Available Online: 14 April 2026 DOI: 10.54254/2977-3903/2026.32815

A continual relation extraction method based on prompt learning and similar-relation-aware optimization

Yongji Dai**, *Na Li

University of Shanghai for Science and Technology, Shanghai, China

*Corresponding Author. Email: 232340562@st.usst.edu.cn

Abstract. Continual relation extraction aims to learn newly arriving relation types while preserving previously acquired knowledge, but it remains challenging due to catastrophic forgetting and severe confusion among semantically similar relations. To address this problem, we propose a continual relation extraction method that integrates prompt learning, feature enhancement, anchoring loss, and similarity-aware prototype contrastive learning. Specifically, relation-specific soft prompts and dynamic prompt selection are introduced to provide targeted semantic guidance for different samples. Based on the matched prompts, both relation prototypes and sample representations are further enhanced to improve feature quality and discriminability. In addition, an anchoring loss and a similarity-aware prototype contrastive objective are designed to explicitly optimize the boundaries between semantically similar relations during memory replay. Experiments on the FewRel and TACRED datasets show that the proposed method consistently outperforms representative baseline methods, especially in later task stages where continual learning becomes more challenging. Ablation studies and visualization results further verify the effectiveness of prompt-guided feature enhancement and similar-relation-aware optimization. These results indicate that explicitly strengthening fine-grained semantic boundaries, in addition to preserving historical knowledge, is important for continual relation extraction, and the proposed method provides an effective solution for improving both knowledge retention and relation discrimination.

Keywords: continual relation extraction, prompt learning, feature enhancement, similarity-aware prototype contrastive learning, anchoring loss

1. Introduction

Relation Extraction (RE) aims to identify the semantic relation between a pair of entities from unstructured text, and it serves as a fundamental technique for knowledge graph construction, question answering, information retrieval, and text understanding. Its core objective is to determine the relation category expressed between a given entity pair according to contextual semantics. As illustrated in Figure 1, the model needs to infer the semantic relation between entities from the surrounding context, such as recognizing the relation "per:city_of_death" between a person entity and a city entity, or "org:founded_by" between an organization and a person. Therefore, relation extraction depends not only on entity information itself, but also heavily on the ability to model fine-grained contextual semantic cues.

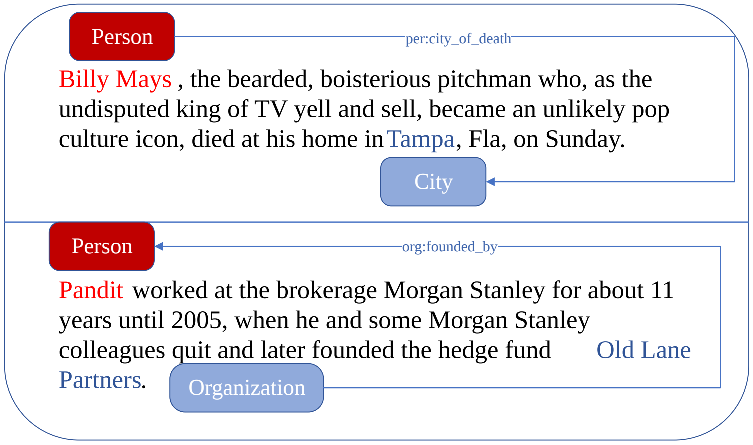


Figure 1. Relation extraction task on TACRED dataset

In recent years, with the rapid development of pre-trained language models, relation extraction under fixed closed-set settings has achieved remarkable progress. However, in real-world applications, relation categories are usually not given once and kept unchanged; instead, they continually expand as application requirements evolve and data accumulate. As a result, models are required to continually acquire new relation categories in a continual learning manner. To address this issue, researchers have proposed a variety of continual relation extraction methods, including EMAR, which is based on episodic memory activation and reconsolidation [1], RP-CRE, which refines sample representations with relation prototypes [2], and CRL, which combines consistent representation learning with memory replay [3]. Subsequent studies further explored improvements from the perspectives of adversarial class augmentation, prototypical contrastive learning, and analogous semantic discrimination [4-6]. Building on these advances, later work has extended continual relation extraction toward classifier decomposition, rehearsal-free ensemble-of-experts learning, memory structure preservation, static relation prototypes, and prompt-based modeling [7-13]. These developments have gradually established continual relation extraction as an important research direction in relation extraction. In essence, continual relation extraction requires a model to learn new relational knowledge while preserving the ability to recognize previously learned relations, thereby alleviating performance degradation caused by incremental learning [1, 3, 6].

Compared with conventional static relation extraction, continual relation extraction faces more severe challenges. On the one hand, when the model learns new tasks sequentially, parameter updates may overwrite previously acquired knowledge, resulting in catastrophic forgetting. On the other hand, many relation categories in relation extraction are semantically similar, and the boundaries between fine-grained relations are often ambiguous, making the model more vulnerable to relation confusion during continual learning [6-10]. As shown in Figure 2, on the FewRel dataset [14], the CEAR model [6] based on the pre-trained language model BERT [15] exhibits clear local overlap among some learned relation features. The red-circled regions indicate that several relation categories are not effectively separated in the representation space. This suggests that although existing methods can learn relation representations to some extent, they still fail to construct sufficiently clear discriminative boundaries for semantically similar relations. Such feature overlap can be further aggravated in continual learning scenarios, eventually weakening both old-knowledge retention and new-relation discrimination.

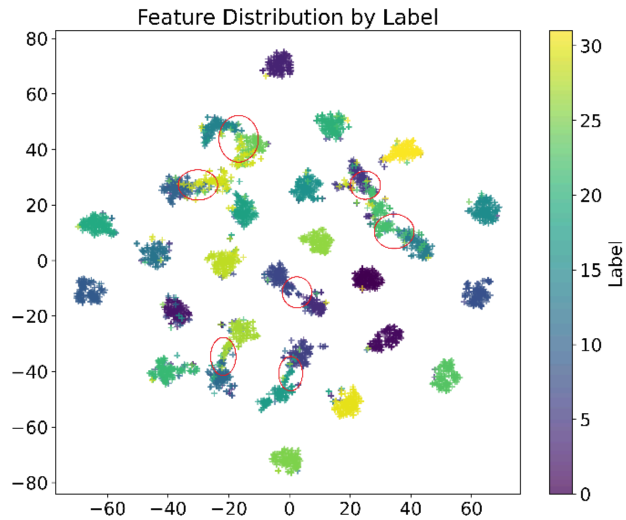


Figure 2. Feature distribution of the CEAR model on FewRel dataset

Most existing continual relation extraction methods mainly rely on sample replay, parameter constraints, or prototype memory. Although these methods have shown effectiveness in mitigating forgetting, they are still insufficient for modeling fine-grained boundaries among semantically similar relations. Meanwhile, prompt learning [16-18], as an important paradigm for adapting pre-trained models in recent years, introduces trainable prompt vectors to guide the model toward task-relevant semantic information and thus provides a new optimization perspective for continual relation extraction. If discriminative prompt representations can be constructed for different relation categories and dynamically matched with input samples for semantic enhancement, it may become possible to simultaneously improve new-knowledge acquisition and old-knowledge retention during continual learning.

To this end, we propose a continual relation extraction method that integrates prompt learning, feature enhancement, anchoring regularization, and similarity-aware prototype contrastive learning. Specifically, we introduce relation-specific soft prompts and dynamic prompt selection to provide relation-aware semantic guidance for input samples. Based on the matched prompts, we further enhance both prototype representations and sample representations to improve the quality and discriminability of relation features. In addition, we design a similarity-aware prototype contrastive learning objective and an anchoring loss to explicitly enlarge the boundaries between semantically similar relations. By jointly incorporating these mechanisms, the proposed method improves relation representation learning, alleviates catastrophic forgetting, and reduces confusion among similar relations in continual relation extraction.

The main contributions of this paper are as follows:

(1) We introduce a prompt learning mechanism for continual relation extraction, where relation-specific soft prompts are constructed and dynamically matched to input samples to provide more targeted semantic guidance during continual learning.

(2) We propose feature enhancement strategies for both relation prototypes and sample representations, which improve feature quality and strengthen discriminability from class-level and instance-level perspectives.

(3) We design an anchoring loss and a similarity-aware prototype contrastive learning objective to explicitly optimize the boundaries between semantically similar relations, thereby alleviating catastrophic forgetting and similar-relation confusion.

2. Related work

Existing research on continual relation extraction has evolved from early studies centered on sample replay, parameter regularization, and prototype memory to more recent directions such as representation regularization, knowledge distillation, prompt enhancement, and data augmentation [19]. Early methods mainly alleviated catastrophic forgetting and improved the retention of previously learned relations through mechanisms including episodic memory activation and reconsolidation, relation prototype refinement, consistent representation learning, and analogous semantic discrimination [1-6]. Building on these foundations, later studies further advanced continual relation extraction from the perspectives of parameter isolation, structural preservation, and task adaptation, including classifier decomposition, spectral analysis, rehearsal-free ensemble-of-experts learning, memory structure preservation, and static relation prototype constraints [7-11]. Meanwhile, MREM, which incorporates historical memory samples as prompts during modeling [12], and adaptive prompting based on within-task variance [13], also indicate that prompt learning has begun to demonstrate clear potential in continual relation extraction.

Knowledge distillation and knowledge transfer have also enriched the methodological landscape of continual few-shot relation extraction in recent years [20-25]. Existing studies improve cross-task knowledge retention and representation discrimination by combining distillation with contrastive learning, prototype augmentation, adaptive gradient correction, and knowledge decomposition [22-25].

Substantial progress has likewise been made in prototype modeling and contrastive learning. These studies enhance the stability of class prototype representations, refine relation representations through dynamic prototypes and contrastive fine-tuning, and promote a clearer relation representation space, thereby improving performance in low-resource and few-shot settings [26-30].

More recently, prompt learning and data augmentation have become important research directions . Prior work has improved semantic modeling in few-shot relation extraction through prompt-based adaptation, relation-aware prompt learning, graph-based prompt tuning, and mutual-pairing data augmentation [31-34]. However, recent research shows that replay- or prototype-based constraints alone are still insufficient to fully resolve the fine-grained confusion among semantically similar relations in continual learning [35]. Different from previous methods, the proposed approach jointly incorporates relation-specific soft prompts, dynamic prompt matching, prototype and instance feature enhancement, and a similar-relation representation anchoring constraint to improve both relation discrimination and historical knowledge retention in continual relation extraction.

3. Method

3.1. Problem definition

We study continual relation extraction under a sequence of tasks $\{T_1, T_2, \ldots, T_K\}$. Each task contains a new relation set, a training set, and a test set, and the relation sets of different tasks are mutually exclusive. The goal is to learn new relations in the current task while preserving the ability to recognize all previously learned relations. To this end, we maintain a memory buffer that stores Typical sample of learned relations and replay them in subsequent tasks.

3.2. Overall framework

The overall architecture of our method is shown in Figure 3. we build a closed-loop framework with three stages: new-task learning, relation-prompt dynamic selection with feature enhancement, and memory replay optimization. In the first stage, we learn initial sample representations for the current task and initialize relation prototypes. In the second stage, we construct relation-specific soft prompts and dynamically match them to input samples, followed by prototype and sample feature enhancement. In the third stage, we replay stored sample and jointly optimize the model with classification, similarity-aware prototype contrastive learning, representation anchoring, and prompt matching. This design allows the model to balance new-knowledge acquisition, historical-knowledge retention, and discrimination among semantically similar relations.

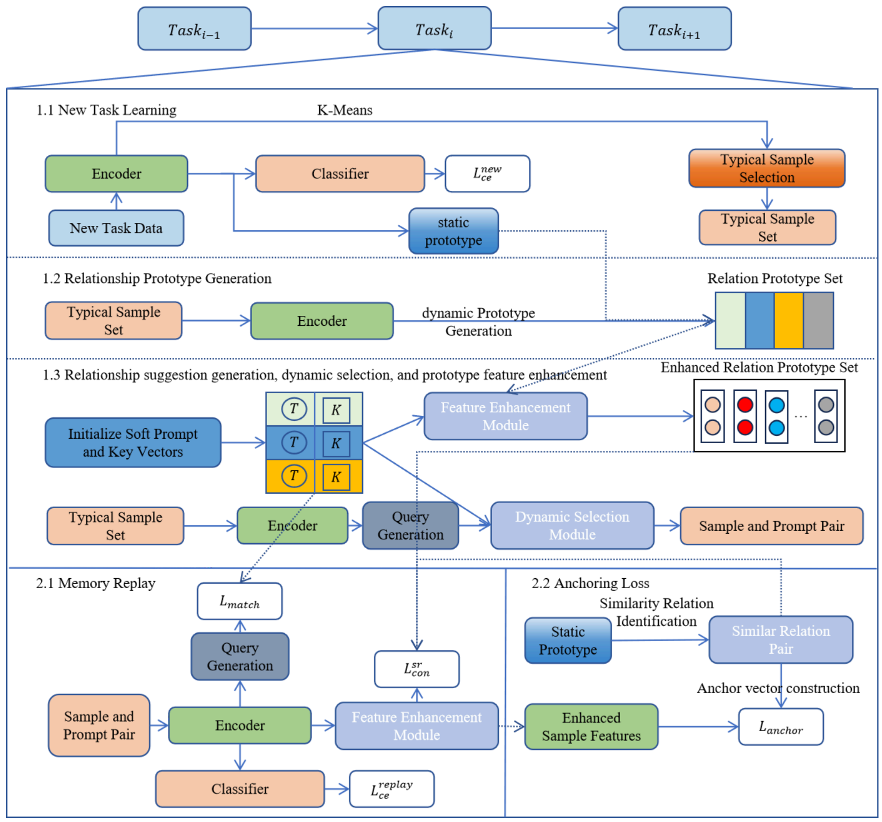


Figure 3. Overall architecture of the model

3.3. New-task learning

3.3.1. Data preprocessing and encoder initialization

Since relation extraction focuses on the semantic interaction between a target entity pair rather than the sentence as a whole, data preprocessing should explicitly highlight the positions of the head entity and the tail entity. Given an input sample $x=(s,e_h,e_t)$, where s denotes the sentence, and e_h and e_t denote the head and tail entities, respectively, we insert special entity markers $[E11]$, $[E12]$, $[E21]$, and $[E22]$ around the two entities. In this way, the encoder can focus more directly on the target entity pair and its contextual semantics.

The marked sentence is then fed into the encoder to obtain the hidden representations of the two entity-start markers. We concatenate these two vectors to form the semantic representation of the sample:

$$h_x = Concat(BERT(x)_{[E11]}, BERT(x)_{[E21]}) \tag{1}$$

where $Enc(\cdot)$ denotes the pre-trained encoder SimCSE-BERT [36]. Based on h_x , a linear classifier followed by a Softmax layer is used to predict the relation label:

$$P_{cls} = Softmax(W \cdot h_x + b) \tag{2}$$

where W and b are trainable parameters. During this stage, the model is optimized only on the current task by the standard cross-entropy loss:

$$L_{ce}^{new} = -\frac{1}{|D_k^{train}|} \sum_{(x_i,y_i)}^{\sum_{cls}ik} log \tag{3}$$

where D_k^{train} is the training set of task k , and y_i is the gold relation label of sample x_i . This stage provides the model with the basic discriminative ability for newly introduced relations and establishes the initial representation space for the current task.

3.3.2. Typical sample selection

In continual relation extraction, the model cannot preserve all historical samples due to limited memory. Therefore, it is necessary to select a small number of representative sample for each relation. The motivation is that, under a fixed memory budget, the model should retain the core semantic patterns of each relation as much as possible. If samples are selected randomly, the stored memory may overfit local expressions and fail to represent the overall semantic distribution of the relation.

To alleviate this issue, we adopt the common K-Means-based exemplar selection strategy. For each relation r , we cluster all training samples of that relation into several clusters according to the assigned memory quota, and then choose the sample nearest to each cluster center as a representative exemplar. The resulting exemplar set of relation r is denoted by M_r . All exemplar sets together form the memory buffer used in the replay stage.

3.3.3. Relation prototype initialization

Relation prototypes serve as class-level semantic centers and provide the basis for subsequent prompt construction, similar-relation identification, and replay optimization. Following the thesis design, we initialize three types of relation-level representations for each relation r : a static prototype, a dynamic prototype, and a final prototype.

The static prototype is computed as the average representation of all training samples belonging to relation r :

$$P_r^{static} = \frac{1}{|D_r|} \sum_{(x_i,y_i=r)\in D_r} h_{x_i} \tag{4}$$

where D_r denotes the training sample set of relation r . Since the static prototype is computed from all current-task samples and remains fixed afterwards, it can preserve the original core semantics of the relation and avoid interference from later feature drift.

The dynamic prototype is initialized from the representative sample of relation r :

$$P_r^{dynamic} = \frac{1}{|D_r|} \sum_{(x_i,y_i=r)\in M_r} h_{x_i} \tag{5}$$

The final prototype is then defined by a weighted fusion of the static and dynamic prototypes:

$$P_r = \alpha \cdot P_r^{static} + (1-\alpha) \cdot P_r^{dynamic} \tag{6}$$

where α is a balancing coefficient. In this way, the final prototype preserves the stable core semantics captured by the static prototype while also reflecting the memory-adaptive characteristics encoded by the dynamic prototype.

3.4. Relation-prompt dynamic selection and feature enhancement

The main motivation of this stage is that traditional replay-based methods mainly rely on memory samples or prototypes, but they are still limited in modeling subtle semantic differences among highly similar relations. Prompt learning provides a new way to inject relation-aware semantic cues into the model. Instead of using fixed prompts, we construct relation-specific soft prompts from relation prototypes and dynamically match them to input samples, so that the model can use more targeted semantic guidance during continual learning.

3.4.1. Relation-specific soft prompts and key vectors

For each relation r , we construct a relation-specific soft prompt T_r and a corresponding key vector k_r based on its static prototype P_r^{sta} . The static prototype preserves relatively stable core semantics of the relation and is therefore suitable as the basis for prompt initialization.

The relation-specific soft prompt and key vector are generated by

$$T_r = Reshape(W_t P_r^{static} + b_t) \tag{7}$$

$$k_r = W_k P_r^{static} + b_k \tag{8}$$

where W_t and W_k are learnable parameter matrices, b_t and b_k are bias terms, and $Reshape(\cdot)$ transforms the mapped vector into a soft-prompt sequence matrix. After initialization, all relation-specific soft prompts and key vectors are stored in a global prompt pool.

3.4.2. Dynamic prompt selection

For a sample x , we first generate a query vector from its semantic representation h_x by a learnable MLP: $q_x = MLP_q(h_x)$, Then, we compute the cosine similarity between q_x and each key vector k_r in the global prompt pool:

$$sim(q_x, k_r) = \frac{q_x \cdot k_r}{\|q_x\| \cdot \|k_r\|} \tag{9}$$

To simultaneously capture the core semantics of the target relation and the semantic differences from confusing relations, we adopt a Top-2 prompt selection strategy. The soft prompt corresponding to the most similar key vector is selected as the main prompt, which is used to activate the core semantics of the target relation. The soft prompt corresponding to the second-highest similarity is selected as the contrastive prompt, which provides additional contrastive cues against highly similar relations.

Compared with directly using a fixed prompt, this dynamic selection mechanism allows the model to adaptively retrieve more suitable semantic guidance for different samples, thereby improving fine-grained relation modeling.

3.4.3. Prototype and sample feature enhancement

After obtaining the matched prompts, we further enhance both prototype features and sample features.

For prototype enhancement, we fuse the relation prototype with its matched soft prompt through a feature interaction function $F(\cdot,\cdot)$:

$$\widetilde{P}_r = \mu P_r + (1-\mu) F(P_r, T_r) \tag{10}$$

where $\widetilde{P}_r$ denotes the enhanced prototype, and μ is a fusion coefficient. The feature interaction function is defined as:

$$F(h,p) = (h^\top p)p \tag{11}$$

where h denotes the input feature and p denotes the prompt feature. This operation projects the input feature onto the prompt direction and uses the prompt semantics to enhance the original representation. In this way, the prototype can better integrate both sample-level statistical semantics and prompt-guided relation semantics.

For sample enhancement, we use the main prompt and the contrastive prompt matched to sample x, denoted by t_{r1} and t_{r2}, respectively. The enhanced sample representation is computed as:

$$h_x^e = h_x + \beta_1 \cdot F(h_{x_r}, t_{r1}) - \beta_2 \cdot F(h_{x_r}, t_{r2}) \tag{12}$$

where β_1 and β_2 are weighting coefficients. In this formulation, the main prompt is used to strengthen relation-relevant semantics, while the contrastive prompt is used to suppress overlapping semantics shared with similar relations. As a result, the final enhanced sample representation becomes more discriminative for continual relation extraction.

3.5. Memory replay optimization

After completing new-task learning and prompt-based feature enhancement, we enter the memory replay stage. The replay set consists of all representative sample accumulated so far. The motivation of this stage is not only to preserve old knowledge, but also to explicitly reduce fine-grained confusion among semantically similar relations. To this end, we jointly optimize the model with four objectives, namely replay classification loss, similarity-aware prototype contrastive learning, representation anchoring constraint, and prompt matching loss.

3.5.1. Replay classification loss

Based on the enhanced sample representation h_x^e, we first compute the replay classification loss over all learned relations:

$$L_{ce}^{replay} = -\frac{1}{|\widetilde{M}_k|} \sum_{(x_i, y_i) \in \widetilde{M}_k}^{\sum_{cls\,i\,k}} log \tag{13}$$

where $\widetilde{M}_k$ denotes the replay exemplar set at task k. This loss maintains the overall classification ability of the model on both old and new relations.

3.5.2. Similarity-aware prototype contrastive learning

Conventional prototype contrastive learning treats all negative prototypes equally. However, in continual relation extraction, confusing errors often come from a small number of semantically close hard negative relations rather than from all negative classes equally. Therefore, ordinary prototype contrastive learning is insufficient for modeling fine-grained boundaries among similar relations.

To address this issue, we first compute the semantic similarity between any two relations r_a and r_b based on their static prototypes:

$$s(r_a, r_b) = sim(P_{r_a}^{static}, P_{r_b}^{static}) \tag{14}$$

Given a similarity threshold δ, if $s(r_a, r_b) > \delta$, the two relations are regarded as a similar-relation pair. Based on this, we define a relation similarity coefficient:

$$\omega(r_a, r_b) = \begin{cases} 1 + s(r_a, r_b), & r_a \neq r_b, s(r_a, r_b) > \delta \\ 1, & otherwise \end{cases} \tag{15}$$

where no extra weighting is imposed on the positive class itself. Then, the similarity-aware prototype contrastive loss is defined as

$$L_{con}^{sr} = -\frac{1}{|\widetilde{M}_k|} \sum_{(x_i, y_i) \in \widetilde{M}_k}^{\sum \frac{exp(sim(h_{x_i}^e, \widetilde{P}_{y_i}))}{\sum_{r \in \widetilde{R}_k}^{\sum(\omega(y_i, r) \cdot sim(h_{x_i}^e, \widetilde{P}_r))} exp}} log \tag{16}$$

where $\widetilde{R}_k$ denotes the set of all learned relations up to task k . This loss enlarges the competitive effect of semantically similar hard negative prototypes and forces the model to learn clearer boundaries against confusing relations.

3.5.3. Representation anchoring constraint

Although similarity-aware prototype contrastive learning [37] enhances discrimination at the prototype level, it mainly acts on relative similarity values and does not explicitly constrain the geometric distribution of samples in the representation space. To further reduce local overlap between similar relations, we introduce a representation anchoring constraint.

For any similar-relation pair (r_a, r_b) , we construct an anchor direction vector based on their static prototypes:

$$a_{r_a,r_b} = \frac{P^{static}_{r_a} - P^{static}_{r_b}}{\|P^{static}_{r_a} - P^{static}_{r_b}\|_2} \tag{17}$$

This vector can be viewed as the discriminative direction between the two relations. For a sample label y_i , we define the direction indicator function:

$$\gamma(y_i; r_a, r_b) = \begin{cases} 1, & y_i = r_a \\ -1, & y_i = r_b \end{cases} \tag{18}$$

Then, the anchoring loss is formulated as

$$L_{anc} = \frac{1}{|D_s|} \sum_{(x_i,y_i)\in D_s}^{\sum(\tau_a - \gamma(y_i;r_a,r_b)\cdot sim(h^e_{x_i}, a_{r_a,r_b}))} ReLU \tag{19}$$

where D_s is the sample set associated with similar-relation pairs, and τ_a is a safety margin. This constraint encourages samples from similar relations to stay on opposite sides of the discriminative direction, thereby explicitly improving local separability in the representation space.

3.5.4. Prompt matching loss

To ensure that each sample can retrieve the correct relation-specific prompt from the prompt pool, we further introduce a prompt matching loss. Its goal is to maximize the similarity between the sample query vector and the key vector of its ground-truth relation while minimizing its similarity to irrelevant keys. Using the InfoNCE objective, the loss is defined as

$$L_{match} = -\frac{1}{|\widetilde{M}_k|} \sum^{\sum \frac{exp(sim(q_{x_i}, k_{y_i}))}{\sum_{r\in\tilde{R}_k}^{\sum sim_{x_i r}} exp}} log \tag{20}$$

where q_i is the query vector of sample x_i , and k_{y_i} is the key vector corresponding to its gold relation label. This loss strengthens semantic alignment between samples and prompts and improves the quality of prompt retrieval during continual learning.

3.6. Overall objective and relation prediction

The overall training objective is divided according to the learning stage. In the new-task learning stage, the model is optimized by $L_{new} = L^{new}_{ce}$.

In the memory replay stage, we jointly optimize:

$$L_{replay} = L^{replay}_{ce} + \lambda_1 L_{sapc} + \lambda_2 L_{anchor} + \lambda_3 L_{match} \tag{21}$$

where λ_1 , λ_2 , and λ_3 are weighting coefficients.

During inference, we compute the cosine similarity between the enhanced representation of a test sample and the enhanced prototype of each learned relation:

$$score(x_{test}, r) = \frac{h^e_{x_{test}} \cdot \widetilde{P}_r}{\|h^e_{x_{test}}\| \cdot \|\widetilde{P}_r\|} \tag{22}$$

The final predicted relation label is given by

$$y^*_{pred} = score(x_{test}, r) \tag{23}$$

In this way, the model predicts the relation whose enhanced prototype is most similar to the enhanced representation of the test sample.

4. Results

4.1. Datasets

We evaluate the proposed method on two public relation extraction datasets, TACRED and FewRel. These two datasets are highly representative in terms of data scale, relation granularity, and task difficulty, and therefore provide a relatively comprehensive benchmark for assessing both new-knowledge acquisition and historical-knowledge retention in continual relation extraction.

TACRED [38] is a large-scale supervised relation extraction dataset constructed from news articles and web text. It contains 106,264 instances, covering 41 predefined relation types and one additional no_relation category. Compared with FewRel, TACRED is closer to real-world supervised relation extraction scenarios, with more diverse sentence expressions, more obvious noise, and more severe class imbalance. Therefore, it is well suited for evaluating the model's relation discrimination ability and continual learning stability in complex and realistic contexts.

FewRel [15] is a large-scale supervised few-shot relation classification dataset constructed from Wikipedia. It contains 100 relation types and 70,000 instances, with 700 annotated samples for each relation. This dataset is commonly divided into 64 training relations, 16 validation relations, and 20 test relations.

4.2. Experimental settings and evaluation metrics

To verify the effectiveness of our method on continual relation extraction, we conduct experiments on the two public datasets FewRel and TACRED. Following the common setting in continual relation extraction, the relation classes are sequentially divided into 10 continuously arriving subtasks, and the relation sets of different tasks are mutually exclusive. The model learns these tasks in order. At each stage, it can access only the training data of the current task and a limited number of historical memory samples, so as to simulate the incremental learning process in realistic scenarios where relation categories continuously expand.

To ensure a fair comparison, we keep a fixed number of representative sample for each relation in the memory buffer, with 10 samples retained for each relation. These sample are selected by the model according to representation quality and class representativeness, and are used for memory replay and historical knowledge retention in subsequent tasks.

As for evaluation, we use average accuracy (Acc) as the main metric. After learning the k -th task, the model is evaluated on the cumulative test set up to the current stage, in order to measure its overall recognition performance on all learned relations.

4.3. Baseline models

To comprehensively evaluate the effectiveness of our method, we select six representative methods in continual relation extraction as baselines, namely RP-CRE [2], CRECL [5], CRL [3], CEAR [6], MREM [12] and KIP-Frame [39]. These methods address catastrophic forgetting in continual relation extraction from different perspectives, including relation prototype modeling, contrastive learning, knowledge retention, knowledge-enhanced prototypes, and prompt-memory fusion, and therefore provide strong representativeness and comparability.

RP-CRE [2]: A continual relation extraction method based on relation prototype optimization. It introduces relation prototypes to refine sample embeddings, thereby improving the stability of old-relation memory during incremental learning, enhancing the utilization of memory samples, and reducing the model's dependence on large memory capacity.

CRECL [5]: A prototype contrastive learning method for continual relation extraction. It introduces a prototype-level contrastive learning mechanism to explicitly enhance the separability among different relation categories in the feature space, thereby improving the model's discrimination ability over continuously arriving relation tasks.

CRL [3]: A continual relation extraction method that combines supervised contrastive learning with memory replay. During the replay stage, it constrains the relation representation space through contrastive learning to alleviate the disruption caused by new-task learning to historical relation representations, thus improving continual learning stability.

CEAR [6]: A relation prototype enhancement method for continual relation extraction. It combines static prototypes and dynamic prototypes to construct relation-insensitive representations, thereby reducing the model's overfitting to memory samples, and further improves the discrimination of similar relations through a joint training mechanism.

MREM [12]: A continual relation extraction method based on memory-sample prompt enhancement. It treats historical memory samples as adaptive prompt information and employs gated fusion and consistency classification mechanisms to improve the retention of old relation knowledge, thereby reducing catastrophic forgetting during continual learning.

KIP-Frame [39]: A continual relation extraction framework that incorporates knowledge-enhanced prototypes. By introducing knowledge-injected prototypes and a stability–plasticity balancing mechanism, it improves the model's adaptability to new knowledge while preserving the stability of old knowledge, thereby mitigating catastrophic forgetting in continual learning.

4.4. Main results

To comprehensively evaluate the performance of our method throughout the whole continual relation extraction process, we report the average accuracy at each task stage on both FewRel and TACRED. Specifically, Table 1 presents the results on the FewRel dataset, while Table 2 reports the results on the TACRED dataset.

Table 1. Comparison of predictive performance of different models on the FewRel dataset (the best results are shown in bold)

Method	Task									
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
RP-CRE [2]	97.9	92.7	91.6	89.2	88.4	86.8	85.1	84.1	82.2	81.5
CRECL [5]	97.8	94.9	92.7	90.9	89.4	87.5	85.7	84.6	83.6	82.7
CRL [3]	98.2	94.6	92.5	90.5	89.4	87.9	86.9	85.6	84.5	83.1
KIP-Frame [39]	98.4	93.5	92.0	91.2	90.0	88.2	86.9	85.6	84.1	82.5
CEAR [6]	98.1	95.8	93.6	91.9	91.1	89.4	88.1	86.9	85.6	84.2
MREM [12]	98.7	97.5	92.3	90.7	90.7	88.6	86.9	85.7	84.9	83.7
Ours	98.2	96.4	95.5	92.7	92.0	90.1	89.3	88.6	86.5	85.3

The results on FewRel show that all methods maintain relatively high accuracy in the early task stages, indicating that when the number of previously learned relations is still small, most methods already possess a certain ability to learn new knowledge. Although our method does not achieve the best result at T1 and T2, it still reaches 98.2 and 96.4, respectively, showing only a small gap from the best-performing models and demonstrating strong initial learning capability. As the tasks continue to accumulate, our method begins to outperform all baseline models from T3 onward and consistently achieves the best results from T3 to T10. This indicates that the proposed method can better coordinate new-knowledge acquisition and old-knowledge retention as more historical relations are introduced. Further observation of the later tasks shows that the performance of all methods declines to varying degrees, suggesting that catastrophic forgetting remains a persistent challenge in continual learning. Nevertheless, our method maintains the best performance at every later stage and achieves 85.3 at the final task T10, outperforming CEAR and MREM by 1.1 and 1.6 percentage points, respectively. These results demonstrate that, as the task sequence progresses and the number of historical relations increases, our method can more effectively preserve the discriminative ability for learned relations while alleviating confusion among similar relations.

Table 2. Comparison of predictive performance of different models on the TACRED dataset (the best results are shown in bold)

Method	Task									
	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
RP-CRE [2]	97.6	90.6	86.1	82.4	79.8	77.2	75.1	73.7	72.4	72.4
CRECL [5]	96.6	93.1	89.7	87.8	85.6	84.3	83.6	81.4	79.3	78.5
CRL [3]	97.7	93.2	89.8	84.7	84.1	81.3	80.2	79.1	79.0	78.0
KIP-Frame [39]	98.3	95.1	90.8	87.5	85.3	84.3	82.1	80.2	79.6	78.6
CEAR [6]	97.7	94.3	92.3	88.4	86.6	84.5	82.2	81.1	80.1	79.1
MREM [12]	97.5	95.4	93.5	90.5	89.3	85.7	83.0	81.8	80.4	79.5
Ours	98.9	95.1	94.6	91.7	88.9	87.2	85.1	84.0	81.9	81.0

The results on TACRED likewise demonstrate the strong competitiveness of our method. In the early stages T1–T5, our method achieves the best results at T1, T3, and T4, with accuracies of 98.9, 94.6, and 91.7, respectively. Although it does not rank first at T2 and T5, the gaps from the best results are only 0.3 and 0.4 percentage points, respectively. This indicates that our method not only adapts quickly to new tasks but also maintains relatively stable performance as the task sequence gradually expands. A further examination of the later tasks T6–T10 shows that our method consistently outperforms all baseline models at every stage. At the final task T10, it reaches 81.0, surpassing CEAR and MREM by 1.9 and 1.5 percentage points, respectively. Since TACRED contains more complex sentence patterns and a more imbalanced class distribution than FewRel, these results further suggest that our method retains strong generalization ability and historical-knowledge retention even in more challenging continual relation extraction scenarios.

Taken together, the results in Table 1 and Table 2 show that our method already exhibits strong competitiveness in the early task stages on both datasets, and its advantage becomes more pronounced as the tasks continue to accumulate, especially in the later stages. This suggests that the proposed model improves continual learning ability in a collaborative manner from multiple aspects, thereby achieving a better balance among new-knowledge acquisition, historical-knowledge retention, and similar-relation discrimination.

4.5. Ablation study and analysis

4.5.1. Results of module ablation

To verify the effectiveness of the core modules, we conduct ablation studies on both TACRED and FewRel, with the results reported in Tables 3 and 4, respectively. The ablation settings include: removing dynamic prompt selection, which disables semantic-based prompt matching for each input sample (w/o DPS); removing prototype and sample feature enhancement, which disables joint enhancement of prototype and sample representations using prompt information (w/o FE); replacing the proposed similarity-aware prototype contrastive learning with ordinary prototype contrastive learning (w/o SRPCL); removing the similar-relation representation anchoring constraint (w/o SRA); and removing the prompt matching loss, which no longer explicitly aligns prompt representations with relation semantics (w/o PML). Overall, the full model achieves the best performance on the later tasks T_6 to T_{10} on both datasets, indicating that all modules contribute to performance improvement.

Table 3. Ablation results of different modules on the TACRED dataset (the best results are shown in bold)

TACRED	Task				
	T_6	T_7	T_8	T_9	T_{10}
Intact Model	87.2	85.1	84.0	81.9	81.0
w/o DPS	87.0	85.1	83.7	81.5	80.7
w/o FE	86.3	83.9	82.8	81.0	79.9
w/o SRPCL	86.1	84.5	83.3	81.0	80.3
w/o SRA	85.9	83.5	82.0	80.6	79.5
w/o PML	86.7	84.9	83.5	81.5	80.7

On TACRED, removing any module leads to performance degradation, with the largest drops caused by removing the representation anchoring constraint and the feature enhancement module. At T_{10} , the full model reaches 81.0%, while removing the anchoring loss and feature enhancement reduces the result to 79.5% and 79.9%, respectively. This shows that in more challenging settings with complex relation semantics, representation anchoring and feature enhancement are crucial for preserving fine-grained decision boundaries and reducing confusion among similar relations. In contrast, removing dynamic prompt selection or prompt matching causes relatively smaller drops, but still leads to consistent degradation, indicating that both modules provide stable auxiliary gains by improving prompt–sample and prompt–relation alignment.

Table 4. Ablation results of different modules on the FewRel dataset (the best results are shown in bold)

FewRel	Task				
	T_6	T_7	T_8	T_9	T_{10}
Intact Model	90.1	89.3	88.6	86.5	85.3
w/o DPS	89.6	89.1	88.1	86.2	84.9
w/o FE	88.2	87.1	86.7	84.9	84.2
w/o SRPCL	88.9	88.1	87.7	85.9	84.5
w/o SRA	88.2	87.3	87.2	85.0	84.1
w/o PML	89.9	88.9	88.3	85.8	84.7

On FewRel, removing each module also lowers performance, although the fluctuations are slightly smaller than those on TACRED, suggesting that the model is inherently more stable on datasets with relatively regular relation distributions. Among all modules, feature enhancement and similar-relation representation anchoring remain the most influential. At T_{10} , the full model achieves 85.3%, while removing feature enhancement and replacing similarity-aware contrastive learning reduce the result to 84.2% and 84.1%, respectively. This indicates that even on a relatively standard dataset such as FewRel, enhancing prototype and sample representations with prompt information and imposing additional constraints on similar-relation distributions still improve knowledge retention in later tasks.

Replacing similarity-aware contrastive learning with ordinary prototype contrastive learning also causes performance drops on both datasets. This suggests that standard prototype contrastive learning is insufficient to capture fine-grained differences among similar relations, whereas the similarity coefficient helps the model focus on semantically close hard negatives and learn clearer decision boundaries. Although the effects of dynamic prompt selection and prompt matching loss are smaller, they still provide consistent improvements within the overall framework.

In summary, the gains of our method do not come from a single component, but from the collaboration of multiple modules. Among them, feature enhancement and similar-relation representation anchoring contribute the most. This confirms that the main advantage of our method lies in improving prototype and sample representations with prompt-based enhancement while strengthening fine-grained relation boundaries through similar-relation-aware optimization.

4.5.2. Visualization of feature distributions after ablating key modules

To further examine the effect of key modules on representation quality, we visualize feature distributions on the FewRel dataset using t-SNE at three representative task stages, namely T_2 , T_6 , and T_{10}. Specifically, Figure 4(a), 4(c), and 4(e) show the feature distributions of the full model at T_2 , T_6 , and T_{10}, respectively, while Figure 4(b), 4(d), and 4(f) present the corresponding results after removing key modules. The ablated modules mainly include prototype and sample Feature Enhancement (FE), Similarity-aware Prototype Contrastive Learning (SRPCL), and the Similar-relation Representation Anchoring constraint (SRA).

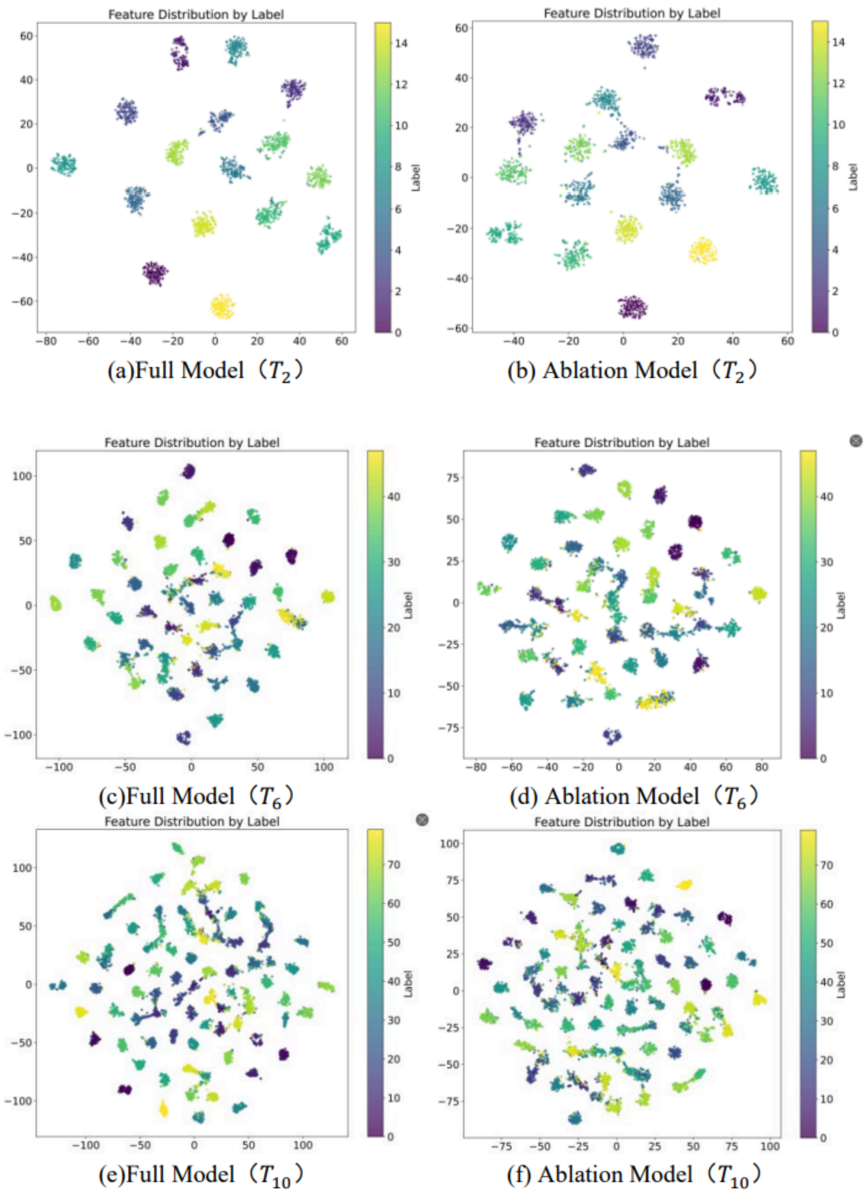


Figure 4. t-SNE visualization of feature distributions on the FewRel dataset for the full model and the ablated model at T_2, T_6, and T_{10}

At the early stage T_2, the full model already forms compact clusters with relatively clear boundaries, and different relations are well separated. After removing the key modules, most classes remain basically distinguishable, but the distances between some clusters become smaller and local boundaries become less clear. This indicates that these modules can already improve feature-space stability and discriminability at the early stage.

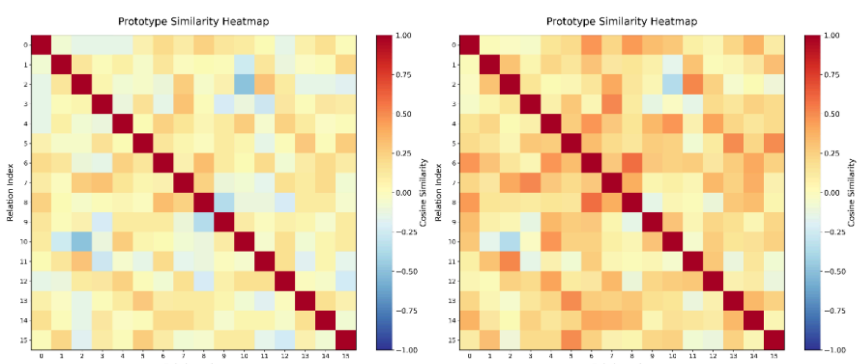
As the task progresses to T_6, the number of relation classes increases and the feature space becomes more complex. As shown in Figs. 4(c) and 4(d), the full model still maintains good intra-class compactness and inter-class separability, whereas the ablated model shows more obvious overlap, especially in central regions and among neighboring classes. This suggests that FE, SRPCL, and SRA play an important role in preserving feature-space structure and reducing representation entanglement among similar relations.

The difference becomes even more pronounced at the final stage T_{10} . As shown in Figs. 4(e) and 4(f), although the full model is exposed to more relation classes and a more challenging continual learning setting, it still preserves relatively clear class boundaries and good cluster separability. In contrast, the ablated model exhibits obvious overlap and cluster stretching in multiple regions, and several previously separable classes become more mixed. This indicates that without feature enhancement and similar-relation-aware optimization, the feature space is more likely to degrade as tasks accumulate, which weakens fine-grained relation discrimination.

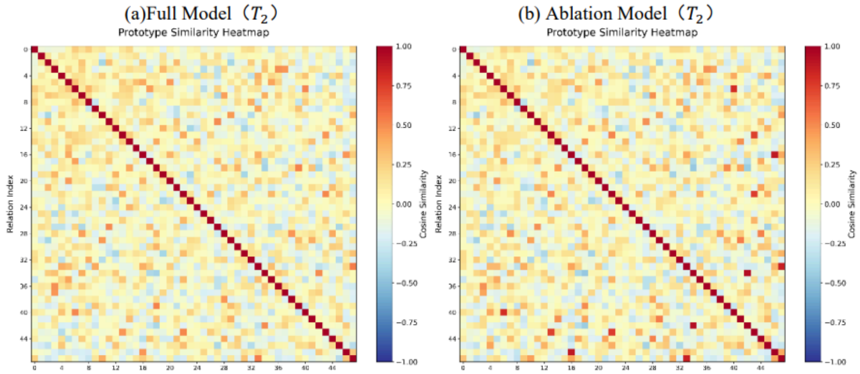
Overall, the visualization results are consistent with the ablation results. Feature enhancement improves intra-class compactness, while SRPCL and SRA further strengthen inter-class separability, especially for semantically similar relations. Their joint effect enables the proposed method to learn clearer, more stable, and more discriminative feature representations during continual relation extraction, thereby alleviating catastrophic forgetting and similar-relation confusion.

4.5.3. Analysis of relation prototype heatmaps after ablating key modules

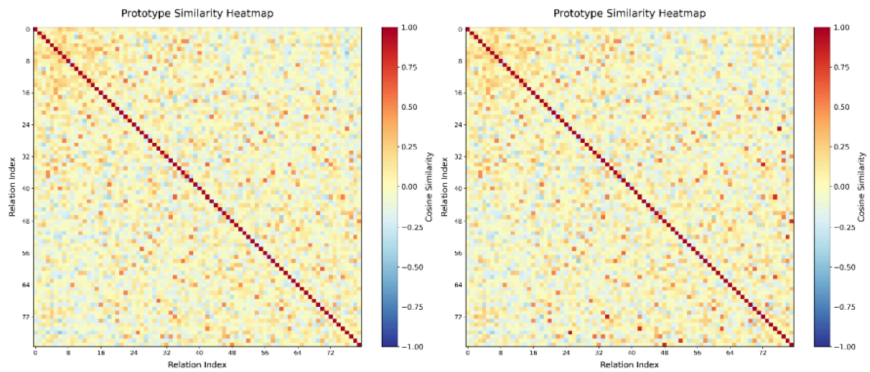
To further examine the effect of key modules on the discriminability of relation prototypes, we visualize the similarity of relation prototypes on the FewRel dataset at different task stages, as shown in Figure 5. Specifically, Figure 5(a), 5(c), and 5(e) show the prototype heatmaps of the full model at T_2 , T_6 , and T_{10}, respectively, while Figure 5(b), 5(d), and 5(f) present the corresponding results after ablating key modules. The ablated modules mainly include prototype and sample Feature Enhancement (FE), Similarity-aware Prototype Contrastive Learning (SRPCL), and the Similar-relation Representation Anchoring Constraint (SRA). Each element in the heatmap represents the cosine similarity between two relation prototypes, ranging from $[-1, 1]$. The diagonal values are always 1, while smaller or negative off-diagonal values indicate better separability between different relations.



(a)Full Model（T_2） (b) Ablation Model（T_2）



(c)Full Model（T_6） (d) Ablation Model（T_6）



(e)Full Model (T_{10}) (f) Ablation Model (T_{10})

Figure 5. Heatmaps of relation prototype similarities on the FewRel dataset at T_2, T_6, and T_{10}

At the early stage T_2, the full model already shows relatively low similarity among most relation prototypes except for the diagonal, indicating that it can learn a clear prototype structure at an early stage. In contrast, after removing the key modules, more warm-colored regions appear in the off-diagonal area, suggesting increased similarity between some relation prototypes. This indicates that without FE, SRPCL, and SRA, relation prototypes are more likely to become closer to each other even in the early stage, which weakens inter-class separability.

The difference becomes more evident at T_6. As shown in Figure 5(c) and 5(d), although more relation classes are introduced, the full model still maintains an overall low-similarity distribution, with only limited local regions showing slightly increased similarity. In comparison, the ablated model exhibits more warm-colored areas and higher positive similarities between some relation pairs, indicating blurred prototype boundaries. This suggests that without the joint effect of the key modules, different relation prototypes are more likely to cluster together, especially for semantically similar relations.

At the final stage T_{10}, this contrast becomes the most pronounced. Figure 5(e) shows that the full model still keeps most relation prototypes within a relatively low similarity range, indicating that our method can maintain a stable prototype structure even as more tasks accumulate. In contrast, Figure 5(f) contains more scattered high-similarity hotspots in the off-diagonal area, showing that prototype overlap becomes more widespread after ablation. In other words, without feature enhancement, similarity-aware prototype contrastive learning, and representation anchoring, the model becomes less capable of preserving clear prototype boundaries as the continual learning process progresses.

Overall, the prototype heatmaps are consistent with both the ablation results and the feature distribution visualizations. The full model maintains lower inter-prototype similarity and forms a clearer prototype-space structure during continual learning. After removing FE, SRPCL, and SRA, more highly similar relation prototypes appear in the heatmaps, indicating that these modules play an important role in suppressing prototype overlap and improving inter-class separability. This effect is especially evident in later tasks, further confirming that our method improves prototype discriminability and provides a more reliable class-level representation basis for stable prediction in continual relation extraction.

5. Conclusion

In this paper, we proposed a continual relation extraction method that integrates prompt learning, feature enhancement, anchoring loss, and similarity-aware prototype contrastive learning. By introducing relation-specific soft prompts and dynamic prompt selection, the proposed method enhances both prototype and sample representations and further improves the discrimination of semantically similar relations during continual learning.

Experimental results on FewRel and TACRED show that the proposed method consistently achieves strong performance, especially in later task stages. The ablation and visualization results further confirm that prompt-guided feature enhancement and similar-relation-aware optimization are effective in alleviating catastrophic forgetting and reducing confusion among similar relations.

These results suggest that continual relation extraction should not only preserve historical knowledge, but also explicitly optimize fine-grained semantic boundaries in the representation space. This provides useful insights for future research on continual information extraction and has practical value for applications such as knowledge graph construction and dynamic text understanding.

References

[1] Han X, Dai Y, Gao T, et al. Continual relation learning via episodic memory activation and reconsolidation [C]//Proceedings of the 58th annual meeting of the association for computational linguistics. 2020: 6429-6440.

[2] Cui L, Yang D, Yu J, et al. Refining sample embeddings with relation prototypes to enhance continual relation extraction [C]//Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2021: 232-243.

[3] Zhao K, Xu H, Yang J, et al. Consistent representation learning for continual relation extraction [C]//Findings of the association for computational linguistics: ACL 2022. 2022: 3402-3411.

[4] Wang P, Song Y, Liu T, et al. Learning robust representations for continual relation extraction via adversarial class augmentation [C]//Proceedings of the 2022 conference on empirical methods in natural language processing. 2022: 6264-6278.

[5] Hu C, Yang D, Jin H, et al. Improving continual relation extraction through prototypical contrastive learning [C]//Proceedings of the 29th international conference on computational linguistics. 2022: 1885-1895.

[6] Zhao W, Cui Y, Hu W. Improving continual relation extraction by distinguishing analogous semantics [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023: 1162-1175.

[7] Xia H, Wang P, Liu T, et al. Enhancing continual relation extraction via classifier decomposition [C]//Findings of the association for computational linguistics: ACL 2023. 2023: 10053-10062.

[8] Nguyen H, Nguyen C, Ngo L, et al. A spectral viewpoint on continual relation extraction [C]//Findings of the Association for Computational Linguistics: EMNLP 2023. 2023: 9621-9629.

[9] Zhou S, Li Y, Miao X, et al. An ensemble-of-experts framework for rehearsal-free continual relation extraction [C]//Findings of the Association for Computational Linguistics: ACL 2024. 2024: 1410-1423.

[10] Huang M, Xiao M, Wang L, et al. DP-CRE: Continual relation extraction via decoupled contrastive learning and memory structure preservation [C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 5338-5349.

[11] Chen J, Wang S, Ma L, et al. SS-CRE: A Continual Relation Extraction Method Through SimCSE-BERT and Static Relation Prototypes [J]. *Neural Processing Letters, 2024, 56*(4): 203.

[12] Zhang Y, Peng H, Zhang Y, et al. Retrieving memory as prompts for continual relation extraction [J]. *Expert Systems with Applications, 2024, 255*: 124542.

[13] Le M, Luu T N, Le T T, et al. Adaptive prompting for continual relation extraction: a within-task variance perspective [C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2025, 39(23): 24384-24392.

[14] Han X, Zhu H, Yu P, et al. FewRel: A large-scale supervised few-shot relation classification dataset with state-of-the-art evaluation [C]//Proceedings of the 2018 conference on empirical methods in natural language

processing. 2018: 4803-4809.

[15] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding [C]//Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers). 2019: 4171-4186.

[16] Lester B, Al-Rfou R, Constant N. The power of scale for parameter-efficient prompt tuning [C]//Proceedings of the 2021 conference on empirical methods in natural language processing. 2021: 3045-3059.

[17] Liu P, Yuan W, Fu J, et al. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing [J]. *ACM computing surveys, 2023, 55*(9): 1-35.

[18] Liu X, Ji K, Fu Y, et al. P-tuning: Prompt tuning can be comparable to fine-tuning across scales and tasks [C]//Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). 2022: 61-68.

[19] SHEN Yining, WANG Yiran, WU Cong. Research Progress of Relation Extraction Based on Deep Learning [J]. *Electronic Science and Technology, 2025, 38*(7): 40-49.

[20] Gou J, Yu B, Maybank S J, et al. Knowledge distillation: A survey [J]. *International journal of computer vision, 2021, 129*(6): 1789-1819.

[21] FENG Jian, WU Peng. A Random Mask and Adaptive Feature Distillation Algorithm [J]. *Electronic Science and Technology, 2025, 38*(10): 1-9.

[22] Wang X, Wang Z, Hu W. Serial contrastive knowledge distillation for continual few-shot relation extraction [C]//Findings of the Association for Computational Linguistics: ACL 2023. 2023: 12693-12706.

[23] Zhang Z, Zeng D, Bai X. Improving continual few-shot relation extraction through relational knowledge distillation and prototype augmentation [C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 8756-8767.

[24] Hu J, Tan C, Xu J C, et al. Continual few-shot relation extraction via adaptive gradient correction and knowledge decomposition [C]//Findings of the Association for Computational Linguistics: ACL 2024. 2024: 11805-11816.

[25] Zhao S, Duan J, He L, et al. Continuous Adaptive Knowledge Distillation for Few-Shot Relation Extraction [J]. *ACM Transactions on Asian and Low-Resource Language Information Processing, 2025, 24*(10): 1-13.

[26] Chen X, Wu H, Shi X. Consistent prototype learning for few-shot continual relation extraction [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023: 7409-7422.

[27] Zhao S M, Tan Z, Pang N, et al. Dynamic-prototype contrastive fine-tuning for continual few-shot relation extraction with unseen relation detection [C]//Proceedings of the 31st International Conference on Computational Linguistics. 2025: 8763-8773.

[28] Zhou W, Zhang S, Naumann T, et al. Continual contrastive finetuning improves low-resource relation extraction [C]//Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2023: 13249-13263.

[29] Pang N, Zhao X, Zeng W, et al. SCL: Selective Contrastive Learning for Data-driven Zero-shot Relation Extraction [J]. *Transactions of the Association for Computational Linguistics, 2024*, 12: 1720-1735.

[30] Meng S, Hu X, Liu A, et al. RAPL: A relation-aware prototype learning approach for few-shot document-level relation extraction [C]//Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. 2023: 5208-5226.

[31] Ma S, Han J, Liang Y, et al. Making pre-trained language models better continual few-shot relation extractors [C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 10970-10983.

[32] Bai G, Lu C, Guo D, et al. Clear up confusion: Advancing cross-domain few-shot relation extraction through relation-aware prompt learning [C]//Proceedings of the 2024 Conference of the North American Chapter of the

Association for Computational Linguistics: Human Language Technologies (Volume 2: Short Papers). 2024: 70-78.

[33] Zhang Z, Yang Y, Chen B. Prompt tuning for few-shot relation extraction via modeling global and local graphs [C]//Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024). 2024: 13233-13242.

[34] Anh N H, Tran Q, Nguyen T X, et al. Mutual-pairing data augmentation for fewshot continual relation extraction [C]//Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). 2025: 4057-4075.

[35] Le A D, Le Hai N, Nguyen T X, et al. Enhancing discriminative representation in similar relation clusters for few-shot continual relation extraction [C]//Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers). 2025: 2450-2467.

[36] Gao T, Yao X, Chen D. Simcse: Simple contrastive learning of sentence embeddings [C]//Proceedings of the 2021 conference on empirical methods in natural language processing. 2021: 6894-6910.

[37] Hu H, Wang X, Zhang Y, et al. A comprehensive survey on contrastive learning [J]. Neurocomputing, 2024, 610: 128645.

[38] Zhang Y, Zhong V, Chen D, et al. Position-aware attention and supervised data improve slot filling [C]//Proceedings of the 2017 conference on empirical methods in natural language processing. 2017: 35-45.

[39] Zhang H, Liang B, Yang M, et al. Prompt-based prototypical framework for continual relation extraction [J]. *IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2022, 30*: 2801-2813.