the proposed neural network model based on PyTorch framework, with the open-source RoBERTa_chinese_ base serving as the encoder. Considering the sequence length and hardware limitations, we set max_length to 128 tokens for FewRel and Bridge-FewRel datasets and 256 tokens for the TinyRel-CM dataset. In addition, the encoding dimension is set to 768, learning rate to $1e-5$, and weigh decay to $1e-5$. The batch size is set to 4 for FewRel and Bridge-FewRel datasets, and the batch size for TinyRel-CM dataset is set to 2. To prevent overfitting, dropout with 0.2 is used after encoder.

For the FewRel dataset, we compare the proposed model with the emerging prototypical network-based few-shot RE approaches that have the comparable number of model parameters, such as LM-RrotoNet [19], TD-Proto [25], Direct-Addition [27], HATT [18], and HCPR [26]. For the TinyRel-CM dataset, the baseline models are prototypical (Proto) networks [30], such as BERT-PAIR(BP) [31], SNAIL [32], HATT [18], GNN [33], and MLMAN [34]. For the Bridge-FewRel dataset, we choose the open-source models as the baselines, which are MTB [24], HCPR [26], BERT-PAIR(BP) [31], and SNAIL [32].

4.3. Experimental Results. First, we followed the experimental method outlined in the research effort [7] to perform comparative experiments on the TinyRel-CM dataset. We evaluated the performance based on the six types of entity pairs, namely, D-D, D-S, D-F, D-U, F-U, and S-F. Table 3 shows the experimental results.

The model was trained with the 5-way-15-shot setting of the TinyRel-CM dataset. The settings of *N*-way-5-shot, *N*-way-10-shot, and *N*-way-15-shot were used for testing, where *N* represents the number of relation types for each entity pair. As shown in Table 3, our approach outperforms all baseline models in 15 out of 18 few-shot RE tasks on the TinyRel-CM dataset, while the HATT model only performs better in the other 3 settings.

Table 1: The 16 types of relations defined in the Bridge-FewRel dataset.

Entity pair	Relation type
Bridge name-bridge entity	Has_ENT and No_ENT
Bridge name-structural defect	Has_DIS and No_DIS
Bridge entity-bridge entity	Has_ENT and No_ENT
Bridge entity-structural defect	Has_DIS and No_DIS
Bridge entity-structural element	Has_ENTE and No_NETE
Structural defect-structural element	Has_LOC
Structural defect-bridge entity	Has_LOC
Structural element-structural element	Has_ENTE and No_NETE
Structural element-structural defect	Has_DIS and No_DIS

To further analyse the proposed model's advantages in domain-specific contexts, we conducted a statistical analysis of the TinyRel-CM dataset. Table 4 shows that for the six types of entity pairs in this dataset over 60% of the sequences have more than 50 tokens and nearly 40% of the sequences are more than 80 tokens long. Moreover, for D-D, D-U, and F-U entity pairs, the number of overlapping relations is 70, 48, and 29, respectively. These results highlight the complex context of the TinyRel-CM dataset and demonstrate that our proposed method is better suited for domain-specific complex contexts than the baseline models. The proposed model enhances the feature representation of relation instances and considers the contribution of more relevant support instances to the prototype representation of relations through an attention prototype network. This results in a more accurate relation prototype and achieves significant performance improvements.

The second experiment on the Bridge-FewRel dataset further validates the effectiveness of the proposed method in domain-specific contexts. Table 5 presents the classification accuracy results of the comparative experiments, which demonstrate that our method outperforms the mainstream open-source baseline models. For example, in the setting of 5-way-1-shot, the proposed model achieves accuracy of 84.11%, outperforming MTB by 3.36%.

We also conducted experiments on the FewRel dataset in the general domain. Table 6 presents the experimental results, where the comparison data of baseline models are sourced from research efforts [26, 28]. It is worth noting that the results of LM-ProtoNet and TD-Proto are from the test set, while others are from the validation set.

The experimental results presented in Table 6 demonstrate that the proposed method also achieves competitive results on few-shot RE tasks in general domain contexts. For example, compared with HATT, BERT + ProtoNET, LM-ProtoNet, and TD-Protocol, our approach achieves better accuracy in all few-shot settings. While HCPR and Direct-Addition outperform our model in the 5-way-1-shot and 10-way-1-shot settings, our model outperforms Direct-Addition in the 5-way-5-shot and 10-way-5-shot settings, with an increased accuracy of relation extraction by 0.93%

and 1.04%, respectively. Furthermore, in these two few-shot settings, our model also outperforms HCPR on the FewRel dataset.

4.4. Case Analysis. To evaluate the actual outputs of different models, we selected several instances from the support sets and query sets and analysed their representative errors. Figure 2 shows the relation prediction results of BERTEM-ProtoNet, MTB, and our model, where the red characters denote errors in prediction results.

As shown in Figure 2, our model achieved better prediction results than the baseline models. Specifically, in Case 1, the BERTEM-ProtoNet and MTB models failed to identify the ENT and ENTE entities, which lead to incorrect prediction of relation type. In Case 2, the input instance contains multiple relations, including "Has_ENTE[ENT-ENTE]," "Has_DIS[ENTE-DIS]," and "Has_LOC[DIS-ENT]." This makes it difficult for the model to distinguish the relation type corresponding to the current entity pair, resulting in wrong relation type prediction in the BERTEM-ProtoNet and MTB models. In Case 3, when adding the given support set instance into the support set in the training stage, there are multiple identical contexts in the support set of the same relation type. In the testing stage, when the given query set instance is input into the few-shot RE model, the baseline models misclassify it as "Contain [Food-Nutrient]" due to the occurrence of too many identical contexts.

5. Ablation Study

To further verify the effect of each module in the proposed method on the few-shot relation extraction performance, some ablation studies are performed on the domain-specific datasets. Table 7 shows the ablation results, where the term "Ours(RoBERTa)" denotes the proposed model and "Ours(BERT)" represents the model that replaces the encoder with BERT. In addition, "w/o. attention-proto" represents the model after replacing the proposed attention prototypical network module with the basic prototypical network. "w/o. BiLSTM" and "w/o. entity-features" denote the models after removing the corresponding feature enhancement modules, respectively.

As shown in Table 7, when using the BERT model as the encoder, the accuracy of the model decreased in all three few-shot settings. In addition, when we replace the attention-based prototypical network with the original prototypical network, the classification accuracy decreases by 0.35% and 0.65% on the Bridge-FewRel dataset, respectively. The accuracy is also reduced by 1.19% in the 5-way-5-shot setting of D-F entity pairs on the TinyRel-CM dataset. When the BiLSTM module and the entity feature enhancement module are not used, the performance of the model also decreases. Especially, when the entity feature enhancement module is removed, the performance decreases by 8.73%, 7.56%, and 3.49% in the three settings,

TABLE 2: Number and proportion of text length (in terms of number of characters) intervals in the three datasets.

Datasets	Length	0–20	20–30	30–40	40–60	60–80	>80
Bridge-FewRel	Num.	7926	3850	2195	4280	142	57
	Prop.	43%	21%	12%	23%	0.8%	0.2%
TinyRel-CM	Num.	28	93	174	348	290	424
	Prop.	2%	7%	12.8%	25.6%	21%	31.6%
FewRel 1.0	Num.	14905	27857	13238	0	0	0
	Prop.	27%	50%	23%	0%	0%	0%

TABLE 3: Experimental results on the TinyRel-CM dataset.

Models	Group	D-D	D-S	D-F	D-U	F-U	S-F
	N	$N=5$	$N=2$	$N=6$	$N=6$	$N=2$	$N=6$
GNN	$K=5$	26.02	66.22	37.48	44.47	55.02	37.13
	$K=10$	27.96	58.75	37.88	45.15	55.53	38.39
	$K=15$	28.36	70.10	43.25	51.40	56.45	40.37
SNAIL	$K=5$	24.64	58.55	28.57	29.05	53.30	27.88
	$K=10$	27.38	57.12	27.85	33.50	55.60	30.52
	$K=15$	23.84	57.40	30.63	35.75	56.35	30.43
Proto	$K=5$	31.50	69.56	34.36	44.76	57.43	40.36
	$K=10$	35.45	72.54	37.11	48.28	58.71	43.55
	$K=15$	38.28	72.79	37.90	50.03	60.09	44.97
HATT	$K=5$	38.21	**75.00**	46.67	54.17	**73.75**	44.17
	$K=10$	44.45	71.24	49.37	51.67	75.00	49.17
	$K=15$	49.61	**80.41**	44.33	56.67	72.58	49.17
MLMAN	$K=5$	38.17	64.98	45.32	51.30	64.24	48.34
	$K=10$	44.89	70.02	50.83	56.60	70.79	53.69
	$K=15$	49.23	72.41	54.25	59.67	74.58	56.67
BERT-PAIR	$K=5$	26.87	59.52	52.54	43.28	63.33	49.95
	$K=10$	28.19	62.32	55.86	46.13	66.70	52.92
	$K=15$	29.25	63.98	57.60	47.81	68.83	54.16
Ours	$K=5$	**52.47**	66.98	**69.25**	**71.03**	71.21	**65.20**
	$K=10$	**60.63**	**72.17**	**71.41**	**75.76**	**78.36**	**69.15**
	$K=15$	**65.40**	73.73	**71.76**	**78.20**	**80.38**	**71.22**

Acc. %. The bold values represent the state-of-the-art results under the current settings.

TABLE 4: Statistics on the TinyRel-CM dataset.

	Tokens	D-D	D-S	D-F	D-U	F-U	S-F
Proportion of long sequences	>50	0.64	0.71	0.63	0.76	0.75	0.63
	>60	0.53	0.55	0.49	0.63	0.59	0.49
	>80	0.29	0.33	0.28	0.38	0.36	0.34
Number of overlapping relations	—	70	0	0	48	29	0

respectively. These ablation results illustrate the effect of feature fusion on accuracy improvement in domain-specific complex contexts.

At last, we explored the performance changes from two aspects of the choice of metric function and the number of instances in the support set. Table 8 shows the comparative results of different metric functions in relation prediction. Table 9 represents the effect of different number of instances in the support sets on the performance.

As shown in Table 8, when the Euclidean distance is used as the metric for relation classification, the classification accuracy of the model is better than that of the similar-based metric. Moreover, as shown in Table 9, as the value of K increases from 1 to 5, the performance of the model improves greatly on both datasets. The experimental results show that the number of instances in the support set is also the main factor affecting the performance of the few-shot RE model.

TABLE 5: Experimental results on the Bridge-FewRel dataset.

Models	5-way-1-shot	5-way-5-shot
BERT-PAIR	42.19	44.70
SNAIL	61.20	63.35
BERT-ProtoNet	69.45	80.74
HCPR	71.64	84.53
BERT$_{EM}$-ProtoNet	77.42	85.08
MTB	80.75	90.93
Ours	84.11	92.12

Acc. %.

TABLE 6: Experimental results on the FewRel dataset.

Model	5-way-1-shot	5-way-5-shot	10-way-1-shot	10-way-5-shot
BERT + ProtoNet	82.92	91.32	73.24	83.68
LM-ProtoNet	76.60	89.31	65.31	82.10
TD-Proto	84.76	92.38	74.32	85.92
HATT	75.01	87.09	62.48	77.50
HCPR	90.90	93.22	84.11	87.79
Direct-Addition	**91.29**	94.05	**86.09**	89.68
Ours	85.31	**94.98**	76.28	**90.72**

Acc. %. The bold values represent the state-of-the-art results under the current settings.

- Case Analysis 1: The long-text instance (Translated from the Bridge-FewRel Dataset)
  Query set instance: [The Bitou Bridge]BRI is a 3×7.5m solid-web masonry arch bridge. [Superstructure]ENT : the span of main arch ring is 7.5m, the height of rise arch is 2.5m, the thickness of the main arch ring is 0.45m, and the width of the arch ring is 5.0m.
  True Answer: Has_ENT[BRI-ENT]   MTB: Has_ENTE[BRI-ENTE]
  BERTEM-ProtoNet: Has_ENTE[BRI-ENTE]   Ours: Has_ENT[BRI-ENT]

- Case Analysis 2: One context contains multiple relations (Translated from the Bridge-FewRel Dataset)
  Query set instance: [Transverse cracks]DIS exist in [the wing plate]ENTE of [box girder 28-1#]ENT, located in the [right wing Plate]ENTE 29m away from [28# pier]ENT.
  True Answer:Has_DIS[ENTE-DIS]   MTB: Has_DIS[ENT-DIS]
  BERTEM-ProtoNet: Has_DIS[ENT-DIS]   Ours: Has_DIS[ENTE-DIS]

- Case Analysis 3: One relation contains multiple instances of the same context (Translated from the TingyRel-CM dataset)
  Support set instance: Fourthly, due to the consumption caused by the disease and the decline in digestive and absorptive functions, most patients will experience malnutrition. Therefore, their diet should include sufficient carbohydrates and protein-rich foods, such as rice, noodles, soy products, lean meat, poultry, fish, etc., as well as adequate intake of vegetables and fruits to supplement vitamins and minerals.
  Contain[Food-Nutrient]: Rice-Carbohydrate, Rice-Protein, Noodles-Carbohydrate, Noodles-Protein, Soy products-Carbohydrate, Soy products-Protein, Lean meat-Carbohydrate, Lean meat-Protein, Fish-Carbohydrate, Fish-Protein, Vegetables-Carbohydrate, Vegetables-Minerals, Fruits-Vitamins, Fruits-Minerals.
  Query set instance: [The lecithin]Nutrient contained in tofu strips can remove cholesterol attached to blood vessel walls, prevent arteriosclerosis, and protect the heart, thus preventing [cardiovascular disease]Disease.
  True Answer: Prevent[Disease-Nutrient]   MTB: Contain[Food-Nutrient]
  BERTEM-ProtoNet: Contain[Food-Nutrient]   Ours: Prevent[Disease-Nutrient]

FIGURE 2: Case studies on the Bridge-FewRel and TinyRel-CM datasets.

TABLE 7: Ablation results on the Bridge-FewRel and TinyRel-CM datasets.

| | Bridge-FewRel | | TinyRel-CM |
	5-way-1-shot	5-way-5-shot	D-F (5-way-5-shot)
Ours (RoBERTa)	84.11	92.12	78.66
Ours (BERT)	83.04	91.09	76.88
w/o. attention-proto	83.76	91.47	77.47
w/o. BiLSTM	82.32	90.26	76.33
w/o. entity-features	75.38	84.56	75.17

Acc. %.

TABLE 8: Comparison of different metric functions in relation prediction.

| | Bridge-FewRel | | FewRel | |
	5-way-1-shot	5-way-5-shot	5-way-1-shot	5-way-5-shot
Similarity-based metric	83.07	91.74	84.96	92.29
Euclidean-based metric	84.11	92.12	85.31	94.98

Acc. %.

TABLE 9: Experimental results in 5-way-K-shot settings.

	1-shot	2-shot	3-shot	4-shot	5-shot
Bridge-FewRel	84.11	85.18	86.74	89.90	92.12
FewRel	85.31	90.80	93.07	93.60	94.98

Acc. %.

6. Conclusion

Few-shot relation extraction is an important research task in NLP, but existing methods have shortcomings in complex context feature representation and relation prediction performance. This paper presents a novel few-shot RE approach via the entity feature enhancement and attention-based prototypical network. Our model uses the pretrained RoBERTa model as the encoder while using the BiLSTM module for directional feature extraction. We further incorporate the entity feature enhancement module to improve the feature representation ability of the model. At last, the attention-based prototypical network is used to predict relations. The experimental results show that the proposed method not only outperforms the baseline models on the datasets from the bridge inspection and health domains but also achieves competitive results on the FewRel dataset in the general domain.

Data Availability

The source codes and data are available at https://github.com/Institute-of-BDKE-CQJTU/FRE.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

References

[1] T. H. Nguyen and R. Grishman, "Relation extraction: perspective from convolutional neural networks," in *Proceedings of the 1st workshop on vector space modeling for natural language processing*, pp. 39–48, Denver, CO, USA, June 2015.

[2] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, "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*, pp. 4171–4186, Minneapolis, MN, USA, June 2019.

[3] Y. Liu, M. Ott, N. Goyal et al., "Roberta: a robustly optimized bert pretraining approach," 2019, https://arxiv.org/abs/1907.11692.

[4] P. Su and K. Vijay-Shanker, "Investigation of bert model on biomedical relation extraction based on revised fine-tuning mechanism," in *Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2522–2529, Seoul, South Korea, December 2020.

[5] K. Xue, Y. Zhou, Z. Ma, T. Ruan, H. Zhang, and P. He, "Fine-tuning BERT for joint entity and relation extraction in Chinese medical text," in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 892–897, San Diego, CA, USA, November 2019.

[6] X. Han, H. Zhu, P. Yu et al., "FewRel: a large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 4803–4809, Brussels, Belgium, November 2018.

[7] X. Geng, X. Chen, K. Q. Zhu, L. Shen, and Y. Zhao, "Mick: a meta-learning framework for few-shot relation classification with small training data," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 415–424, Virtual Event, Ireland, October 2020.

[8] X. Liu, K. Zhou, and Y. He, "Survey of metric-based few-shot classification," *Pattern Recognition and Artificial Intelligence*, vol. 34, no. 10, pp. 909–923, 2021.

[9] P. Zhou, J. Xu, Z. Qi, H. Bao, Z. Chen, and B. Xu, "Distant supervision for relation extraction with hierarchical selective attention," *Neural Networks*, vol. 108, pp. 240–247, 2018.

[10] Q. Guo, S. Cao, and Z. Yi, "A medical question answering system using large language models and knowledge graphs," *International Journal of Intelligent Systems*, vol. 37, no. 11, pp. 8548–8564, 2022.

[11] S. Li, T. Yao, S. Li, and L. Yan, "Semantic-enhanced multimodal fusion network for fake news detection," *International Journal of Intelligent Systems*, vol. 37, no. 12, pp. 12235–12251, 2022.

[12] Y. Yuan, X. Zhou, S. Pan, Q. Zhu, Z. Song, and L. Guo, "A relation-specific attention network for joint entity and relation extraction," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pp. 4054–4060, Yokohama Yokohama, Japan, January 2021.

[13] D. Zeng, K. Liu, S. Lai, G. Zhou, and J. Zhao, "Relation classification via convolutional deep neural network," in *Proceedings of the COLING 2014, the 25th international conference on computational linguistics: technical papers*, pp. 2335–2344, Dublin, Ireland, August, 2014.

[14] M. Zhang, Y. Zhang, and G. Fu, "End-to-end neural relation extraction with global optimization," in *Proceedings of the 2017 conference on empirical methods in natural language processing*, pp. 1730–1740, Copenhagen, Denmark, September 2017.

[15] G. Bekoulis, J. Deleu, T. Demeester, and C. Develder, "Adversarial training for multi-context joint entity and relation extraction," in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pp. 2830–2836, Brussels, Belgium, November 2018.

[16] M. Eberts and A. Ulges, "Span-based joint entity and relation extraction with transformer pre-training," in *Proceedings of the ECAI 2020, the 24th European Conference on Artificial Intelligence*, pp. 2006–2013, Santiago de Compostela, Spain, September 2020.

[17] R. Li, D. Li, J. Yang et al., "Joint extraction of entities and relations via an entity correlated attention neural model," *Information Sciences*, vol. 581, pp. 179–193, 2021.

[18] T. Gao, X. Han, Z. Liu, and M. Sun, "Hybrid attention-based prototypical networks for noisy few-shot relation classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 01, pp. 6407–6414, 2019.

[19] M. Fan, Y. Bai, M. Sun, and P. Li, "Large margin prototypical network for few-shot relation classification with fine-grained features," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, pp. 2353–2356, Beijing, China, November 2019.

[20] H. Ren, Y. Cai, R. Y. K. Lau, H. Leung, and Q. Li, "Granularity-aware area prototypical network with bimargin loss for few shot relation classification," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 1–4866, 2022.

[21] Y. Li, Z. Ma, L. Gao, Y. Wu, F. Xie, and X. Ren, "Enhance prototypical networks with hybrid attention and confusing loss function for few-shot relation classification," *Neurocomputing*, vol. 493, pp. 362–372, 2022.

[22] B. Lv, L. Jin, X. Li et al., "DPNet: domain-aware prototypical network for interdisciplinary few-shot relation classification," *Applied Intelligence*, vol. 52, no. 13, pp. 15718–15733, 2022.

[23] X. Chen, G. Wang, H. Ren, Y. Cai, H. Leung, and T. Wang, "Task-adaptive feature fusion for generalized few-shot relation classification in an open world environment," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 30, pp. 1003–1015, 2022.

[24] L. B. Soares, N. Fitzgerald, J. Ling, and T. Kwiatkowski, "Matching the blanks: distributional similarity for relation learning," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2895–2905, Florence, Italy, July 2019.

[25] K. Yang, N. Zheng, X. Dai, L. He, S. Huang, and J. Chen, "Enhance prototypical network with text descriptions for few-shot relation classification," in *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pp. 2273–2276, Virtual Event, Ireland, October 2020.

[26] J. Han, B. Cheng, and W. Lu, "Exploring task difficulty for few-shot relation extraction," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 2605–2616, Punta Cana, Dominican Republic, November 2021.

[27] Y. Liu, J. Hu, X. Wan, and T. H. Chang, "A simple yet effective relation information guided approach for few-shot relation extraction," in *Proceedings of the Findings of the Association for Computational Linguistics: ACL 2022*, pp. 757–763, Dublin, Ireland, May 2022.

[28] H. Peng, T. Gao, X. Han et al., "Learning from context or names? An empirical study on neural relation extraction," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, pp. 3661–3672, November 2020.

[29] R. Li, T. Li, and J. Yang, "Bridge inspection named entity recognition based on Transformer-BiLSTM-CRF," *Journal of Chinese Information Processing*, vol. 35, no. 4, pp. 88–96, 2021.

[30] J. Snell, K. Swersky, and R. Zemel, "Prototypical networks for few-shot learning," in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4080–4090, California, CA, USA, December 2017.

[31] T. Gao, X. Han, and H. Zhu, "FewRel 2.0: towards more challenging few-shot relation classification," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, pp. 6250–6255, Hong Kong, China, November 2019.

[32] N. Mishra, M. Rohaninejad, and X. Chen, "A simple neural attentive meta-learner," in *Proceedings of the International Conference on Learning Representations*, Vancouver, Canada, May 2018.

[33] V. Garcia and J. Bruna, "Few-shot learning with graph neural networks," in *Proceedings of the International Conference on Learning Representations*, Vancouver, Canada, May 2018.

[34] Z. X. Ye and Z. H. Ling, "Multi-level matching and aggregation network for few-shot relation classification," in *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pp. 2872–2881, Florence, Italy, July 2019.