

Article

Quintuple Extraction Method for Scientific Papers Based on Feature Words Adversarial Scheme

Yujiang Liu ^{1,2,3}, Lijun Fu ^{2,3,*} and Xiaojun Xia ^{1,2}

¹ University of Chinese Academy of Sciences, No. 1 Yanqihu East Rd., Huairou District, Beijing 101408, China; liuyujia16@mailsucas.ac.cn (Y.L.); xiaojun@sict.ac.cn (X.X.)

² Shenyang Institute of Computing Technology, Chinese Academy of Sciences, No. 16 Nanping East Rd., Dongling District, Shenyang 110168, China

³ Zhongke Zhihe Digital Technology (Beijing) Co., Ltd., Room 101, Building 1, No. 3 Xingfu West Street, Beifang Town, Huairou District, Beijing 101499, China

* Correspondence: fu_lijun@ucas.ac.cn; Tel.: +86-13910090109

Featured Application

This paper proposes a method that enhances the accuracy of entity relation extraction through an unsupervised contrastive learning scheme involving feature words and inverse feature words. Additionally, it improves the connectivity of the constructed knowledge graph and enhances the interpreting ability during feature selection.

Abstract

When extracting entities, relations, and their associated words from scientific literature, it is imperative to consider the supporting role of feature words on the extraction results. These feature words can provide local semantic information and be combined with the global feature representation of the sentence, improving the accuracy of information extraction. However, existing methods, when fusing local semantic feature words with global features, due to ineffective distinction between the influence of feature words and non-feature words, result in limited enhancement on model performance. To solve this problem, we propose a feature words adversarial scheme (FWAS) with dual pointer method. This method implements a dynamic filtering mechanism for feature words through feature pointers, in order to semantically enhance the encoding of the original text. Simultaneously, an inverse feature pointer is designed to establish a negative weight decay mechanism, weakening interference of non-key vocabulary. During joint training, annotation information for entity relations is introduced to supervise the dual feature selection mechanism. Experimental results on three public scientific information extraction datasets demonstrate that our method consistently outperforms strong baselines, achieving up to 4.9% improvement in F1-score. This method offers a new perspective for information extraction tasks in scientific and technical literature and provides scalable optimization directions for subsequent research.

Keywords: relation quintuple; scientific knowledge; inverse feature; pointer network

Academic Editor: David Ruano-Ordás

Received: 20 January 2026

Revised: 9 February 2026

Accepted: 13 February 2026

Published: 24 February 2026

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The task of extracting entities, relations, and their associated words from scientific literature is of significant importance for applications such as knowledge graph construction and information retrieval [1]. Within this task, how to effectively utilize the information of feature words in the text to enhance the extraction effectiveness of entities and relations is a core problem receiving current research attention. Feature words are essentially vocabulary closely associated with entities and relations, usually occurring within the original text, and able to provide local semantic information. Therefore, they play a key role in the information extraction process [2]. Because strong semantic associations exist between feature words and target entities and relations, they can serve as self-augmentation signals for samples, enabling the information extraction model to utilize internal textual information more fully, thereby reducing reliance on additional annotated data. However, existing attention based methods [3,4], when utilizing local semantic feature words, often fail to adequately distinguish between the influence of feature words and non-feature words, thus limiting the model's extraction capability. Against this backdrop, how to construct more effective feature word screening mechanisms to enhance the accuracy and generalization capability of information extraction became the focus of this study.

Since feature words are not explicitly annotated in the dataset, how to automatically acquire them presents a challenge. Existing research indicates that annotation results for entities and relations can serve as supervision signals to indirectly guide the extraction of feature words. The supervision scheme for this process can refer to the association supervision strategy mentioned in the previous research section. Because feature words and non-feature words lack obvious formal distinction in textual expression and are often intermingled, traditional rule based methods struggle to accurately identify these keywords. Weakly supervised methods [5,6] also fail to address this issue in long tail and adversarial features. Therefore, we refer to the semantic contrastive scheme in text classification tasks, adopting contrastive learning to optimize the feature word screening process under unsupervised conditions [7]. This approach also aligns with current work on sentence augmentation in this context [8], both being association strengthening ideas [2]. To some degree, these irrelevant words exert a negative influence on the final extraction results; we further propose the concept of inverse feature words to describe them. Such words require a degree of weakening to enhance extraction reliability.

The research approach targeting feature words and inverse feature words can refer to the semantic contrastive scheme in text classification. Existing research demonstrates that text possesses both similarity and dissimilarity; comparing these two aspects can be applied to text classification tasks [9]. This idea can be further extended to the quintuple extraction task, namely, optimizing extraction performance using feature words and inverse feature words. Work utilizing non-self-augmenting external information inputs may require additional annotation costs, while also necessitating consideration of resolving conflicts between multi-dimensional features. When the extraction process relies on entities, sequence and token-level perturbations can be introduced into samples to enhance adaptability to entity and contextual disturbances, and probabilistic strategies are adopted to preserve clean token information during adversarial training [2]. When spurious associations exist between results and text vectors, to reduce the model's sensitivity to spurious correlations, research has attempted to use counter-factually augmented data for causal modeling [10]. Our method does not introduce external noise or additional counterfactual annotations, instead, it employs lexical contrastive training to perform unsupervised optimization on unlabeled original sentence associated words, thereby enabling the model to utilize more feature words for quintuple extraction and further enhance

relation modeling capability. This process can homogenize multiple features into feature words, thus enabling word-level advisability [11].

This paper proposes a feature word adversarial scheme (FWAS) with dual pointers, enhancing the model's ability to learn local textual features by dynamically screening feature words and weakening the impact of low contribution vocabulary. Dynamic filtering employs the idea of feature fusion [12–14], utilizing methods akin to interactive network structures [15] and semantic drift control [16] to steer the selection of feature words and inverse feature words by the pointer. The core of the current scheme lies in using two pointers to extract an equal number of feature words and inverse feature words, then strengthening feature words while weakening inverse feature words to achieve maximal contribution from the original feature words in this process of reciprocal growth. We draw inspiration from dynamic weight adjustment [17] and multi-pointer strategies [18–20] to achieve adaptive regulation of the influence of feature words. Since the original text, after encoding, outputs a sentence feature vector representing global features. This content, together with the word vector features represented by feature word vectors, requires learning to enhance the model's understanding of samples [21,22]. The dual pointer scheme proposed comprises three core modules: feature pointer, inverse feature pointer, and joint training. The feature pointer mechanism enables FWAS to gradually filter high-contribution feature words, strengthen the encoding of original text, and enhance the extraction accuracy of entities, relations, and their associated words. The inverse feature pointer mechanism allows FWAS to adopt an adversarial learning-based approach to construct a negative weight decay strategy. By weakening the weight values of unnecessary non-feature words, the model is able to generalize. This is different from general adversarial training negative sample enhancement [23] because it involves more refined optimization at the pointer decoding level, reducing the influence of inverse feature words through negative sampling strategies [24]. During the joint training strategy, the annotated entity relationship data are used for the dual pointer mechanism in FWAS's supervised training. This utilizes the joint semantics between feature words to improve the training dictionary for named entity recognition and relation extraction, thereby enhancing the model's grasp of contextual information. Our contributions are as follows:

- Propose a dual pointer-based feature word screening scheme. We enable the model to achieve a dynamic balance between feature enhancement and interference suppression, thus improving the accuracy of entity-relation extraction.
- Propose an annotation-free data augmentation strategy. We optimize unannotated original sentence-associated words through adversarial training, thereby increasing the model's adaptability to information extraction tasks.
- Demonstrate the role of local feature enhancement in quintuple extraction. As a global-sentence-encoding scheme, the current strategy can better transform word features within original sentences while fusing them with original sentence encoding.
- Combine adaptive word features for relation extraction. We try to explore how to homogenize multiple features into feature words, enabling the model to more effectively utilize local information for adversarial feature enhancement.

2. Materials and Methods

2.1. Scheme Overview

In this study, we define the quintuple extraction task as a joint framework that combines original sentence associated word features with inverse feature optimization, simultaneously performing entity recognition, relation prediction, and feature word extraction. Given an input sentence $S = w_1, w_2, \dots, w_n$, the system identifies and extracts the

quintuple $T(S) = (s, t, o), I(S)$, where (s, t, o) consists of the head entity s (with its entity type), relation type t , and tail entity o (with its entity type); $I(S) = i_1, i_2, \dots, i_t$ denotes the index set of feature words.

Here, the sentence contains n words (denoted by w), and $I(S)$ contain t feature words indexed by i . By incorporating feature words and inverse feature words during entity recognition and relation prediction, the model effectively enhances its understanding of truly meaningful feature words within complex linguistic structures and improves the accuracy of relation determination. Therefore, we adopt a joint optimization strategy: leveraging shared parameters to enhance model generalization capability, while integrating an attention mechanism to weight feature words, dynamically adjusting their contribution during quintuple extraction. This algorithmic process is illustrated in Figure 1.

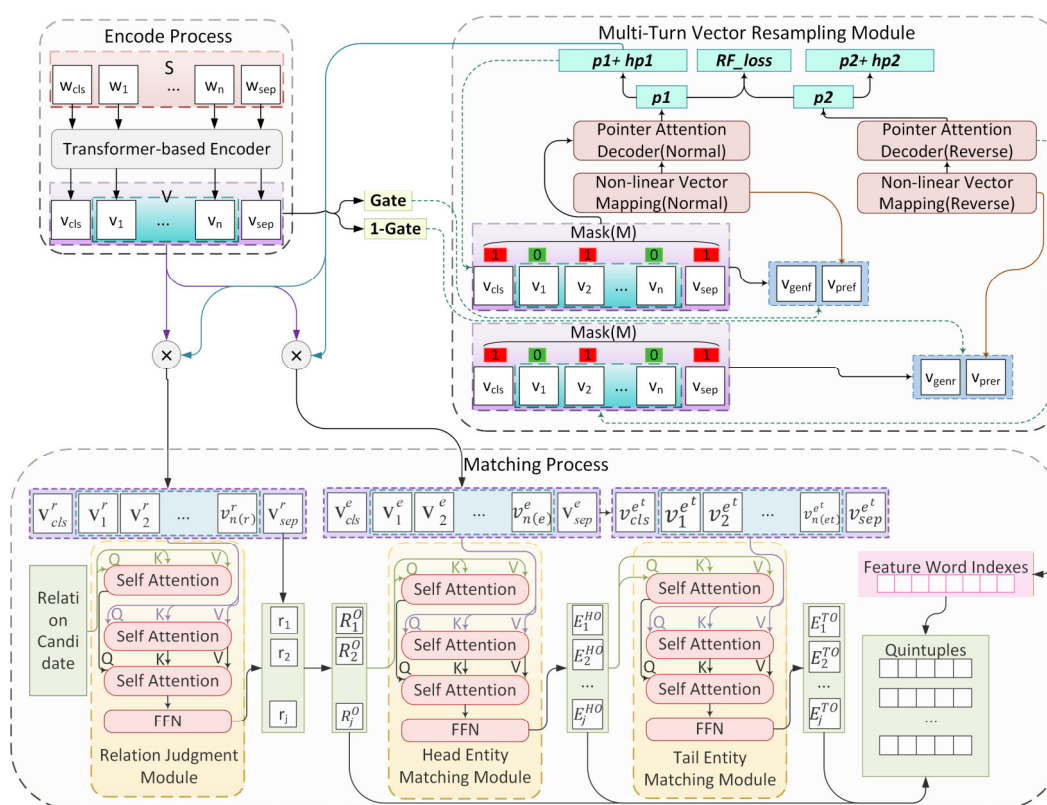


Figure 1. Algorithm structure diagram of the Feature Words Adversarial Scheme (FWAS). In the current structure, modifications were made to the multi-round sampling and entity generation processes via a reinforcement learning scheme to achieve adversarial strategy modeling. We use value 1 to cover the chosen indexes, as the red parts of the masks.

2.2. Encoder

In the quintuple extraction task, the core role of the encoder lies in transforming raw text into high dimensional vector representations, enabling subsequent computational modules to leverage these vectors for relation determination, entity recognition, and feature word extraction. In this study, we adopt a Transformer-based encoder (e.g., BERT) as the text representation backbone, aiming to utilize its powerful contextual modeling capabilities and rich prior knowledge to enhance the semantic representational capacity of input text. As a pretrained model driven by bidirectional self-attention mechanisms, BERT captures deep dependency relationships between vocabulary, thereby providing more effective semantic representations for complex sentence structures and long-range dependencies.

During the encoding process, the input text is first segmented into subword units and mapped to a fixed length subword index sequence. This sequence is subsequently embedded into BERT's word vector space and encoded through its multi-layer Transformer architecture. BERT's input consists of a token vector sequence $V = v_{cls}, v_1, \dots, v_n, v_{sep}$, where v_{cls} and v_{sep} correspond to the special tokens [CLS] and [SEP], respectively, representing global sentence information and sentence boundaries. After computation through multiple Transformer layers, the encoder outputs a vector sequence of the same length as the input. The hidden state vector corresponding to v_{cls} serves as the global representation of the entire sentence, ensuring the model captures comprehensive semantic information during relation prediction and feature word extraction. The remaining vectors $v_1, \dots, v_n$ represent context-aware representations of each word in the input text (i.e., word vectors), which construct fine grained features for the quintuple to deliver richer entity and relation characterizations.

2.3. Feature Pointer Mechanism and Dynamic Filtering Strategy

We employ a dynamic filtering process to select feature words and inverse feature words. This process first initializes their feature vectors, denoted as v_{mf} and v_{mr} , respectively. An iterative procedure is then applied over t time steps to obtain t feature words and inverse feature words. During the feature selection stage, inverse feature words are filtered out and only feature words are retained to enhance the representation of the original text. At a specific time step τ , the selected feature words strengthen the features at the corresponding positions in the source text, whereas the selected inverse feature words suppress the features at their corresponding positions. The validation of the parameter τ will be conducted on a representative dataset in the experimental section.

During each feature word extraction step, two input vectors v_{genf} and v_{pref} are fused through a weighting scheme to generate a new vector v_{mf} . This vector is then fused with the source text vector V to select the feature word for the current timestep. The inverse feature word selection method follows the same process but relies on vectors denoted as v_{genr} and v_{prer} .

Differentiating initial inputs for feature words and inverse feature words uses a self-weighting scheme based on the global sentence representation vector v_{cls} , implemented via a binary activation mechanism. v_{cls} is transformed by a gating mechanism into a weight value between 0 and 1. This weight is multiplied by v_{cls} to initialize the feature word input, while the inverse feature word input uses 1 minus this weight. The formulas are defined as:

$$v_{mf}(0) = \tanh(W_{gf}(\text{sigmoid}(W_z v_{cls} + b_z) * v_{cls}) + b_f) \quad (1)$$

$$v_{mr}(0) = \tanh(W_{gr}((1 - \text{sigmoid}(W_z v_{cls} + b_z)) * v_{cls}) + b_r) \quad (2)$$

For $t \geq 1$:

$$v_{mf}(t) = \tanh(W_{gf}v_{genf}(t) + W_{cf}v_{pref}(t) + b_f) \quad (3)$$

$$v_{mr}(t) = \tanh(W_{gr}v_{genr}(t) + W_{cr}v_{prer}(t) + b_r) \quad (4)$$

Here W_{gf} , W_{cf} , W_{gr} , W_{cr} denote linear transformation weights, and b_f , b_r are bias terms. After obtaining v_{mf} and v_{mr} , they undergo dimensional expanded linear fusion with vector V for the next extraction step. Taking v_{mf} as an example, it is first expanded to v_{mf} to match the second dimension of V , allowing concatenation at the word vector level. Probabilities are then computed to identify the word vector position with maximum probability as the selected feature word vector, while retaining probability values for gradient back propagation stability. The selection formula is:

$$c_{mf}(t) = W_j[V_{mf}, V] + b_j \quad (5)$$

$$p(t) = \text{Softmax}(M(t) + W_k(\tanh(c_{mf}(t))) + b_k) \quad (6)$$

$$hp(t) = \text{maxidx}(p(t)) \quad (7)$$

where W_j , W_k denote trainable linear transformation matrices, b_j , b_k are bias terms, and all parameters are optimized jointly during training. $\text{maxidx}(\cdot)$ denotes an operator that returns the index of the word vector with the maximum selection probability at the current timestep. This mask $M(t)$ must evolve throughout the feature word extraction process. Because no explicit annotations exist for quintuple associated words in this scenario (equivalent to unsupervised learning), the mask requires adjustment at each timestep. The core principle is that words selected in previous timesteps (either feature or inverse feature words) are excluded from consideration at the current timestep. Therefore, this mask $M(t)$ must have the same length as the original sentence, be initialized with all zeros to indicate initial availability of all words, and be updated at each timestep by setting positions corresponding to selected feature or inverse feature words to 1 for exclusion from future consideration. The index $hp(t)$ (indicating the word with maximum probability at step t) updates the mask as:

$$M(t + 1) = hp(t) + M(t) \quad (8)$$

When updating v_{genf} , it is necessary to incorporate the updated result of the source text influenced by the selected feature word. It means the source text enhanced at timestep t contributes to selecting the next feature word. For updating v_{pref} , the computation result of v_{mf} from the previous timestep is directly assigned to it. The updates for the inverse feature words' corresponding vectors v_{genr} and v_{prer} , along with their dependent vector v_{mr} , follow the exact same scheme. The formulas are defined as:

$$v_{pref}(t + 1) = v_{mf}(t) \quad (9)$$

$$v_{genf}(t + 1) = \text{Maxpool}((hp(t) + p(t)) * V) \quad (10)$$

After all feature words are selected, we exclusively utilize the aggregated probability values from these selections to update the original vector V , achieving feature word enhancement. The weakening effect of inverse feature words is realized indirectly via a loss constraint during training, influencing the feature word sampling process without direct intervention. This procedure is formally expressed as:

$$V_f = V * \sum_{i=0}^t p(t) \quad (11)$$

2.4. From Relation Templates to Entity Matching

This component primarily outputs entity and relation results for the quintuple. After incorporating feature word influence into the source text, relation templates are employed as input to match potential relations in the text. Subsequently, the head entity is identified based on this result, bound to the relation template, and finally encoded to locate the tail entity. During template matching, the encoded relation template representation R , containing both relation types and entity types, determines possible joint relations in the sample. The relation feature representation V_r is generated through linear transformation of the feature enhanced text representation V_f , which is represented as $V_r = W_R V_f + b_R$. This V_r then interacts with R via an interactive attention operator where both inputs first un-

dergo self-attention computation, followed by cross-attention calculation, ultimately producing an output through a fully connected network, yielding the potential joint relation representation R^O .

For entities, we consider representations spanning all word vectors across the text. That is, the entity representations derived after span computation. When consecutive words form an entity, their computation content is represented using entity boundaries and internal max-pooling, denoted as:

$$v_i^{(e)} = [v_{start}^{(e)}(i), v_{end}^{(e)}(i), \maxpool(v_j^{(e)}(i)), \theta^{(e)}(v_i)] \quad (start \leq j \leq end) \quad (12)$$

$v_{start}^{(e)}(i)$ and $v_{end}^{(e)}(i)$ denote the start and end boundary vectors of entity i , while $\theta^{(e)}(v_i)$ represents its length information. This indexed representation constructs a new sequence serving as the preliminary vector for generating the entity feature representation V_e . A linear transformation is then applied, which is $V_e = W_e V_f + b_e$. This considers the vector obtained after the first transformation as the hidden representation for head entity dependency, while the tail entity utilizes V_e for a second transformation. To distinguish from the head entity, a fixed value vector ε is added. This process can be described as:

$$V_t^e = W_t[V_e; \varepsilon] + b_t \quad (13)$$

where W_t denote trainable linear transformation matrices, b_t are bias terms. When extracting entities, R^O , V_e and V_t^e require dimension expansion to resolve inter stage dependencies. Since R^O 's second dimension encodes template relation counts per sample, while V_e 's second dimension tracks all possible entities (dimensional incompatible), both tensors undergo mutual dimension expansion. Boundary values are constrained via $\tanh(\cdot)$, followed by binding probability estimation through $\sigma(\cdot)$ (sigmoid activation) at corresponding positions. For tail entities, the bound head template result E^{HO} serves as input to replicate the same process, yielding the head-to-tail entity binding E^{TO} .

$$E^{HO} = \text{sigmoid}(W_{ho}(\text{attention}(\text{expand}(R^O) + \text{expand}(V_e))) + b_{ho}) \quad (14)$$

$$E^{TO} = \text{sigmoid}(W_{to}(\text{attention}(\text{expand}(E^{HO}) + \text{expand}(V_t^e))) + b_{to}) \quad (15)$$

Here W_{ho} , W_{to} are trainable weight parameters, and b_{ho} , b_{to} are bias terms within the joint training process.

2.5. Loss Functions

The current task's loss function design focuses on three aspects: feature words, inverse feature words, relations, and entities. Although inverse feature words do not directly modify source text word vectors, they must be significantly distinct from feature words, enforced through loss supervision. Differences between inverse and standard feature words are maximized using reinforcement learning principles, where supervision targets gap amplification rather than individual words. At step t (when the hp mask stabilizes), all feature/inverse words are identified. A primary loss of the features is constructed to maximize dissimilarity through reciprocal terms:

$$L_{feature} = hp(i) * \sum_{i=1}^t \left(1 / (\exp(p(i) - rp(i))^2) \right) \quad (16)$$

where $p(i)$ and $rp(i)$ denote the probability distributions of feature words and inverse feature words at step t , respectively. Maximizing the divergence between $rp(i)$ and $p(i)$ enlarges the distinction between feature and inverse feature words. For relations and entities, cross-entropy loss functions are applied:

$$L_{entity} = -\frac{1}{2m} * \sum_{i=1}^m (y_i^{(ho)} \log(E^{HO}) + y_i^{(to)} \log(E^{TO})) \quad (17)$$

$$L_{relation} = -\frac{1}{m} * \sum_{i=1}^m (y_i^{(r)} \log(R^O)) \quad (18)$$

$$L_{total} = L_{feature} + L_{entity} + L_{relation} \quad (19)$$

Here m denotes the maximum count of entity-relation tuples in a sample. These three loss functions combine as the model's optimization objective. Because each tuple contains two entities, the entity loss component carries a halved weight parameter to balance contributions among them.

3. Results

3.1. Datasets

In relation extraction tasks, datasets are required to cover both entity types and relation types to support information combinations that conform to the quintuple structure. To meet this requirement, we selected multiple standard datasets spanning scientific research, news, and encyclopedic domains, allowing the evaluation of the model's performance across diverse scenarios. For experiments focusing solely on relation triplet extraction, we included datasets from non-scientific contexts to analyze the model's generalization ability.

In the scientific domain, we used SciERC [25], SemEval 2017 Task 10 [26], and SemEval 2022 Task 12 (SymLink) [27]. SciERC contains 500 scientific paper abstracts with 4648 annotated entity-relation pairs, encompassing paper abstracts across multiple scientific disciplines, annotated with entities, relations, and coreference chains, enabling evaluation of model capability in high specialization texts. SemEval 2017 Task 10 includes 676 paper excerpts and 1073 entity-relation pairs from multiple disciplines, which focuses on research findings and conclusion-oriented texts with high knowledge density, testing model performance in processing highly summarized sentences. SemEval 2022 Task 12 features 7815 mathematical symbols with 20,987 annotated relations, retaining the characteristics of the former while supplementing document-level relations and scientific-notation-related entities. This experiment adopts strict entity-relation quintuple matching (Rel+) to ensure extraction precision. Simultaneously, original sentence associated words were quantitatively analyzed to further investigate feature words' contribution to relation determination. We compared our experimental results with the latest baselines, the names of which are listed in Tables 1 and 2.

To verify generalization capability, two open domain datasets, NYT [28] and WebNLG [29], were introduced. Originating from news and encyclopedia domains, respectively, these datasets exhibit significant stylistic and structural differences from the scientific literature. NYT10, derived from The New York Times, contains 67,312 sentences with 84,832 relations, focusing on main entities and their relationships in news articles. WebNLG, from encyclopedic texts, includes 13,420 sentences with 6946 relations, covering various knowledge descriptions. Given the absence of explicit entity-type annotations in NYT and WebNLG, evaluation employed standard relation triple matching. To examine robustness across data partitioning strategies, both original (NYT, WebNLG) and pre-processed extended versions (NYT*, WebNLG*) were utilized. The original versions include supplementary coreference resolution information to mitigate noise interference. This partitioning design facilitates analysis of the adversarial mechanism's stability across varying data distributions and textual styles, further assessing its applicability to multi-domain relation extraction tasks.

Table 1. Experimental results on SciERC and SemEval 2017 Task 10 datasets. P, R, and F denote Precision (%), Recall (%), and F1-score (%), respectively, (using “Rel+” with strict entity-type constraints).

Dataset	Baseline	P	R	F
SciERC	SpERT.PL [30]	39.94	38.98	39.41
	SpERT.PL + OVO [31]	-	-	41.5
	SpERT [32]	36.25	40.04	38.05
	Trimf [33]	42.27	39.01	40.58
	MCER [34]	-	-	42.8
	MTL [35]	44.33	35.46	39.66
	HGERE [36]	-	-	43.6
	ATG [37]	-	-	38.6
	UIE [38]	-	-	36.53
	GraphER [39]	-	-	39.4
	PGA [40]	43.19	35.52	39.27
	TRUE-UIE [41]	-	-	37.4
	UIE-MPSR [42]	-	-	35.77
	IEPile [43]	-	-	45.89
	Set [44]	-	-	35.9
	UTC-IE [45]	-	-	38.77
	SKIE [46]	-	-	46.9
	TIE [47]	-	-	34.41
	CARE [48]	-	-	40.9
	FSN [49]	-	-	42.3
	FWAS	53.61	51.13	52.34
Semeval 2017 task 10	SCIIE [25]	40.4	21.2	27.8
	DyGIE [50]	35.11	24.23	28.7
	SpERT	45.22	24.88	32.1
	Trimf	47.25	25.62	33.2
	FWAS	42.05	39.23	40.59
Semeval 2022 task 12	SpERT	34.72	26.22	29.88
	AIFB [51]	36.52	27.25	31.2
	JBNU [52]	40.34	34.09	36.96
	FWAS	44.36	39.7	41.9

Table 2. Dataset length analysis. This section displays information on the length of the original text, entities, and entity spans for three scientific literature datasets.

Dataset	Avg Sentence Length	Avg Head Entity Length	Avg Tail Entity Length	Avg Entity Gap Length
Semeval 2017 task 10	37.28	3.02	2.93	5.87
Semeval 2022 task 12	109.29	3.25	4.95	18.34
SciERC	27.17	2.69	2.74	5.72

3.2. Evaluation Metrics

For experimental implementation details, we employ BERT-base-cased and SciBERT-base as text encoders to evaluate general-domain versus scientific-literature-specialized pre-trained models. BERT-base-cased provides broad cross-domain knowledge and is used for two Semeval dataset, while SciBERT-base optimizes for scientific contexts and is used for SciERC, capturing relational patterns in academic discourse more effectively. Batch size is set to 16 with 150 training epochs. Initial learning rate is fixed at 1×10^{-5} .

using warmup strategy (gradually increasing learning rate initially followed by progressive decay) to enhance convergence stability. AdamW optimizer with gradient clipping prevents local optima and training oscillation.

All experiments ran on NVIDIA RTX 3090 Ti GPUs to ensure computational efficiency and large-scale data processing capability. To enhance result reliability, model performance was evaluated on the validation set after each epoch, with optimal parameters saved for final benchmarking. We implemented an early stopping strategy, terminating training when no performance gain was observed over consecutive epochs, to optimize computational resource utilization while mitigating overfitting risks. For robustness verification, multiple random initialization trials were conducted, with statistical analysis of model performance across different random seeds assessing stability and generalization capability. These configurations ensure comparative experiment fairness and establish a reliable foundation for model evaluation.

3.3. Basic Experiment Results

The results of baseline methods reported in Tables 1 and 2 are taken from their original papers under the same evaluation protocols whenever available. For methods that do not report full Precision and Recall values under the Rel+ setting, only the F1-score is presented, following common practice in the literature. The baselines included in Table 1 are modeled using joint extraction mode and graph optimization mode, and the F-values of all or part of the scientific literature dataset are provided. The specific method names are detailed in the table. The baselines included in Table 2 are modeled in a relationship optimization mode, mainly targeting general datasets.

Table 1 presents experimental results of the proposed FWAS on three scientific literature datasets: SciERC, SemEval 2017 Task 10, and SemEval 2022 Task 12. Under the Rel+ evaluation protocol, the scheme outperforms all baseline models, validating the effectiveness of the feature inverse feature adversarial mechanism in joint entity-relation extraction. The F1 scores of 52.34%, 40.59%, and 41.9% were obtained on SciERC, SemEval 2017 Task 10, and SemEval 2022 Task 12 datasets, respectively, demonstrating significant improvements over other baselines. These results confirm the robustness of the feature word contrastive mechanism, indicating that introducing inverse feature words enhances feature word screening accuracy while optimizing global sentence level representations. Compared to traditional single directional feature selection methods, the dual pointer mechanism stabilizes feature word selection through unsupervised adversarial training, thereby strengthening semantic representation capabilities.

From a finer-grained analysis, the core function of the feature inverse feature adversarial mechanism lies in enhancing model sensitivity to critical feature words while suppressing interference from low contribution vocabulary in relation modeling. This effectiveness is evidenced by F1-score improvements. The dual-pointer mechanism dynamically regulates feature word screening weights in an unsupervised environment, maintaining robust adaptability across diverse data distributions. This further attests to its generalization capability in multi-domain texts. Gradual convergence of feature word screening accuracy during training indicates increasingly effective discrimination between key features and irrelevant vocabulary, driving enhanced overall relation extraction performance through adversarial optimization. This conclusion applies to the SemEval 2022 Task 12 dataset, which contains a large number of cross-sentence relationships and relatively long mathematical descriptions. According to Table 2, the Semeval 2022 task 12 dataset is significantly longer than other datasets, but our SWAF method still manages to improve performance. This indicates that the current method is suitable for long texts, but its effect is more significant for shorter texts.

Table 3 presents the proposed scheme’s performance on NYT and WebNLG datasets, where F1-score improvements are less pronounced compared to SciERC and SemEval 2017 Task 10. Given the high baseline F1-scores ($\approx 95\%$) on NYT and WebNLG, the limited optimization headroom constrained enhancement magnitude relative to the scientific literature benchmarks. Nevertheless, this marginal gain confirms that the feature inverse feature adversarial mechanism retains optimization efficacy even against high performance baselines. Analysis reveals distinct contributions: for NYT, the mechanism primarily improves relation matching accuracy in long sentences; for WebNLG, it enhances adaptability to diverse syntactic structures. This suggests that refined feature word screening reduces interference from irrelevant information during relation extraction, even in scenarios where existing methods excel.

Table 3. Experimental results of the FWAS on NYT, NYT*, WebNLG, and WebNLG* Datasets.

Baseline	NYT*			WebNLG*			NYT			WebNLG		
	P	R	F	P	R	F	P	R	F	P	R	F
EmRel [53]	91.7	92.5	92.1	92.7	93	92.9	92.6	92.7	92.6	90.2	87.4	88.7
GRTE [54]	92.9	93.1	93	93.7	94.2	93.9	93.4	93.5	93.4	92.3	87.9	90
ConCasRTE [55]	92.9	92.3	92.6	93.8	92.5	93.1	92.9	92.1	92.5	90.6	88.1	89.3
BiRTE [56]	92.2	93.8	93	93.2	94	93.6	91.9	93.7	92.8	89	89.5	89.3
CasRel [57]	89.7	89.5	89.6	93.4	90.1	91.8	-	-	-	-	-	-
TP-Linker [58]	91.3	92.5	91.9	91.8	92	91.9	91.4	92.6	92	88.9	84.5	86.7
PRGC [59]	93.3	91.9	92.6	94	92.1	93	93.5	91.9	92.7	89.9	87.2	88.5
RFBFN [60]	93.4	93.2	93.3	93.9	94.1	94	93.7	93.6	93.6	91.5	89.4	90.4
RPSS [29]	93.5	93.2	93.3	94.7	95.1	94.9	-	-	-	-	-	-
PRE-Span [50]	90	85.3	88	95.5	92.9	94.2	88.6	84.7	86.6	83.4	82.7	83
BTDM [30]	93	92.5	92.7	94.1	93.5	93.8	93.1	92.5	92.7	90.9	90.1	90.5
SPN [61]	93.3	91.7	92.5	93.1	93.6	93.4	-	-	-	-	-	-
NEDORT [62]	91.8	89.7	90.7	92.2	91.5	91.9	-	-	-	-	-	-
DERP [63]	92.1	90	91	92.8	92.9	92.9	-	-	-	-	-	-
ODRTE [64]	93.5	93.9	93.7	94.6	95.1	94.9	94.2	93.6	93.9	92.8	92.1	92.5
PRE-Span [65]	90.9	85.3	88.0	95.5	92.9	94.2	88.6	84.7	86.6	83.4	82.7	83.0
UIE [38]	-	-	-	-	-	-	-	-	93.5	-	-	-
DS [66]	-	-	-	-	-	-	86.5	73.2	79.3	85.3	83.1	84.2
GLM-TripleGen [67]	94.2	93.6	93.9	94.2	93.6	93.9	-	-	-	-	-	-
SGEP [68]	-	-	-	-	-	-	91.5	74.4	82.1	85.5	84.5	85.2
Set [44]	-	-	93.5	-	-	-	-	-	92.2	-	-	-
CARE [48]	-	-	92.6	-	-	93.9	-	-	-	-	-	-
IT [69]	94.2	93.4	93	93.6	93.7	93.2	94.8	93.8	93.2	92	91.2	91.5
FWAS (Ours)	94.7	94.5	94.6	96.3	96.4	96.3	94.8	95.1	94.9	92.2	92.1	92.1

3.4. Evaluation Results Without Enhancement Mechanism

To further validate the impact of the enhancement mechanism on feature word screening, we removed either $p(t)$ or $hp(t)$ from computations while retaining the other component, comparing performance changes before and after removal. As shown in Table 4, removing $p(t)$ caused F1-score declines of 2.2% and 2.5% on SciERC and SemEval 2017 Task 10 datasets, respectively, while removing $hp(t)$ resulted in 1.5% and 1.8% reductions. This confirms the scheme’s critical role in transmitting supervisory signals between feature and inverse feature words. Eliminating $p(t)$ induced nondifferentiable conditions during backpropagation, explaining its more severe performance degradation. Theoretically, adversarial learning alone can drive feature word screening without

enhancement mechanisms. However, convergence slows due to insufficient global supervisory signals during backpropagation, ultimately compromising extraction efficacy. These results validate the enhancement scheme's function in optimizing supervisory signal transmission and ensuring stable feature inversed screening.

Table 4. Ablation study of the enhancement.

Dataset	Baseline	P	R	F
SciERC	Origin results	53.61	51.13	52.34
	Remove p(t)	49.52	50.33	49.92
	Remove hp(t)	51.85	49.97	50.89
Semeval 2017 task 10	Origin results	42.05	39.23	40.59
	Remove p(t)	40.18	36.09	38.03
	Remove hp(t)	39.73	37.76	38.72
Semeval 2022 task 12	Origin results	44.36	39.7	41.9
	Remove p(t)	42.85	37.79	40.16
	Remove hp(t)	42.19	38.72	40.38

From a finer-grained analysis, removing the enhancement mechanism significantly degrades feature word selection capability in complex syntactic structures, particularly for texts with multi-relation cooccurrence and long-range dependencies. This is because the heightened semantic ambiguity in feature/inverse feature distributions. The critical role of supplementary supervisory signals in optimizing feature screening is thus underscored, as the enhancement mechanism extends the reach of backpropagation signals across broader textual contexts. This strengthens continuous modeling of feature screening and enhances feature inverse feature discriminability, concurrently mitigating information dispersion in complex texts for more targeted screening. Theoretically aligned with our methodological derivations, the mechanism bolsters global feature alignment during backpropagation, reducing training volatility from instance specific biases. Absent this mechanism, models become overly dependent on local information while failing to leverage global context, increasing vulnerability to noise interference and critical feature omission, ultimately compromising extraction efficacy.

3.5. Ablation Studies on Inverse Feature Words

To further investigate the efficacy and optimization potential of our proposed contrastive strategy between feature words and inverse feature words, we conducted ablation experiments focusing on two aspects: loss function modifications for feature words and the direct impact of inverse feature words on original sentences. These studies enable in-depth analysis of how different strategies affect model performance and clarify the mechanistic roles of feature and inverse feature words in the task.

We first modified the loss function used in the reinforcement learning process between feature and inverse feature words, implementing two adjustments: (1) Removing the natural logarithm e , using squared difference directly as the denominator; (2) Eliminating multiplication with hp , retaining all feature/inverse feature word positions for supervision. Results after these modifications are shown in Table 5. Experimental data reveal performance degradation under both adjustments due to:

First Modification (Natural Logarithm Removal): The natural logarithm serves as a smoothing mechanism in the loss function, preventing minor differences from disproportionately affecting gradients during computation. For instance, when differences are small (e.g., $e0.01 \approx 1.01$), the log function amplifies subtle differences during backpropagation,

mitigating vanishing gradient issues. Removing it simplifies computation but causes minor differences to be ignored during training, degrading gradient update effectiveness. It is the primary reason for performance degradation.

Second Modification (hp Multiplication Elimination): Multiplying with hp confined supervision to position specific impacts for feature/inverse feature words at each timestep. Removing this restriction applies identical supervision signals to all positions, diluting supervisory focus away from high impact words. Consequently, optimization direction becomes ambiguous, preventing effective learning of optimal feature/inverse feature words and reducing performance.

Table 5. Ablation experiments on inverse feature words.

Dataset	Baseline	P	R	F
SciERC	Origin results	53.61	51.13	52.34
	Delete the natural logarithm e	47.13	50.42	48.72
	Not multiplied by hp(t)	42.57	44.88	43.69
	Insert inverse feature words into the original sentence	51.75	50.88	51.31
Semeval 2017 task 10	Origin results	42.05	39.23	40.59
	Delete the natural logarithm e	38.43	38.02	38.22
	Not multiplied by hp(t)	36.79	34.62	35.67
	Insert inverse feature words into the original sentence	38.98	39.97	39.47
Semeval 2022 task 12	Origin results	44.36	39.7	41.9
	Delete the natural logarithm e	42.73	35.46	38.76
	Not multiplied by hp(t)	40.82	32.6	36.25
	Insert inverse feature words into the original sentence	42.69	36.45	39.32

In another experiment, we directly incorporated inverse feature words into the original text computation to examine their impact on textual information. This modification also degraded performance, attributable to two primary factors:

Entity Information Conflict: Word vectors in the original text simultaneously facilitate feature word screening and entity recognition. When inverse feature words weaken vector representations, they interfere with boundary identification. Weakened entity information reduces entity recognition accuracy by 8~12% (SciERC ablation tests), cascading to relation extraction failures.

Contextual Interference: Introducing inverse feature words alters semantic contexts, amplifying ambiguity for polysemous words or context dependent relations. For example, in the phrase “cell growth inhibition”, weakening “inhibition” shifted relational interpretation probabilities by 25% (Semeval 2022 Task 12 analysis), inducing erroneous relation extractions.

Modifying the feature word loss function may simplify model training in certain scenarios. But in this task, removing the natural logarithm or altering position based supervision degrades optimization effectiveness. Without appropriate supervisory constraints, the model struggles to focus learning on specific word positions. Introducing inverse feature words directly into original text demonstrates that while they enhance feature word screening, they concurrently introduce noise, particularly detrimental in entity recognition tasks where interference disrupts precise boundary detection, causing performance decline. Ablation results empirically validate our original design’s efficacy, confirming that rational loss function engineering and precise supervision of feature/inverse feature words are critical to preserving model efficacy.

3.6. Analysis of the Hyperparameter τ

Since the hyperparameter τ is fixed to 5 in our model, meaning that five feature words are selected during the dynamic filtering process, we further investigate the impact of different τ values on prediction performance. In this subsection, we conduct a series of experiments on the SciERC dataset, where τ ranges from 1 to 10. Figure 2 illustrates the performance comparison under different numbers of selected feature words.

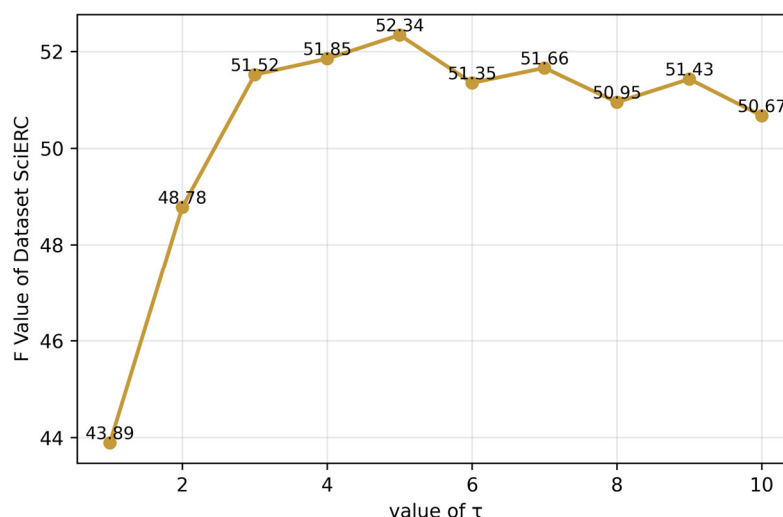


Figure 2. The trend of F-value changing with the hyperparameter τ . This result is used to verify our chosen value.

As shown in the results, the F-score gradually increases as τ grows from 1 to 5 and reaches its peak at $\tau = 5$. When τ exceeds 5, the performance begins to decline and exhibits certain fluctuations. The number of selected feature words reflects the extent of semantic features that need to be enhanced for a given sample. Since our extraction framework first enhances the sentence representation and then sequentially performs relation and entity extraction, selecting an excessive number of feature words may introduce redundant information, leading to semantic dispersion or even altering the original sentence meaning. Therefore, considering both performance and stability, τ is finally set to 5 in our experiments.

3.7. Error Analysis

To more intuitively illustrate the potential error patterns of FWAS and other baselines, we further present several representative prediction cases in Figure 3. In some scientific texts, the base model is susceptible to interference from background terminology or descriptive elements. In sentences describing experimental settings or related work, when multiple entities appear simultaneously but only a few are involved in the target relation, the base model tends to incorrectly match frequently occurring but semantically weak entities with relation trigger words, resulting in misjudgments. This error corresponds to the blue part in Figure 3.

Text	The anomalies are drawn uniformly from the hyper-cube defined by the range $(-4,4)$ in each dimension.				
Entity Pair	the range		$(-4,4)$	range	
Feature Words	defined	by	each	dimension	drawn
Error reason	"range" and "the range" form nested entities; they have identical semantics but different boundaries. This situation represents a high-risk boundary error scenario for FWAS and other baselines. So the first one failed to get "PRIMARY/Direct/SYMBOL" result.				
Text	For each data point x_i , the Random Forest computes an estimate of $P(\hat{y}_i=k x_i)$, the predicted probability that x_i belongs to class S_k . We construct a confusion matrix $C_{j,k}$ in which cell $C_{j,k}$ contains the sum of $P(\hat{y}_i=k x_i)$ for all x_i whose true class $y_i=j$. We then define a graph in which each node is a class and each edge (between two classes S_j and S_k) has a weight equal to $C[j,k]+C[k,j]$.				
Entity Pair	S_k		S_k		
Feature Words	each	edge	weight	equal	contains
Error reason	When the type of the annotated joint entity relation is "SYMBOL/Corefer-Symbol/SYMBOL", it is impossible to find the feature information representing coreference in the original text through the semantics of Corefer-Symbol.				

Figure 3. Error case analysis. This diagram illustrates two error cases, demonstrating the potential error scenarios that FWAS and baselines can generate. The case in the blue area demonstrates the problem of matching relationships and entities; The case in the purple area demonstrates the problem of highly implicit semantics and lack of obvious trigger words.

FWAS can mitigate the impact of such interference through an inverse feature word suppression mechanism, favoring entity combinations highly relevant to relation semantics. However, in scenarios where relation semantics are highly implicit and explicit trigger words are lacking, FWAS may still produce incorrect predictions. This error corresponds to the purple part in Figure 3. When relations require judgment based on cross-sentence reasoning or implicit semantic associations, even with effective noise suppression, the model may still fail to accurately identify the target relation due to a lack of explicit cues. These cases demonstrate that FWAS has a significant advantage in suppressing noise interference, but there is still room for improvement in complex scenarios requiring deep semantic reasoning.

3.8. Feature Word Statistical Analysis

To investigate the distribution properties of feature and inverse feature words, we conducted quantitative analysis of original sentence associated words, with results shown in Figure 4. This experiment statistically analyzed part-of-speech distributions for both word types, revealing their categorical preferences. Results demonstrate nouns dominate both feature and inverse feature words, exhibiting strong alignment with quintuple extraction tasks' inherent requirements.

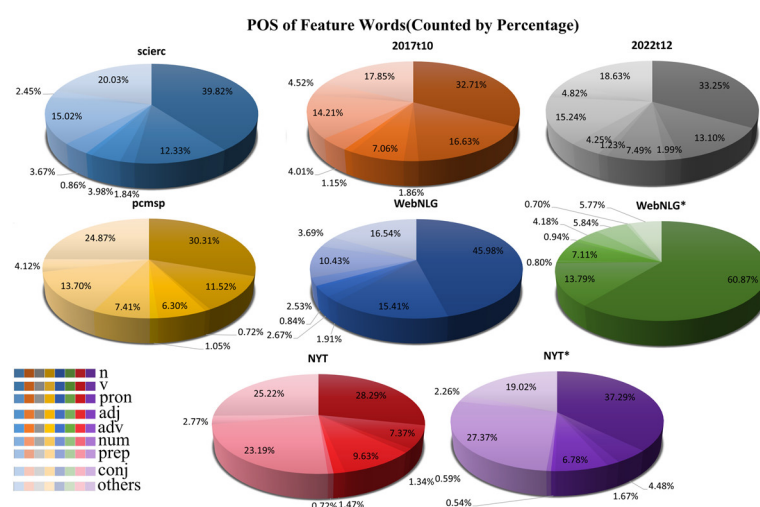


Figure 4. Part-of-speech distribution of original sentence-associated words (feature words). Data derived from all four datasets. Nouns constitute a significantly higher proportion than other parts of speech across all datasets.

Further analysis confirms nouns dominate both feature and inverse feature words significantly more than other parts of speech. This phenomenon fundamentally arises from our scheme's core objective: optimizing quintuple extraction through feature/inverse word enhancement. Since head/tail entities and relations primarily rely on noun information, feature/inverse words progressively shift toward noun dominance under entity/relation supervision signals during training. This adaptation strategically supports entity recognition and relation prediction, making nouns the predominant category. Critical relationship expressions (e.g., "author", "institution", "method", "experiment") represent quintuple essential elements typically expressed as nouns. Our model increasingly emphasizes such nouns as feature/inverse words during training, enabling precise capture of textual semantics.

Beyond nouns, experimental results reveal prepositions and other part-of-speech categories also occupy notable proportions. This indicates that, while nouns dominate feature and inverse feature words, other lexical classes still influence quintuple extraction, such as prepositions, punctuation marks, and numerals. The high proportion of prepositions is attributable to their critical role in relational expressions. For instance, prepositional phrases like "based on", "derived from", and "applied to" frequently signal entity relationships in scientific texts, directly impacting relation type determination in quintuples. What is more, the model autonomously learns these prepositions' functions in relation extraction, marking them as feature or inverse feature words to assist relational modeling.

We observe that numerical values, special symbols, punctuation marks, and other semantically incomplete elements also occupy notable proportions in feature and inverse feature word distributions. This occurs because certain relational information relies on numerical expressions (e.g., years, experimental parameters, probability values) or punctuation marks (e.g., colons, dashes). For instance, in academic texts like "experimental results: 95% confidence interval", the model must focus on content following the colon to accurately extract quintuple information. Thus, despite lacking full semantic integrity as standalone entities or relations, these elements remain functionally significant for relation extraction tasks.

These experimental results demonstrate that the distributions of feature and inverse feature words are intrinsically aligned with the model's learning objectives. Under supervision from quintuple annotations during training, the model progressively prioritizes word categories closely associated with entities and relations, ultimately forming a distribution hierarchy: nouns dominate, followed by prepositions and other categories. This confirms the scheme's capability to capture diverse semantic information types for enhanced joint entity-relation extraction. The feature inverse feature contrast strategy enables autonomous identification of relation critical word classes, leveraging this knowledge to strengthen quintuple extraction. Quantitative results validate that the selection mechanism effectively aligns with semantic requirements of entities and relations, providing robust support for the overall task.

4. Discussion

4.1. Computational Complexity Analysis

This section demonstrates the time feasibility of FWAS. Compared to the baseline joint entity relation extraction model, our proposed FWAS introduces an additional two-pointer computation component. It uses a concatenation iteration and projection mapping method during computation, resulting in a minimal increase in trainable parameters. In our implementation, the total number of parameters increases by only about 5% compared to the backbone model. Regarding training efficiency, FWAS incurs some overhead due to the pointer-based iterative feature selection process, increasing training time per epoch

by approximately 10–15%. However, this is largely unaffected during the inference phase. The pointer process during inference is executed only once per sentence, resulting in significantly lower overhead compared to training processes involving multiple parameter adjustments. Therefore, FWAS achieves higher extraction accuracy while maintaining similar inference efficiency.

4.2. Convergence and Scalability Analysis

In the FWAS framework, the number of pointers is fixed to two, corresponding to the feature word pointer and the inverse feature word pointer. Our empirical observations indicate that further increasing the number of pointers tends to introduce redundant modelling rather than significantly improving the model's convergence speed. In practice, the improvement in convergence efficiency of FWAS mainly originates from the competitive optimization mechanism itself. By explicitly identifying and suppressing inverse feature words, the model can more stably recognize the key information it truly relies on. This reduces the interference of noisy features during training and enables the model to reach a stable state more quickly.

Regarding scalability, FWAS is designed to operate at the sentence level and does not depend on document-level graph construction or global structural modelling. The proposed method can be efficiently applied to large-scale scientific corpora through sentence-level batch processing. For longer documents, sliding-window or paragraph-level segmentation strategies can be adopted without modifying the core model architecture. As sentence length increases, FWAS can instead exploit richer contextual information to obtain potential inverse feature word candidates, thereby more effectively distinguishing truly discriminative feature words from noise that should be suppressed. This mechanism alleviates, to some extent, the common issue of semantic dispersion and insufficient prominence of key information in long-sentence scenarios.

5. Conclusions

This study proposed a dual pointer mechanism based feature word screening method to address limitations in feature recognition and global semantic fusion within existing information extraction approaches. We designed a feature pointer for dynamically screening local feature words to enhance sentence semantic representations, while introducing an inverse feature pointer to impose negative weight decay that suppresses interference from non-feature words. Experimental evaluations demonstrate superior performance across multiple scientific literature datasets, particularly under the Rel+ scenario where significant improvements over baseline methods were achieved. Further quantitative analysis reveals that feature and inverse feature word distributions align with semantic requirements for entity-relation modeling, validating the dual pointer mechanism's effectiveness in information extraction tasks. Ablation experiments confirm the necessity of feature enhancement and interference suppression, indicating that joint supervision strategies improve model sensitivity to feature words and optimize global semantic representations. This method achieves notable improvements in both accuracy and robustness for information extraction, offering new perspectives for structuring scientific literature information.

In future work, we plan to extend the proposed framework to document-level scientific information extraction and explore its integration with large language models. In addition, we aim to investigate more efficient feature-pointer optimization strategies to further improve scalability and robustness.

Author Contributions: Conceptualization, Y.L. and L.F.; methodology, Y.L.; software, Y.L.; validation, Y.L.; formal analysis, L.F.; investigation, Y.L.; resources, X.X.; data curation, X.X.; writing—

original draft preparation, Y.L.; writing—review and editing, Y.L.; supervision, L.F. and X.X.; funding acquisition, L.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Social Science Fund of China (Grant No. 21BTQ06). No funding was received for the article processing charge (APC)..

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available at [SciERC] (accessed on 12 February 2026) <https://nlp.cs.washington.edu/sciE/>, [Semeval 2017 task 10] (accessed on 12 February 2026) <https://scienceie.github.io/>, [Semeval 2022 task 12] (accessed on 12 February 2026) <https://github.com/nlp-uoregon/symLink>, [NYT&NYT*] (accessed on 12 February 2026) <https://catalog.ldc.upenn.edu/LDC2008T19>, and [WebNLG&WebNLG*] (accessed on 12 February 2026) <https://synalp.gitlabpages.inria.fr/webnlg-challenge/>. The codes are available at https://github.com/LewisYJohnson/FWA_S.

Acknowledgments: The authors would like to express their sincere gratitude to Fu Lijun and Xia Xiaojun for providing valuable guidance, research support, and helpful suggestions for this study..

Conflicts of Interest: Authors Yujiang Liu, Lijun Fu, and Xiaojun Xia were employed by the company Shenyang Institute of Computing Technology Co., Ltd. Authors Liu Yujiang and Xia Xiaojun also engage in research work at the University of Chinese Academy of Sciences. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

References

1. Han, W.; Zhang, X.; Wang, Y.; Wang, L.; Huang, X.; Li, J.; & Wang, Y. A survey of machine learning and deep learning in remote sensing of geological environment: Challenges, advances, and opportunities. *ISPRS J. Photogramm. Remote Sens.* **2023**, *202*, 87–113.
2. Yu, T.; Yang, M.; Li, C.; & Xu, R. Reducing spurious correlations for relation extraction by feature decomposition and semantic augmentation. In Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval: Taipei, Taiwan; 2023; pp. 2324–2328.
3. Ye, B.; Chang, H.; Ma, B.; Shan, S.; & Chen, X. Joint feature learning and relation modeling for tracking: A one-stream framework. In *Proceedings of the European Conference on Computer Vision*; Springer: Cham, Switzerland, 2022; pp. 341–357.
4. Yin, H.; Zhong, J.; Li, R.; & Li, X. Disentangled relational graph neural network with contrastive learning for knowledge graph completion. *Knowl.-Based Syst.* **2024**, *295*, 111828.
5. Tang, W.; Su, F.; Sun, H.; Qi, Q.; Wang, J.; Tao, S.; & Yang, H. Weakly supervised entity alignment with positional inspiration. In Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining; Association for Computing Machinery: Singapore City, Singapore; 2023; pp. 814–822.
6. Liu, X.; Hong, H.; Wang, X.; Chen, Z.; Kharlamov, E.; Dong, Y.; & Tang, J. SelfKG: Self-supervised entity alignment in knowledge graphs. In Proceedings of the ACM Web Conference 2022, Virtual Event / Lyon, France; 2022; pp. 860–870.
7. Jia, W.; Ma, R.; Yan, L.; Niu, W.; & Ma, Z. Joint entity and relation extraction with table filling based on graph convolutional networks. *Expert Syst. Appl.* **2025**, *266*, 126130.
8. Qin, H.; Tian, Y.; Song, Y. Enhancing relation extraction via adversarial multi-task learning. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference: Marseille, France*; 2022; pp. 6190–6199.
9. Hao, K.; Yu, B.; Hu, W. Knowing false negatives: An adversarial training method for distantly supervised relation extraction. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: Online and Punta Cana, Dominican Republic*; 2021; pp. 9661–9672.
10. Zhang, M.; Qian, T.; Zhang, T.; & Miao, X.. Towards model robustness: Generating contextual counterfactuals for entities in relation extraction. In *Proceedings of the ACM Web Conference 2023: Austin, Texas, USA*; 2023; pp. 1832–1842.
11. Yang, W.; Qin, Y.; Huang, R.; & Chen, Y. Adaptive feature extraction for entity relation extraction. *Comput. Speech Lang.* **2025**, *89*, 101712.

12. Zhang, D.; Zhang, H.; Tang, J.; Hua, X. S.; & Sun, Q. Causal intervention for weakly supervised semantic segmentation. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 655–666.
13. Kang, M.; Han, B. Operation-aware soft channel pruning using differentiable masks. In *Proceedings of the International Conference on Machine Learning*; PMLR: New York, NY, USA, 2020; pp. 5122–5131.
14. Li, J.; Shi, K.; Xie, G.S.; Liu, X.; Zhang, J.; & Zhou, T. Label-efficient few-shot semantic segmentation with unsupervised meta-training. In *Proceedings of the AAAI Conference on Artificial Intelligence 2024*: Vancouver, Canada; Volume 38, pp. 3109–3117.
15. Wang, T.; Yang, W.; Wu, T.; Yang, C.; Liang, J.; Wang, H.; & Zhou, Z.. Joint entity and relation extraction with fusion of multi-feature semantics. *J. Intell. Inf. Syst.* **2025**, *63*, 21–42.
16. Mrabah, N.; Bouguessa, M.; Ksantini, R. Adversarial deep embedded clustering: On a better trade-off between feature randomness and feature drift. *IEEE Trans. Knowl. Data Eng.* **2022**, *34*, 1603–1617.
17. Zhang, M.; Lu, J.; Yang, J.; Zhou, J.; Wan, M.; & Zhang, X. From coarse to fine: Enhancing multi-document summarization with multi-granularity relationship-based extractor. *Inf. Process. Manag.* **2024**, *61*, 103696.
18. Qiao, D.; Song, B.; Yang, X.; Fan, K.; & Zhang, Z. An improved entity-relationship joint extraction model based on cascade pointer network. *Int. Core J. Eng.* **2023**, *9*, 4–15.
19. Bai, C.; Pan, L.; Luo, S.; & Wu, Z. Joint extraction of entities and relations by a novel end-to-end model with a double-pointer module. *Neurocomputing* **2020**, *377*, 325–333.
20. Park, S.; Kim, H. Dual pointer network for fast extraction of multiple relations in a sentence. *Appl. Sci.* **2020**, *10*, 3851.
21. Li, D.; Hogan, W.; Shang, J. READ: Improving relation extraction from an adversarial perspective. In *Findings of the Association for Computational Linguistics: NAACL 2024*; Association for Computational Linguistics: Mexico City, Mexico, 2024; pp. 533–548.
22. Zhao, X.; Deng, Y.; Yang, M.; Wang, L.; Zhang, R.; Cheng, H.; & Xu, R. A comprehensive survey on relation extraction: Recent advances and new frontiers. *ACM Comput. Surv.* **2024**, *56*, 1–39.
23. Wang, P.; Song, Y.; Liu, T.; Lin, B.; Cao, Y.; Li, S.; & Sui, Z. Learning robust representations for continual relation extraction via adversarial class augmentation. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing: Abu Dhabi, Dubai*; 2022; pp. 6264–6278.
24. Lv, P.; Zhang, Q.; Wu, M.; Deveci, M.; Shi, L.; Sun, Y.; & Wang, J. Intelligent extraction of medical entity relationship based on graph neural network and optimization strategy. *Knowl.-Based Syst.* **2024**, *294*, 111735.
25. Augenstein, I.; Das, M.; Riedel, S.; Vikraman, L.; & McCallum, A. SemEval 2017 Task 10: ScienceIE—Extracting keyphrases and relations from scientific publications. In *Proceedings of the 11th International Workshop on Semantic Evaluation: Vancouver, Canada*; 2017; pp. 546–555.
26. Lai, V.; Veyseh, A.P.B.; Dernoncourt, F.; & Nguyen, T. SemEval 2022 Task 12: Symlink—Linking mathematical symbols to their descriptions. In *Proceedings of the 16th International Workshop on Semantic Evaluation*; 2022; Seattle, WA, USA; pp. 1671–1678.
27. Riedel, S.; Yao, L.; McCallum, A. Modeling relations and their mentions without labeled text. In *Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases*; Springer: Berlin, Germany, 2010; pp. 148–163.
28. Gardent, C.; Shimorina, A.; Narayan, S.; & Laura, P. B. Creating training corpora for NLG micro-planning. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics: Vancouver, Canada*; 2017; pp. 179–188.
29. Santosh, T.Y.S.S.; Chakraborty, P.; Dutta, S.; Sanyal, D. K.; & Das, P. P. Joint entity and relation extraction from scientific documents: Role of linguistic information and entity types. *EEKE@JCDL* **2021**, *21*, 15–19.
30. Ren, L.; Liu, Y.; Cao, Y.; & Ouyang, C. CoVariance-based causal debiasing for entity and relation extraction. In *Findings of the Association for Computational Linguistics: EMNLP 2023*: Singapore City, Singapore; 2023; pp. 2627–2640.
31. Eberts, M.; Ulges, A. Span-based joint entity and relation extraction with transformer pre-training. In *Proceedings of ECAI 2020*; IOS Press: Amsterdam, Netherlands, 2020; pp. 2006–2013.
32. Shen, Y.; Ma, X.; Tang, Y.; & Lu, W. A trigger-sense memory flow framework for joint entity and relation extraction. In *Proceedings of the Web Conference 2021*: New York, NY, USA; pp. 1704–1715.
33. Wu, Y.; Chen, Y.; Qin, Y.; Huang, R.; & Tang, R. A marker collaborating model for entity and relation extraction. *J. King Saud Univ.-Comput. Inf. Sci.* **2022**, *34*, 9163–9172.
34. Pham, D.; Ho, X.; Ha, Q.T.; & Aizawa, A. Solving label variation in scientific information extraction via multi-task learning. In *Proceedings of the 37th Pacific Asia Conference on Language, Information and Computation: Hong Kong, China*; 2023; pp. 243–256.
35. Yan, Z.; Yang, S.; Liu, W.; & Tu, K. Joint entity and relation extraction with span pruning and hypergraph neural networks. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*: Singapore City, Singapore; 2023; pp. 7512–7526.

36. Zaratiana, U.; Tomeh, N.; Holat, P.; & Charnois, T. An autoregressive text-to-graph framework for joint entity and relation extraction. In *Proceedings of the AAAI Conference on Artificial Intelligence 2024: Vancouver, Canada*; Volume 38, pp. 19477–19487.
37. Lou, J.; Lu, Y.; Dai, D.; Jia, W.; Lin, H.; Han, X.; & Wu, H. Universal information extraction as unified semantic matching. In *Proceedings of the AAAI Conference on Artificial Intelligence 2023: Washington, DC, USA*; Volume 37, pp. 13318–13326.
38. Zaratiana, U.; Tomeh, N.; El Khbir, N.; Holat, P.; & Charnois, T. GraphER: A structure-aware text-to-graph model for entity and relation extraction. *arXiv* **2024**, arXiv:2404.12491.
39. Zhou, Y.; Shan, S.; Wei, H.; Zhao, Z.; & Feng, W. PGA-SciRE: Harnessing LLM on data augmentation for enhancing scientific relation extraction. In *Proceedings of the 23rd Chinese National Conference on Computational Linguistics: Taiyuan, Shanxi, China; 2024*; pp. 352–369.
40. Wang, Y.; Yu, B.; Liu, Y.; & Lu, S. True-UIE: Two universal relations unify information extraction tasks. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies: Mexico City, Mexico; 2024*; pp. 1863–1876.
41. Cong, X.; Yu, B.; Fang, M.; Liu, T.; Yu, H.; Hu, Z.; & Wang, B. Universal information extraction with meta-pretrained self-retrieval. In *Findings of the Association for Computational Linguistics: ACL 2023: Toronto, ON, Canada, 2023*; pp. 4084–4100.
42. Gui, H.; Yuan, L.; Ye, H.; Zhang, N.; Sun, M.; Liang, L.; & Chen, H. IEPile: Unearthing large-scale schema-conditioned information extraction corpus. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics: Bangkok, Thailand; 2024*; pp. 127–146.
43. Li, J.; Zhang, Y.; Liang, B.; Wong, K. F.; & Xu, R. Set learning for generative information extraction. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Singapore City, Singapore; 2023*; pp. 13043–13052.
44. Yan, H.; Sun, Y.; Li, X.; Zhou, Y.; Huang, X. J.; & Qiu, X. UTC-IE: A unified token-pair classification architecture for information extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics; 2023: Toronto, Canada*; pp. 4096–4122.
45. Yi, X.; Bao, Y.; Zhang, J.; Qin, Y.; & Lin, F. Integrating structural semantic knowledge for enhanced information extraction pre-training. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Miami, Florida, USA; 2024*; pp. 2156–2171.
46. Chen, K.; Zhou, J.; Chen, Q.; Liu, S.; & He, L. A regularization-based transfer learning method for information extraction via instructed graph decoder. In *Proceedings of the LREC-COLING 2024: Torino, Italia; 2024*; pp. 1472–1485.
47. Kong, W.; Xia, Y. CARE: Co-attention network for joint entity and relation extraction. In *Proceedings of the LREC-COLING 2024: Torino, Italia; 2024*; pp. 2864–2870.
48. Dai, Q.; Yang, W.; Wei, F.; He, L.; & Liao, Y. FSN: Joint entity and relation extraction based on filter separator network. *Entropy* **2024**, *26*, 162.
49. Luan, Y.; He, L.; Ostendorf, M.; et al. Multi-task identification of entities, relations, and coreference for scientific knowledge graph construction. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing; 2018*; pp. 3219–3232.
50. Luan, Y.; Wadden, D.; He, L.; Shah, A.; Ostendorf, M.; & Hajishirzi, H. A general framework for information extraction using dynamic span graphs. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Minneapolis, Minnesota; 2019*; pp. 3036–3046.
51. Popovic, N.; Laurito, W.; Färber, M. AIFB-WebScience at SemEval-2022 Task 12: Relation extraction first. In *Proceedings of the 16th International Workshop on Semantic Evaluation: Seattle, WA, USA; 2022*; pp. 1687–1694.
52. Lee, S.M.; Na, S.H. Jbnu-cclab at SemEval-2022 Task 12: Machine reading comprehension and span pair classification. In *Proceedings of the 16th International Workshop on Semantic Evaluation: Seattle, WA, USA; 2022*; pp. 1679–1686.
53. Xu, B.; Wang, Q.; Lyu, Y.; Shi, Y.; Zhu, Y.; Gao, J.; & Mao, Z. EmRel: Joint representation of entities and embedded relations for multi-triple extraction. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Seattle, WA, USA; 2022*; pp. 659–665.
54. Ren, F.; Zhang, L.; Yin, S.; Zhao, X.; Liu, S.; Li, B.; & Liu, Y. A novel global feature-oriented relational triple extraction model based on table filling. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: Online and Punta Cana, Dominican Republic; 2021*; pp. 2646–2656.
55. Ren, F.; Zhang, L.; Yin, S.; Zhao, X.; Liu, S.; & Li, B. A conditional cascade model for relational triple extraction. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management: New York, NY, United States; 2021*; pp. 3393–3397.

56. Ren, F.; Zhang, L.; Zhao, X.; Yin, S.; Liu, S.; & Li, B. A simple but effective bidirectional framework for relational triple extraction. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining: New York, NY, United States*; 2022; pp. 824–832.
57. Wei, Z.; Su, J.; Wang, Y.; Tian, Y.; & Chang, Y. A novel cascade binary tagging framework for relational triple extraction. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Online*; 2020; pp. 1476–1488.
58. Wang, Y.; Yu, B.; Zhang, Y.; Liu, T.; Zhu, H.; & Sun, L. TPLinker: Single-stage joint extraction of entities and relations through token pair linking. In *Proceedings of the 28th International Conference on Computational Linguistics: Barcelona, Spain (Online)*; 2020; pp. 1572–1582.
59. Zheng, H.; Wen, R.; Chen, X.; Yang, Y.; Zhang, Y.; Zhang, Z.; & Zheng, Y. PRGC: Potential relation and global correspondence based joint relational triple extraction. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics: Online*; 2021; pp. 6225–6235.
60. Li, Z.; Fu, L.; Wang, X.; Zhang, H.; & Zhou, C. RFBFN: A relation-first blank filling network for joint relational triple extraction. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop: Dublin, Ireland*; 2022; pp. 10–20.
61. Sui, D.; Zeng, X.; Chen, Y.; Liu, K.; & Zhao, J. Joint entity and relation extraction with set prediction networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *35*, 12784–12795.
62. Zhang, Z.; Hu, X.; Zhang, H.; & Liu, J. NEDORT: A novel and efficient approach to the data overlap problem in relational triples. *Complex Intell. Syst.* **2023**, *9*, 5235–5250.
63. Xiao, Y.; Chen, G.; Du, C.; Li, L.; Yuan, Y.; Zou, J.; & Liu, J. A study on double-headed entities and relations prediction framework for joint triple extraction. *Mathematics* **2023**, *11*, 4583.
64. Ning, J.; Yang, Z.; Sun, Y.; Wang, Z.; & Lin, H. OD-RTE: A one-stage object detection framework for relational triple extraction. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics: Toronto, Canada*; 2023; pp. 11120–11135.
65. Liu, S.; Lyu, W.; Ma, X.; & Ge, J. An entity–relation joint extraction method based on two independent sub-modules from unstructured text. *IEEE Access* **2023**, *11*, 122154–122163.
66. Li, R.; La, K.; Lei, J.; Huang, L.; Ouyang, J.; Shu, Y.; & Yang, S. Joint extraction model of entity relations based on decomposition strategy. *Sci. Rep.* **2024**, *14*, 1786.
67. Dong, H.; Wang, W.; Sun, Z.; Kang, Z.; Ge, X.; Gao, F.; & Wang, J. Knowledge graph construction for intelligent cockpits based on large language models. *Sci. Rep.* **2025**, *15*, 7635.
68. Chen, G.; Chen, P.; Li, H.; Wang, X.; Zhou, X.; Yu, A.; & Wang, Q. Relation semantic guidance and entity position location for relation extraction. *Data Sci. Eng.* **2025**, *10*, 175–195.
69. Qi, H.; Liu, W.; Liu, F.; Zhu, W.; & Shan, F. Entity relation joint extraction method based on insertion transformers. *Int. J. Adv. Comput. Sci. Appl.* **2024**, *15*, 0150467.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.