IFS/ACM Transactions on Machine Learning, vol. 2 , no. 1, pp. 16-29, 2025

A Span-Constrained Decoding Method for Relation Extraction

Rui Chen[a], Hang Li[a,*], Chu Zhao[a], Shoulin Yin[a] and Asif Ali Laghari[b]

[a]College of Artificial Intelligence, Shenyang Normal University, Shenyang 110034 Liaoning, China
[b]Department of Computer Science, Sindh Madressatul Islam University, Karachi 74200 Sindh, Pakistan

ARTICLE INFO

Keywords:
entity recognition
relation extraction
knowledge modeling
semantic understanding
CasRel framework

ABSTRACT

Driven by the accelerating digitalization of education, artificial intelligence technologies have been progressively embedded into critical application scenarios, including experimental pedagogy and the development of educational resources. The automated extraction of entities and relations from unstructured experimental texts constitutes a foundational prerequisite for knowledge modeling and semantic comprehension, with its performance directly governing the usability and robustness of downstream systems. To tackle the challenges confronting existing relation extraction methods in this domain—specifically, the difficulty of precisely identifying entity boundaries and the compounding of errors stemming from erroneous candidate propagation—this paper proposes SCOPE-CasRel, a span-constrained decoding approach for relation extraction specifically tailored to secondary school experimental texts. Grounded in the CasRel framework, the proposed method introduces a span-constrained pairing mechanism that jointly models the start and end boundaries of entities, complemented by a length-constrained strategy to refine entity span selection, thereby enhancing boundary consistency. Additionally, a candidate filtering strategy is devised for the decoding phase, wherein low-confidence candidates are pruned via a dynamic threshold coupled with a Top-K mechanism, effectively mitigating the accumulation of false positives. Empirical validation is performed on the publicly available NYT and WebNLG datasets. The experimental results reveal that the proposed method surpasses several mainstream baselines across precision, recall, and F1-score, achieving substantial gains in extraction accuracy and overall performance while preserving strong recall capability.

1. Introduction

Driven by the accelerating digitalization of education, artificial intelligence (AI) technologies–such as classfication, clustering, and natural language processing–have been progressively embedded into a broadening spectrum of educational resource development and knowledge service applications [1, 2, 3]. As a pivotal research area within the field of information extraction for natural language processing, relation extraction seeks to automatically identify entities and their semantic relationships from unstructured text [4]. It serves as a foundational technology for constructing structured knowledge representations and enabling semantic understanding [5]. In the educational domain, particularly in texts pertaining to experimental pedagogy, relation extraction plays a pivotal role in the organization and representation of experimental knowledge.

By automatically extracting entities and their relationships among elements such as experimental apparatus, operational procedures, experimental phenomena, and experimental conclusions from textual sources, fragmented knowledge embedded within unstructured content can be transformed into structured representations, thereby furnishing foundational support for the construction of experimental knowledge graphs. Building upon this, relation extraction can further facilitate intelligent retrieval of experimental teaching resources, semantic comprehension of experimental processes, and the development of educational support systems. For instance, in experimental pedagogy scenarios, knowledge structures derived from extracted relations can assist students in more intuitively grasping the causal dependencies between experimental procedures and outcomes. Moreover, such structures furnish essential data support for downstream applications, including intelligent question answering and experiment recommendation systems. Recent studies have demonstrated that converting educational materials into knowledge graphs can enhance the

DOI: https://doi.org/10.70891/TML.2025.060059
ISSN of IFS/ACM TML: 3078-5510

*Corresponding author
lihangsoft@163.com (H. Li)

structured organization of digital educational content and provide a basis for intelligent retrieval, question answering, and personalized learning support [6, 7].

In recent years, relation extraction methods have gradually evolved from traditional approaches reliant on manually crafted features to joint extraction models grounded in deep learning. In contrast to pipeline-based methods that perform entity recognition and relation classification as separate stages, joint extraction models are capable of modeling the interactions between entities and relations within a unified framework, thereby mitigating the problem of error propagation to a certain extent. Cascade-based relation extraction methods, exemplified by CasRel, accomplish triple extraction through subject identification followed by relation-specific object prediction, achieving promising performance in scenarios involving overlapping relations.

Although joint extraction approaches have significantly advanced relation modeling capabilities, existing cascade-based models still exhibit several limitations when handling complex textual scenarios. First, the start and end positions of subjects and objects are typically predicted independently, lacking explicit constraints on the overall span of an entity. Consequently, boundary mismatch problems frequently arise. As shown in Figure 1, in experimental texts, entities may suffer from boundary truncation or boundary extension, resulting in incomplete semantic representations and subsequently compromising relation prediction accuracy.

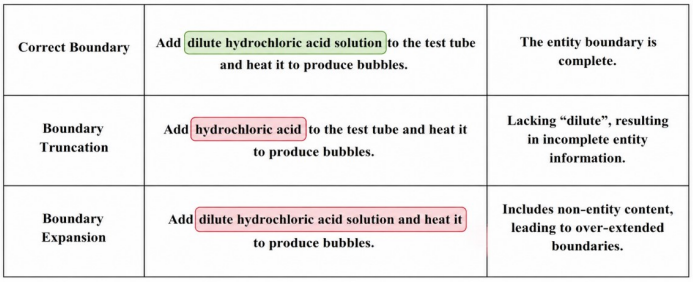

Figure 1: Illustration of entity boundary mismatch problem.

Furthermore, the cascade decoding process lacks an effective candidate constraint mechanism. Erroneous candidates generated in earlier stages may propagate into subsequent relation prediction, leading to the compounding of false-positive results. As shown in Figure 2, an incorrectly identified subject may further generate multiple spurious triples during subsequent decoding stages, thereby amplifying prediction errors. It can be observed that entity boundary mismatches primarily undermine the accuracy of entity recognition, whereas the propagation of erroneous candidates further degrades the overall performance of relation extraction.

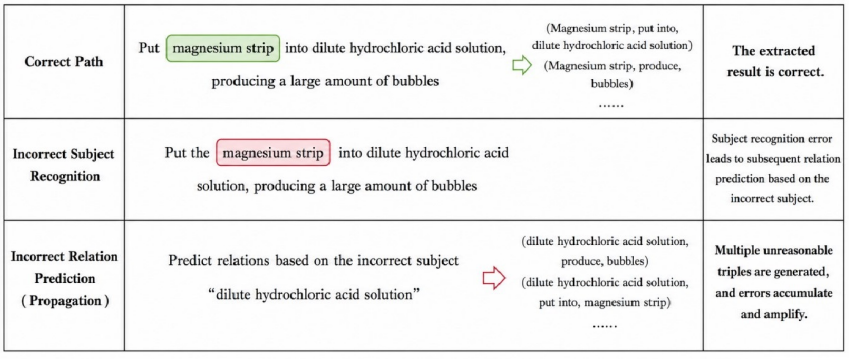

Figure 2: Illustration of the error candidate propagation and amplification problem.

To address the aforementioned issues, this paper investigates the task of relation extraction and proposes a span-constrained decoding-based relation extraction method, which improves upon the traditional CasRel model. The main contributions are summarized as follows:

- To address the insufficient entity boundary modeling in cascade-based relation extraction models, a span-constrained decoding mechanism is proposed. By jointly modeling the start and end boundaries of entities and introducing a length constraint, the accuracy of entity boundary identification is enhanced.
- To mitigate the problem of error accumulation during the decoding process, a candidate filtering strategy is designed. By employing a dynamic threshold and a Top-K mechanism to filter candidate results, the propagation of incorrect predictions is effectively suppressed.
- Empirical validation is conducted on the publicly available NYT and WebNLG datasets, with comparative studies performed to verify the effectiveness of the proposed method.

The remainder of this paper is organized as follows. Section 1 presents the introduction, encompassing the research background and significance, an analysis of the principal challenges in relation extraction, and the research content and contributions of this work. Section 2 reviews related work, systematically summarizing the research progress in relation extraction from three perspectives: pipeline-based methods, joint extraction methods, and domain-specific relation extraction approaches. Section 3 elaborates the model architecture design in detail, including the proposed SCOPE-CasRel model, which comprises a contextual encoding module, a subject labeling module, a span-constrained pairing module, an object and relation prediction module, and a candidate filtering module, along with the detailed implementation process. Section 4 presents the experiments and result analysis, including the experimental settings, datasets, and evaluation metrics. It further validates the effectiveness of the proposed method through comparative experiments and ablation studies, followed by a detailed discussion of the results. Section 5 concludes the work and outlines directions for future research.

2. Related Work

Entity-relation extraction aims to automatically identify entities and their semantic relationships from unstructured text. It serves as a fundamental basis for tasks such as knowledge graph construction, semantic retrieval, and intelligent question answering. According to different modeling paradigms, existing entity-relation extraction methods can be broadly divided into two categories: pipeline-based methods and joint extraction methods. Pipeline-based methods typically first perform entity recognition and then conduct relation classification on the identified entity pairs. In contrast, joint extraction methods simultaneously model entity and relation information within a unified framework.

In recent years, with the advancement of pre-trained language models and structured decoding techniques, relation extraction research has gradually transitioned from early stage-wise processing to joint modeling of entities and relations. Existing studies have demonstrated that, under consistent span representation settings, well-designed joint models generally outperform pipeline-based counterparts. However, joint modeling also introduces challenges in entity-relation interaction and decoding complexity [8].

Pipeline-based methods decompose the relation extraction task into two subtasks: named entity recognition and relation classification. These methods are characterized by a clear structure and relatively straightforward implementation. Early approaches often relied on manually engineered features, syntactic structures, or convolutional neural networks to model the contextual information between entity pairs. For instance, Zeng et al. [9] proposed a distant supervision-based relation extraction method using piecewise convolutional neural networks, which enhances the representation of local semantic features between entities through piecewise pooling. Miwa and Bansal [10] further employed LSTMs over both sequential word representations and dependency tree structures to jointly model entity and relation information, demonstrating that both word-order information and syntactic structures play important roles in relation extraction. These studies advanced the development of neural network-based relation extraction methods; however, under the typical pipeline framework, entity recognition and relation classification remain two relatively independent stages. Once errors such as boundary truncation, boundary expansion, or missing entities occur in the entity recognition stage, they are usually difficult to rectify in the subsequent relation classification stage. In recent years, some studies have attempted to improve entity-relation extraction from a pipeline perspective. Chen et al. [11] proposed a pattern-first pipeline method and pointed out that stage-wise extraction strategies that prioritize either entities or relations tend to lead to entity redundancy and candidate combination problems. Tang et al. [12] proposed a boundary regression model, which reduces the impact of inaccurate entity spans on relation extraction results through boundary filtering and boundary regression. These studies indicate that although pipeline-based methods are straightforward to implement and facilitate modular optimization, their performance is often constrained by the quality of entity boundaries and the scale of candidate entities.

Joint extraction methods model entity recognition and relation prediction within a unified framework, enabling more effective utilization of the interactions between entities and relations. Zheng et al. [13] proposed an early joint entity-relation extraction method based on a unified labeling scheme, which encodes both entity and relation information into a single labeling framework, thereby promoting the development of joint extraction methods. CasRel [14] models relations as a mapping function from subject to object and adopts a cascade binary tagging scheme to extract relational triples. It demonstrates strong performance in handling overlapping triples and has therefore been widely used as a baseline model for subsequent improvements. Building upon this foundation, researchers have further explored approaches such as single-stage modeling, table filling, span pruning, and structured decoding. TPLinker [15] formulates joint extraction as a token-pair linking problem and introduces a handshaking tagging scheme to jointly model entity boundaries and relations, alleviating error accumulation caused by inconsistencies between the training and inference stages. OneRel [16] formulates joint extraction as a fine-grained triple classification task, emphasizing the holistic dependency among the subject, relation, and object within a triple, thereby reducing redundant information and error propagation introduced by cascade steps. Yan et al. [17] proposed a joint extraction model combining span pruning and hypergraph neural networks, which improves extraction performance in complex relation scenarios through a high-recall pruning mechanism and higher-order interaction modeling. Li et al. [18] proposed a joint extraction model based on a pointer-based decomposition strategy, which decomposes the task into head entity recognition and tail entity–relation extraction, and strengthens the connection between entity boundaries and relation types through boundary-aware modeling. These studies indicate that joint extraction methods have evolved from simple joint prediction of entities and relations to fine-grained modeling approaches that integrate boundary modeling, span selection, and structured decoding.

Entity-relation extraction in domain-specific texts has also garnered increasing attention. Compared with general news texts, domain-specific texts typically contain a substantial number of specialized terms, where entity boundaries rely more heavily on contextual semantics, and relation types are constrained by specific application scenarios. In the medical domain, Lu et al. [19] proposed a feature-enhanced cascade binary tagging framework for Chinese electronic medical records to address the complexity of medical entity relations and the specialized nature of semantic expressions in clinical texts. Li et al. [20] focused on the joint extraction of Chinese medical entities and relations, highlighting challenges such as complex entity types, overlapping relations, and difficulties in boundary recognition, and adopted pointer networks and CasRel-inspired approaches for modeling. In the railway domain, Chen et al. [21] proposed a joint entity-relation extraction model integrating Global Pointer and tensor learning for track circuit texts, supporting railway maintenance knowledge graph construction and on-site maintenance decision-making, and demonstrating that entities and relations in technical domain texts often exhibit strong structural dependencies. In the legal domain, related studies [22, 23] have also attempted to extract entities and relations from legal texts to construct legal knowledge graphs, supporting tasks such as case analysis and similar case retrieval. This indicates that domain-specific relation extraction plays a crucial role in knowledge organization and intelligent applications.

3. Model Architecture Design

The relation extraction task aims to extract triples comprising entities and their semantic relations from a given text sequence. Given an input text sequence $X = \{x_1, x_2, \ldots, x_n\}$, the objective is to predict a set of triples: $\mathcal{T} = \{(s, r, o)\}$, where s denotes the subject entity, o denotes the object entity, and r represents the semantic relation between them.

Based on the principle described in Equation (1), this paper proposes an improved joint extraction model SCOPE-CasRel.

$$\prod_{(s,r,o)} p((s,r,o)|x) = \prod_{s\in T} p(s|x) \prod_{(r|o)} p((r,o)|x,s) = \prod_{s\in T} p(s|x) \prod_{r\in T|s} p(o|x,s,r) \prod_{r\in R\setminus T|s} p(o_\phi|x,s,r) \tag{1}$$

In Equation (1), $s \in T$ denotes the subject entity in overlapping relations; $T|s$ represents the set of triples in T that take s as the subject. $(r, o) \in T|s$ denotes the relation-object pair within $T|s$. r represents the set of all relation types in the training set. $R \in T|s$ denotes all relations excluding those associated with the subject s in T. o_ϕ is the "null" object, indicating that, except for the relations included in $T|s$, all other relations have no corresponding object in the sentence s.

The model comprises five modules: a contextual encoding module, a subject labeling module, a span-constrained pairing module, an object and relation prediction module, and a candidate filtering module. The overall workflow proceeds as follows. First, a pre-trained language model is employed to obtain contextual representations of the

input text; subsequently, the subject labeling module identifies potential subject entities. Building upon this, a span-constrained pairing mechanism is introduced to optimize entity boundary modeling. Next, the object and relation prediction module jointly predicts object entities and their associated relations conditioned on the identified subject. Finally, a candidate filtering strategy is applied to refine the outputs and yield the final set of extracted triples. The overall architecture of the model is shown in Figure 3.

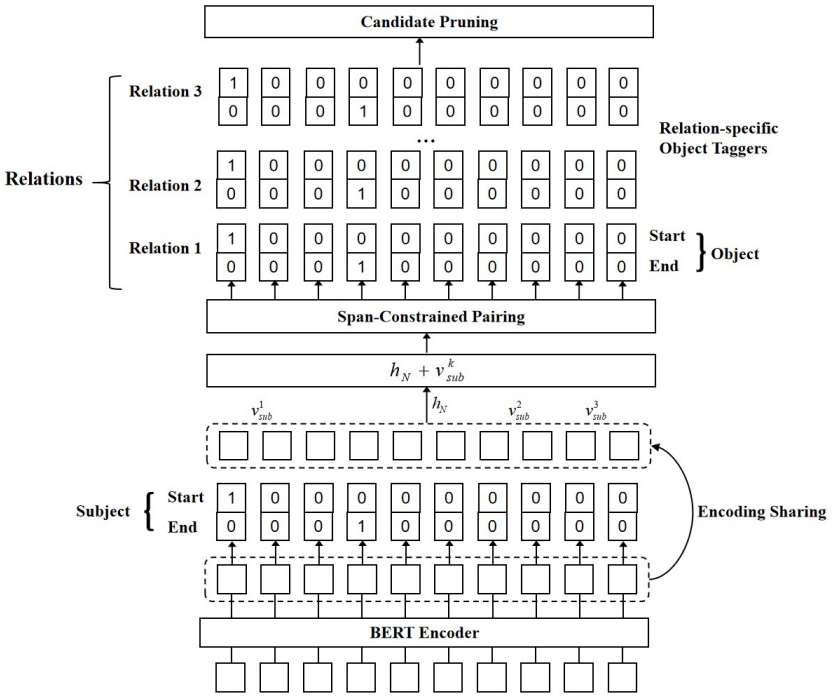

Figure 3: Overall architecture of the SCOPE-CasRel model.

3.1. Contextual Encoding Module

In the relation triple extraction task, the contextual semantic representation of text directly governs the performance of entity boundary recognition and relation modeling. Therefore, constructing high-quality contextual representations constitutes a foundational component of the entire model. In this work, the pre-trained language model BERT is adopted as the contextual encoder to perform deep semantic modeling of the input text, thereby obtaining representations imbued with global semantic information.

As shown in Figure 3, given an input text sequence:

$$X = \{x_1, x_2, \ldots, x_n\} \tag{2}$$

First, each token in the sequence is mapped into a continuous vector representation through word embedding and positional encoding, transforming discrete textual symbols into dense vectors. Specifically, x_i denotes the i-th token, and its corresponding input representation is denoted as e_i, which is obtained by summing the word embedding and positional embedding. In this manner, both the semantic information of the word and its positional information within the sequence are encoded simultaneously. Accordingly, the input embedding sequence can be formulated as:

$$E = \{e_1, e_2, \ldots, e_n\} \tag{3}$$

Next, the embedding representations are fed into the BERT encoder, where a multi-layer bidirectional Transformer is employed to model the sequence, capturing contextual dependencies between words as well as long-range semantic

information. Upon encoding, the model produces contextual semantic representations for each position in the sequence as follows:

$$H = \text{BERT}(X) = \{h_1, h_2, \ldots, h_n\}, h_i \in R^d \tag{4}$$

where h_i means the i-th token semantic representation vector in the global context. Unlike traditional static word embeddings, this representation can dynamically adapt according to the context, thereby more accurately reflecting the semantic role of a word in the current context–a capability that is particularly critical for subsequent entity boundary recognition. Furthermore, to capture sentence-level global semantic information, the model typically selects the output vector corresponding to the special token [CLS] as the global representation, denoted as h_N. This vector can comprehensively capture the semantic characteristics of the entire input sequence and is employed in subsequent modules to assist in constructing conditional representations. For instance, during the relation prediction stage, it is fused with the subject representation vector to enhance the model's ability to capture global semantic information.

As shown in Figure 4, the representation sequence output by the contextual encoding module is shared across multiple sub-modules through an "Encoding Sharing" mechanism. On the one hand, the subject labeling module performs boundary prediction for each position. On the other hand, the object and relation prediction module also relies on the same representation for conditional modeling. This sharing mechanism avoids redundant encoding computations while ensuring that different subtasks are modeled within a unified semantic space, thereby enhancing overall model consistency and computational efficiency.

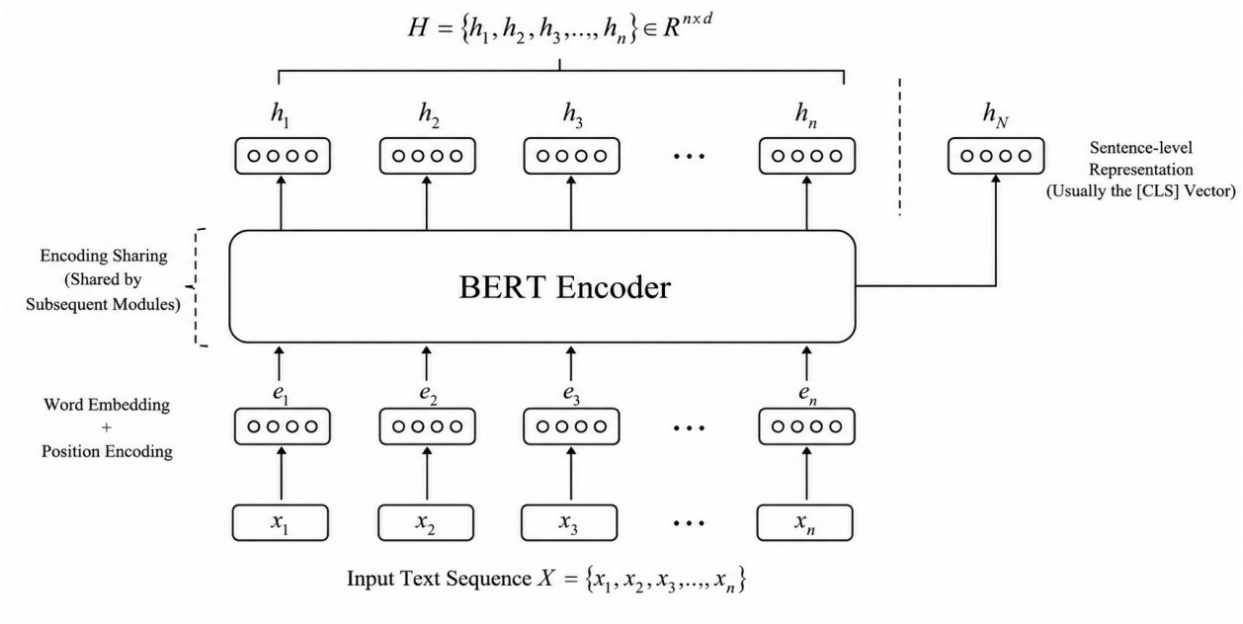

Figure 4: Contextual encoding module architecture diagram.

Through the aforementioned encoding process, the model obtains representations that incorporate both local semantic information and global contextual information, furnishing a reliable feature foundation for subsequent subject identification, span-constrained modeling, and relation extraction. This pre-trained model-based contextual encoding approach demonstrates strong generalization capability in complex textual scenarios and is well-suited to handle the entity-dense and relation-complex characteristics of experimental texts.

3.2. Subject Labeling Module

As a core component of a relation triple, the subject plays a central role, and the accuracy of its boundary identification directly affects subsequent relation recognition and object extraction. Based on the CasRel framework, this work formulates subject extraction as a boundary-based binary classification problem, where the probabilities of

each position in the sequence being the start or end of a subject are predicted independently, so as to determine the span of the subject entity. This formulation avoids the complex label dependencies inherent in traditional sequence labeling methods, enabling the model to handle multiple subjects and overlapping entities more flexibly. Building upon this, the model directly leverages the contextual representations obtained in the preceding section and performs two independent binary classification tasks for each position in the sequence to determine whether it constitutes the start or end of a subject. Through two separate linear transformations with independent parameters, followed by a Sigmoid activation function, the corresponding probability values are obtained as follows:

$$P_{sub}^{start}(i) = \sigma(W_s h_i + b_s) \tag{5}$$

$$P_{sub}^{end}(i) = \sigma(W_e h_i + b_e) \tag{6}$$

This prediction mechanism operates on each position independently, which facilitates the identification of multiple candidate subjects within a single sequence and enhances the model's expressive power when confronted with intricate textual contexts. Once the start and end probability distributions for subjects have been computed, the model employs a threshold strategy with parameter τ to sieve candidate positions, yielding distinct sets of plausible subject start and end positions, respectively:

$$S_{start} = \{i | P_{sub}^{start}(i) > \tau\}, S_{end} = \{j | P_{sub}^{end}(j) > \tau\} \tag{7}$$

Furthermore, by cross-combining elements from the start-position set and the end-position set, the model assembles a collection of candidate subject spans as follows:

$$C_{sub} = \{(i, j) | i \in S_{start}, j \in S_{end}, i \leq j\} \tag{8}$$

It bears noting that the aforementioned process merely conducts an initial generation of candidate boundaries without establishing optimal correspondences between start and end positions. In the original CasRel framework, a relatively rudimentary combination strategy is typically adopted. In contrast, this work introduces a boundary consistency constraint mechanism in subsequent modules to further refine candidate spans, thereby mitigating boundary mismatches and unreasonable span configurations. During the training phase, the subject labeling module is supervised via a binary cross-entropy loss function applied independently to the start and end positions. The loss function is defined as follows:

$$\mathcal{L}_{start} = -\sum_{i=1}^{n} \left[y_i^{start} \log P_{sub}^{start}(i) + (1 - y_i^{start}) \log(1 - P_{sub}^{start}(i)) \right] \tag{9}$$

$$\mathcal{L}_{end} = -\sum_{i=1}^{n} \left[y_i^{end} \log P_{sub}^{end}(i) + (1 - y_i^{end}) \log(1 - P_{sub}^{end}(i)) \right] \tag{10}$$

The overall loss function is defined as:

$$\mathcal{L}_{sub} = \mathcal{L}_{start} + \mathcal{L}_{end} \tag{11}$$

The model effectively extracts potential subject boundary information from contextual representations, furnishing essential conditional inputs for subsequent relation-aware object extraction. Notably, this module is solely responsible for generating candidate subject boundaries and does not engage in complex span pairing operations, thereby providing a clean interface for subsequent constraint-based span optimization strategies. This architectural choice enhances the model's overall extraction performance in complex scenarios involving multiple entities and overlapping relations.

3.3. Span-Constrained Pairing Module

While the preceding module yields predictions for subject start and end positions, the independent modeling of these boundaries leaves them devoid of explicit pairing constraints. This shortcoming readily gives rise to boundary mismatch problems, wherein start and end positions fail to align with the same entity, producing invalid spans. Such defects become especially pronounced in texts characterized by densely distributed entities or intricate syntactic structures, emerging as a critical factor undermining extraction performance. To remedy this deficiency, a span-constrained pairing mechanism is introduced at the candidate subject generation stage to jointly model and refine

candidate boundaries. For any candidate span (i, j), where i means the index of the start position of candidate subjects, j means the index of the end position, and satisfies that $i \leq j$, first, a joint confidence score is constructed based on the boundary prediction results:

$$S(i, j) = P_{sub}^{start}(i) \cdot P_{sub}^{end}(j) \tag{12}$$

where $P_{sub}^{start}(i)$ and $P_{sub}^{end}(j)$ denote positions i and j the probabilities of being predicted as the subject start position and end position, respectively. This score reflects the likelihood that both boundaries are valid simultaneously and serves as the base confidence of the candidate span. Nevertheless, sole reliance on boundary probabilities proves inadequate to guarantee the semantic validity of extracted spans. Hence, a boundary consistency constraint is further incorporated to complement the modeling of candidate spans from a holistic semantic perspective. Specifically, by aggregating the representations within the span interval $[i, j]$, an overall semantic representation of the segment is obtained, which is then leveraged to gauge the coherence between boundary positions and the internal semantics of the span. Whenever the start or end position conflicts with the internal semantics, the score of the corresponding span is penalized, thereby diminishing the likelihood of erroneous boundary matching. Meanwhile, cognizant of the fact that real-world entities typically exhibit certain regularities in length distribution, excessively long or short spans often deviate from valid entity structures. Grounded in this observation, a length constraint term is embedded into the scoring process to regulate span length, thereby suppressing unreasonable candidate spans and enhancing overall prediction stability. Grounded upon this multi-factor modeling, candidate spans are assigned a unified score, which is defined as follows:

$$Score(i, j) = S(i, j) + \lambda_1 C(i, j) + \lambda_2 L(i, j) \tag{13}$$

where $C(i, j)$ denotes the boundary consistency scoring function, which is used to measure the degree of alignment between span boundaries and the overall semantic representation. $L(i, j)$ denotes the length constraint term, which reflects the validity and rationality of the span length. λ_1 and λ_2 are weight parameters used to balance the influence of different constraints. Finally, the model ranks candidate spans according to the scoring function and selects high-scoring results as the subject set.

Relative to the original CasRel framework, the proposed method preserves the boundary prediction architecture while introducing a joint scoring mechanism grounded in semantic consistency and length constraints at the decoding stage, thereby effecting a transition from independent boundary prediction to holistic span modeling. This enhancement not only effectively mitigates the boundary mismatch problem but also markedly improves the accuracy and robustness of subject identification. It exhibits more stable performance, particularly in complex scenarios involving multiple entities and overlapping relations. Analogous span modeling approaches have also been deployed in joint entity-relation extraction tasks in recent studies, demonstrating promising efficacy in intricate entity boundary identification [24].

3.4. Object and Relation Prediction Module

Upon obtaining the optimized set of subject candidates, the model proceeds to identify the corresponding object entities conditioned on each subject and determine the semantic relation between them. Diverging from approaches that first independently extract all entities and subsequently perform relation classification, this module adopts a subject-conditioned decoding strategy, reformulating the relation extraction process as the prediction of object boundaries given a specific subject. In this manner, joint modeling of subjects, relations, and objects is achieved.

For the k-th candidate subject s_k, first, its semantic representation vector is constructed. Specifically, the contextual representations within the subject span are aggregated to obtain the subject representation v_{sub}^k where v_{sub}^k for the k-th semantic vector of the subject, which is used to incorporate subject-conditioned information. Building upon this, the subject representation is fused with the original contextual representations to derive a subject-aware conditional representation:

$$h_i^k = h_i + v_{sub}^k \tag{14}$$

where h_i^k means under the condition of the subject s_k semantic representation at the position. h_i is the representation output by the contextual encoding module. In this manner, the model can explicitly incorporate the current subject information during subsequent prediction, thereby enhancing the specificity of relation modeling. Drawing upon the conditional representation, the model performs object boundary prediction for each relation type separately. Let the

relation set be defined as $\mathcal{R}$, $r \in \mathcal{R}$ denotes a candidate relation. For relation r, the model predicts the probabilities of each position i being the start and end positions of the object entity, respectively:

$$P_{obj}^{start}(i, r|s_k) = \sigma(W_r^{start} h_i^k + b_r^{start}) \tag{15}$$

$$P_{obj}^{end}(i, r|s_k) = \sigma(W_r^{end} h_i^k + b_r^{end}) \tag{16}$$

Under the given subject s_k and r condition, $P_{obj}^{start}(i, r|s_k)$ and $P_{obj}^{end}(i, r|s_k)$ denote the probabilities of i being the start and end positions of the object entity, respectively. W_r^{start}, W_r^{end}, b_r^{start}, and b_r^{end} are learnable parameters corresponding to the relation. By assigning independent parameters to different relations, the model is able to discern the distributional characteristics of object entities under distinct semantic relations.

In complex relation extraction scenarios, a single subject may correspond to multiple objects and multiple relation types. For instance, when describing an experimental reaction process, a reagent may simultaneously participate in a "reaction-with" relation with another substance and a "generation" relation with the resulting product. Grounded in relation-specific object labeling, the model can perform object prediction for different relation types under the same subject condition, thereby effectively accommodating one-to-many relational structures.

When both the predicted start and end position probabilities of an object under a given relation exceed a predefined threshold, the corresponding object span can be combined with the current subject and relation type to assemble a candidate triple. $(s_k, r, o_{k,r})$. $o_{k,r})$ means the predicted object entity s_k under the condition of r. Through this conditional decoding mechanism, the model effectively circumvents spurious combinations between irrelevant entities and enhances the accuracy of triple extraction. During the training phase, this module likewise employs binary cross-entropy loss to supervise the start and end positions of object entities. The overall loss can be formulated as:

$$\mathcal{L}_{obj} = \mathcal{L}_{obj}^{start} + \mathcal{L}_{obj}^{end} \tag{17}$$

where $\mathcal{L}_{obj}^{start}$ and $\mathcal{L}_{obj}^{end}$ denote the prediction losses for the object start and end positions, respectively. Given that their formulation is identical to that of subject boundary prediction, it is not reiterated here.

Through the aforementioned modeling approach, the object and relation prediction module is capable of performing relation-aware object identification conditioned on the subject, thereby enabling effective construction of relational triples. Meanwhile, the high-quality subject boundaries obtained through span-constrained optimization in the preceding module furnish more reliable inputs for this module, which helps curtail error propagation and lays the groundwork for further suppression of false positives in the subsequent candidate filtering module.

3.5. Candidate Filtering Module

Following the object and relation prediction module, the model yields a collection of candidate triples comprising subjects, relations, and objects. Nevertheless, in cascade-based extraction frameworks, low-confidence subjects or objects generated in earlier stages may persist into subsequent combination steps, giving rise to spurious triple generation. This problem becomes especially acute in entity-dense and relation-overlapping scenarios, where entities such as experimental apparatus, reagents, operational procedures, and experimental phenomena frequently co-occur. In the absence of an effective candidate control mechanism, the model may erroneously pair entities with tenuous semantic associations or incomplete boundaries, leading to a proliferation of false positives. Accordingly, this work further devises a candidate filtering module at the decoding stage to sieve and re-rank candidate triples, thereby attenuating the impact of accumulated false positives on the ultimate extraction results.

Let the candidate triple set generated by the previous module be defined as:

$$\mathcal{T}_{cand} = \{(s, r, o)\} \tag{18}$$

where s means candidate subjects, r denotes a candidate relation, o denotes a candidate relation. To gauge the overall reliability of candidate triples, this paper integrates the subject span score, object boundary score, and relation-conditioned prediction score to compute a joint confidence score for each candidate triple. The underlying rationale is that a triple should be retained only when the subject boundaries are reliable, the object boundaries are well-defined, and the relation prediction confidence is sufficiently high.

Specifically, for a candidate triple (s, r, o), its confidence score can be formulated as:

$$Conf(s, r, o) = \beta_1 Score(s) + \beta_2 Score(o|s, r) + \beta_3 P(r|s, o) \tag{19}$$

where $Conf(s, r, o)$ denotes the overall confidence score of the candidate triple; $Score(s)$ denotes the scoring result of the subject span. $Score(o|s, r)$ denotes the score of the object span under the given condition. $P(r|s, o)$ denotes the existence of a relation between the subject and the object. r denotes the confidence between the subject and the object. β_1, β_2, and β_3 are weight parameters used to balance the influence of different scoring components. Through this design, the model ceases to rely solely on individual boundary probabilities for decision-making, and instead conducts a comprehensive evaluation of candidate results from a holistic triple-level perspective. Upon obtaining the confidence scores of candidate triples, this paper adopts a dynamic threshold coupled with a Top-K strategy for filtering. The dynamic threshold is adaptively adjusted according to the distribution of candidate results in the current input text, circumventing the limited adaptability of a fixed threshold across heterogeneous samples. For texts with sparse entities and simpler relational structures, a more stringent threshold can effectively screen out low-confidence candidates. Conversely, for experimental texts with densely distributed entities and intricate relations, judiciously relaxing the threshold helps preserve potentially valid triples, thereby striking a more favorable equilibrium between precision and recall.

The filtering process can be formulated as:

$$\mathcal{T}_{final} = \text{TopK}(\{(s, r, o)|Conf(s, r, o) \geq \theta_{dyn}\}) \tag{20}$$

where $\mathcal{T}_{final}$ means the final retained set of triples. θ_{dyn} denotes the dynamic threshold derived from the distribution of candidate scores. TopK$(\cdot)$ denotes selecting the top-ranked K candidate triples with the highest scores from those satisfying the threshold condition. This strategy effectively filters out patently erroneous candidates while curbing the propagation of false positives stemming from an excessive proliferation of candidate triples.

In the context of complex relation extraction tasks, this module is particularly effective in reducing erroneous triples. For example, in the sentence "A magnesium strip is placed into a dilute hydrochloric acid solution, producing a large number of bubbles," the correct results should mainly focus on the experimental relations between "magnesium strip" and "dilute hydrochloric acid solution," as well as between "magnesium strip" and "bubbles," such as "magnesium strip – placed into – dilute hydrochloric acid solution" and "magnesium strip – produces – bubbles." However, during cascade decoding, if the model incorrectly identifies "dilute hydrochloric acid solution" as the subject, it may further generate unreasonable candidates such as "dilute hydrochloric acid solution – produces – bubbles" or "dilute hydrochloric acid solution – placed into – magnesium strip." The candidate filtering module addresses this issue by jointly considering the overall confidence of triples and their ranking results, preferentially retaining outputs with more salient semantic relations and more reliable boundaries, thereby mitigating false-positive accumulation caused by local prediction errors. Collectively, the candidate filtering module furnishes further constraints and optimization for the outputs of cascade decoding. It operates in tandem with the span-constrained pairing module: whereas the span-constrained pairing module primarily rectifies entity boundary mismatches at the span level, the candidate filtering module suppresses low-confidence combinations at the triple level. Through the synergistic integration of these two mechanisms, the model is able to preserve strong recall while enhancing extraction precision, thereby bolstering its robustness and stability in experimental text scenarios characterized by multiple entities and overlapping relations.

4. Experiments and Result Analysis

4.1. Experimental Environment and Parameter Settings

The proposed model is trained and evaluated in a cloud server environment. The experiments are implemented on the Ubuntu 20.04 operating system using PyTorch 2.0.0 (Python 3.8, CUDA 11.8). All experiments are conducted on a server equipped with a vGPU-48GB graphics card, a 20 vCPU Intel Xeon Platinum 8470Q processor, and 96 GB of memory. The detailed experimental configuration is shown in Table 1.

With respect to model implementation, this study constructs the relation extraction model atop the pre-trained language model BERT, and devises modules encompassing subject labeling, span-constrained pairing, object and relation prediction, and candidate filtering upon this foundation. All input texts undergo tokenization and are subsequently truncated or padded to conform to a maximum sequence length.

The model is trained using the AdamW optimizer for parameter updates. Key hyperparameters are determined based on the experimental implementation and tuned on the validation set. During training, the learning rate is set to 2×10^{-5}, the batch size is set to 6, and the model is trained for 40 epochs. Weight decay is also introduced to improve the generalization ability of the model. The main model hyperparameter settings are summarized in Table 2.

Table 1
Experimental environment configuration.

Configuration Item	Parameter
Operating System	Ubuntu 20.04
Deep Learning Framework	PyTorch2.0.0
GPU	vGPU-48GB×1
CPU	20vCPU Xeon Platinum 8470Q
Memory	96GB

Table 2
Key model hyperparameters.

Parameter	Value
Learning Rate	2×10^{-5}
Batch Size	6
Epoch	40
Maximum Sequence Length	256
Dynamic Threshold Base Value	0.6
EMA Decay Rate	0.999

Table 3
Dataset statistics.

Dataset	Training Set	Validation Set	Test Set	Relations
NYT	56195	5000	5000	24
WebNLG	5019	703	500	239

During the decoding stage, the model employs a dynamic thresholding mechanism to filter subject and object boundaries, while incorporating span length constraints to restrict candidate entities. Concurrently, an exponential moving average (EMA) strategy is introduced during training to smooth parameter updates, thereby enhancing model stability on the validation set. All experiments are conducted within a unified environment under consistent parameter settings to ensure the fairness and comparability of results.

4.2. Datasets and Evaluation Metrics

The datasets used in this study are as follows:

- NYT [25]. The NYT dataset is sourced from news corpora. Its textual content is relatively formal, with clearly articulated entities and relations. It boasts a substantial data scale, encompassing approximately 50,000 training samples and over 5,000 test samples. Owing to its ample size, it is well-suited for evaluating the overall performance of models in general-domain scenarios.
- WebNLG [26]. The WebNLG dataset spans multi-domain knowledge, exhibiting more flexible sentence structures and a more diverse repertoire of relation types. It is comparatively modest in scale, comprising approximately 5,000 training samples and over 500 test samples. Nevertheless, its sentence structures are more intricate, with multiple relations frequently co-occurring within a single sentence, thereby imposing more stringent demands on a model's relation modeling capability.

The evaluation metrics used in this study include precision, recall, and F1-score, which are commonly adopted in relation extraction tasks. Precision measures the proportion of correctly predicted triples among all predicted triples, while recall reflects the proportion of correctly identified triples among all ground-truth triples. The F1-score provides a harmonic balance between precision and recall and serves as the primary evaluation metric in this work. The calculation formulas are as follows:

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}, \mathrm{F1} = \frac{2\mathrm{PR}}{\mathrm{P} + \mathrm{R}} \tag{21}$$

Table 4
Comparison results on the NYT and WebNLG dataset.

	NYT			WebNLG		
Method	P	R	F1	P	R	F1
CopyRE	61.0	56.6	58.7	37.7	36.4	37.1
ETL-Span[27]	85.5	71.7	78.0	84.3	82.0	83.1
RSAN	85.7	83.6	84.6	80.5	83.8	82.1
CasRel[28]	89.7	89.5	89.6	93.4	90.1	91.8
Proposed Method	90.9	90.0	90.4	94.2	90.6	92.4

where P denotes precision. R denotes recall. F1 denotes the overall evaluation metric. True Positive (TP) represents the number of correctly predicted triples, i.e., relations where the predicted results are consistent with the ground truth. False Positive (FP) represent the number of incorrectly predicted triples, i.e., relations predicted by the model that do not exist in the ground truth. False Negative (FN) represents the number of missed ground-truth triples, i.e., relations that exist in the ground truth but are not extracted by the model [27, 28, 29, 30].

Elevated Precision signifies a lower incidence of spurious relations within the predicted outputs, whereas elevated Recall indicates more comprehensive coverage of ground-truth relations. The F1-score harmonizes both dimensions and is therefore adopted as the principal evaluation metric. All subsequent experimental analyses are grounded in these measures.

4.3. Experimental Results and Analysis

To validate the efficacy of the proposed model, comparative experiments are conducted against representative relation extraction approaches. The compared methods encompass traditional pipeline-based models, joint extraction frameworks, and the CasRel baseline. All models undergo training and evaluation on identical datasets under unified experimental conditions to ensure equitable comparability. The experimental results are presented in Table 4.

Table 4 presents the experimental results of different models on the NYT and WebNLG datasets. Overall, the proposed SCOPE-CasRel model attains robust performance across both datasets. Relative to conventional joint extraction approaches, it exhibits superior precision, recall, and F1-score, signifying that the introduced enhancements effectively bolster the overall efficacy of relation extraction models.

On the NYT dataset, the proposed method secures the highest F1-score. The underlying rationale is that the span-constrained pairing mechanism capitalizes on the coherence between entity start and end boundaries to execute optimal matching, thereby curtailing boundary truncation and expansion errors while enhancing entity recognition accuracy. Given that relation extraction performance is critically contingent upon the quality of entity recognition, more precise entity boundaries furnish a dependable foundation for subsequent relation prediction, consequently elevating overall extraction performance.

On the WebNLG dataset, the proposed method likewise surpasses the CasRel baseline. In contrast to NYT, WebNLG typically encompasses a greater density of entities and more intricate relational structures, imposing more exacting demands on the model's decoding capability. The incorporated candidate filtering mechanism effectively sieves low-confidence candidates during relation prediction, diminishing the propagation of erroneous subjects and objects into subsequent inference and mitigating the deleterious effects of error accumulation. Empirical results demonstrate that this strategy enhances precision while preserving a comparatively high recall.

Collectively, experimental outcomes on both datasets corroborate that the proposed span-constrained pairing mechanism and candidate filtering strategy successfully redress the problems of entity boundary misalignment and noisy candidate decoding inherent in the CasRel framework. Notably, in complex relational scenarios, the model is capable of generating more accurate entity spans and attenuating interference from invalid candidates, thereby achieving superior relation extraction performance. This attests to the strong generalization capability and practical effectiveness of the proposed method.

5. Conclusion

To address the challenges of entity boundary mismatch and erroneous candidate propagation in relation extraction tasks, this paper proposes SCOPE-CasRel, a span-constrained decoding relation extraction method built upon the CasRel framework. The proposed method introduces a span-constrained pairing mechanism that explicitly models the coherence between entity start and end boundaries, thereby enhancing the precision of entity boundary identification. Concurrently, a candidate filtering strategy is devised to constrain and eliminate low-confidence candidates, attenuating the impact of error propagation during cascade decoding.

Empirical evaluation on the NYT and WebNLG datasets demonstrates that the proposed method attains strong performance across precision, recall, and F1-score, surpassing the CasRel baseline and other representative relation extraction approaches. The results corroborate that the span-constrained pairing mechanism effectively mitigates boundary mismatch problems, while the candidate filtering strategy curbs false-positive accumulation. In synergy, these components substantially elevate the relation extraction capability of the model.

Future research will further investigate more fine-grained entity span modeling techniques and integrate advanced pre-trained language models and large language models to enrich semantic representations of entities and relations, thereby improving the model's generalization performance and robustness in complex scenarios.

Acknowledgement

The authors declare that there is no funding and no conflict of interest.

References

[1] P. Li, J. Gao, J. Zhang, S. Jin, Z. Chen, Deep reinforcement clustering, IEEE Transactions on Multimedia, 2023, vol. 25, pp. 8183–8193.

[2] J. Zhang, J. Gao, Z. Chen, J. Zhou, Y. Liu, P. Li, A latent adversarial Cauchy-Schwarz autoencoder for medical image segmentation, Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, IEEE, Istanbul, Turkiye, 2023, pp. 3962–3967.

[3] W. Zhang, J. Wang, English text sentiment analysis network based on CNN and U-Net, IFS/ACM Transactions on Machine Learning, 2024, vol. 1, no. 1, pp. 13–18.

[4] Y. Gao, P. Zhang, J. Chen, Aspect sentiment triplet extraction based on semantic enhancement dual encoders, Software Engineering, 2024, vol. 27, no. 12, pp. 5–10.

[5] Y. Lu, Application of AI in the field of documentary heritage: A review of the literature, Journal of Artificial Intelligence Research, 2024, vol. 1, no. 2, pp. 15–21.

[6] M. Canal-Esteve, Y. Gutiérrez, Educational material to knowledge graph conversion: A methodology to enhance digital education, Proceedings of the 1st Workshop on Knowledge Graphs and Large Language Models, Association for Computational Linguistics, Bangkok, Thailand, 2024, pp. 85–91.

[7] S. Choi, Y. Jung, Knowledge graph construction: Extraction, learning, and evaluation, Applied Sciences, 2025, vol. 15, no. 7, pp. 3727:1–3727:41.

[8] Z. Yan, Z. Jia, K. Tu, An empirical study of pipeline vs. joint approaches to entity and relation extraction, Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing, Association for Computational Linguistics, Virtual Event, 2022, pp. 437–443.

[9] D. Zeng, K. Liu, Y. Chen, J. Zhao, Distant supervision for relation extraction via piecewise convolutional neural networks, Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Lisbon, Portugal, 2015, pp. 1753–1762.

[10] M. Miwa, M. Bansal, End-to-end relation extraction using LSTMs on sequences and tree structures, Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Berlin, Germany, 2016, pp. 1105–1116.

[11] Z. Chen, C. Guo, A pattern-first pipeline approach for entity and relation extraction, Neurocomputing, 2022, vol. 494, pp. 182–191.

[12] R. Tang, Y. Chen, Y. Qin, R. Huang, Q. Zheng, Boundary regression model for joint entity and relation extraction, Expert Systems with Applications, 2023, vol. 229, pp. 120441:1–120441:12.

[13] S. Zheng, F. Wang, H. Bao, Y. Hao, P. Zhou, B. Xu, Joint extraction of entities and relations based on a novel tagging scheme, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 1227–1236.

[14] Z. Wei, J. Su, Y. Wang, Y. Tian, Y. Chang, A novel cascade binary tagging framework for relational triple extraction, Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Virtual Event, 2020, pp. 1476–1488.

[15] Y. Wang, B. Yu, Y. Zhang, T. Liu, H. Zhu, L. Sun, TPLinker: Single-stage joint extraction of entities and relations through token pair linking, Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Virtual Event), 2020, pp. 1572–1582.

[16] Y. Shang, H. Huang, X. Mao, OneRel: Joint entity and relation extraction with one module in one step, Proceedings of the 36th AAAI Conference on Artificial Intelligence, AAAI Press, Virtual Event, 2022, pp. 11285–11293.

[17] Y. Li, H. Yan, Y. Zhang, X. Wang, Joint entity and relation extraction combined with multi-module feature information enhancement, Complex & Intelligent Systems, 2024, vol. 10, pp. 6633–6645.
[18] R. Li, K. La, J. Lei, L. Huang, J. Ouyang, Y. Shu, S. Yang, Joint extraction model of entity relations based on decomposition strategy, Scientific Reports, 2024, vol. 14, pp. 1786:1–1786:9.
[19] X. Lu, J. Tong, S. Xia, Entity relationship extraction from Chinese electronic medical records based on feature augmentation and cascade binary tagging framework, Mathematical Biosciences and Engineering, 2024, vol. 21, no. 1, pp. 1342–1355.
[20] D. Li, Y. Yang, J. Cui, X. Meng, J. Qu, Z. Jiang, Y. Zhao, Joint extraction of Chinese medical entities and relations based on RoBERTa and single-module global pointer, BMC Medical Informatics and Decision Making, 2024, vol. 24, pp. 218:1–218:12.
[21] Y. Chen, G. Chen, P. Li, Research on a joint extraction method of track circuit entities and relations integrating global pointer and tensor learning, Sensors, 2024, vol. 24, no. 22, pp. 7128:1–7128:20.
[22] H. Ma, GNN-ERE: A graph neural network framework for entity-relation extraction and intelligent legal case reasoning, Informatica, 2025, vol. 49, no. 25, pp. 431–448.
[23] A. Wang, Graphical learning algorithm based analysis of entity relationships in legal texts, Journal of Combinatorial Mathematics and Combinatorial Computing, 2025, vol. 127b, pp. 1133–1152.
[24] D. Feng, R. Li, J. Wang, S. Yan, L. Ma, Y. Xing, The joint entity-relation extraction model based on span and interactive fusion representation for Chinese medical texts with complex semantics, Proceedings of the 10th China Health Information Processing Conference, Springer, Fuzhou, China, 2024, pp. 196–216.
[25] S. Riedel, L. Yao, A. McCallum, Modeling relations and their mentions without labeled text, Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, Barcelona, Spain, 2010, pp. 148–163.
[26] C. Gardent, A. Shimorina, S. Narayan, L. Perez-Beltrachini, Creating training corpora for NLG micro-planners, Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Vancouver, Canada, 2017, pp. 179–188.
[27] B. Yu, Z. Zhang, X. Shu, T. Liu, Y. Wang, B. Wang, S. Li, Joint extraction of entities and relations based on a novel decomposition strategy, Proceedings of the 24th European Conference on Artificial Intelligence, IOS Press, Santiago de Compostela, Spain, 2020, pp. 2282–2289.
[28] M.-Y. Shih, J.-W. Jheng, L.-F. Lai, A two-step method for clustering mixed categroical and numeric data, Journal of Applied Science and Engineering, 2010, vol. 13, no. 1, pp. 11–19.
[29] F. Song, MPTTF-BERT: Multi-prefix-tuning template fusion by BERT for zero-shot english text relation extraction model, Journal of Applied Science and Engineering, 2025, vol. 28, no. 12, pp. 2561–2571.
[30] T. Wang, Domain adaptive English aspect word extraction method based on bidirectional long and short-term memory network and multi-head attention mechanism, Journal of Applied Science and Engineering, 2025, vol. 28, no. 12, pp. 2661–2669.