Cross-Domain Few-Shot Agricultural Pest and Disease Named Entity Recognition Based on a Pseudo-Label Filtering Self-Training Decomposition Framework

Fengshun Wei[a], Shengquan Liu[†a], Bo Kong[a], Liruizhi Jia[a], Shouhao Yu[a], Lu Wang[a]

[a]*School of Computer Science and Technology, Xinjiang University, Urumqi, China*

Abstract

In cross-domain few-shot named entity recognition tasks, although decomposition framework-based methods have achieved certain successes, they still face challenges related to domain adaptation bias. Fine-tuning with limited labeled samples has inherent limitations in adapting to target domain feature distributions, and traditional pseudo-labeling methods, which mainly rely on model confidence, have not yet fully leveraged the guiding value of the few-shot samples. To address this issue, we design a dual-stage pseudo-label filtering mechanism to enhance the guiding capacity of few-shot samples. We first conduct preliminary selection using a confidence threshold and then use a small set of labeled target domain data as a quality reference to calculate semantic matching scores between pseudo-entities and annotated samples. A minimum matching score standard is established to further filter the pseudo-labels. This mechanism enhances the feature patterns of the limited labeled target domain data with high-quality pseudo-labels, improving the guiding ability of few-shot samples and promoting model adaptation to the target domain feature distribution. We evaluate our method on English public datasets and Chinese agricultural datasets (wheat, cotton, and publicly available rice), conducting cross-domain experiments with rice and wheat as source domains and cotton as the target to validate its effectiveness in real-world transfer scenarios.Experimental results show that our method consistently outperforms the best baseline approaches.

Keywords: Cross-domain few-shot; Named entity recognition; Decomposition framework; Pseudo-label self-training; Agricultural pest and disease domain application;

1. Introduction

Named Entity Recognition (NER) is one of the fundamental tasks in natural language processing, playing a critical role in applications such as information extraction and knowledge graph construction[1]. Traditional NER methods typically rely on large amounts of annotated data for training[2]. However, in specialized domains such as cotton pest and disease management, obtaining sufficient labeled data is often costly and time-consuming. Cross-domain few-shot learning in transfer learning emerges as a solution, aiming to leverage the rich knowledge from the source domain and a few samples from the target domain to achieve effective entity recognition[3].

†Corresponding author: Prof. Shengquan Liu(liu@xju.edu.cn); ORCID:0000-0001-9623-4714)

End-to-end metric learning-based approaches [4,5] have become mainstream for few-shot NER. These methods require the simultaneous learning of complex structures composed of entity boundaries and entity types. When the domain gap is large, it is challenging for models to capture such complex structural information based only on a few domain-adaptive support samples, resulting in significant performance degradation. Consequently, these methods often suffer from insufficient boundary learning and misclassify sequences of words into incorrect entity categories.

Recently, there has been a growing trend toward adopting decomposition frameworks for named entity recognition [6,7], which decompose NER into two subtasks: entity span detection and span classification, with each stage addressing a single task. Since decomposition methods only handle the boundary detection task in the first phase, they can identify more precise entity boundaries and better leverage source domain knowledge for target domain span type classification compared to end-to-end methods, thereby achieving superior performance.

However, in the cross-domain few-shot NER setting, models constructed by such methods still face the challenge of domain adaptation bias. When linguistic expressions and entity characteristics differ between the source and target domains, fine-tuning with only a few target domain samples results in limited model capability to understand the feature distribution of the target domain, thereby restricting feature guidance capability. Specifically, the model tends to detect many erroneous source domain entity spans and struggles to accurately capture the semantic boundaries and type characteristics of target domain entities.

To address the above challenges, we propose a decomposition framework for named entity recognition based on dual-stage pseudo-label filtering self-training, termed Pseudo-label Filtering Self-Training for Decomposed Named Entity Recognition (PFSTDNER). The proposed method extends the guidance capability of few-shot samples through two key steps: first, generating initial pseudo-labels on unlabeled target domain data using the few-shot fine-tuned model, followed by preliminary filtering through a confidence threshold; second, using a few labeled samples as quality reference standards to compute the semantic matching scores between pseudo-entities and few-shot samples, establishing a minimum matching score threshold to further filter out low-quality pseudo-labels. This mechanism effectively expands the coverage of target domain feature patterns, enhances the guidance capability of few-shot samples, and promotes the model's adaptation to the target domain feature distribution.

We evaluate our method on both English public datasets and on Chinese agricultural pest and disease datasets, where we construct wheat and cotton datasets and incorporate a publicly available rice dataset. We design cross-domain experiments using rice and wheat as source domains and cotton as the target domain, verifying the effectiveness and practical value of our approach in real-world domain transfer scenarios.Experimental results demonstrate that the model incorporating this mechanism outperforms optimal baseline methods across various cross-domain few-shot datasets, and validates the practical application value of the proposed model in the Chinese agricultural pest and disease subdomain migration experiments.

2. Related Work

2.1. Few-Shot Named Entity Recognition

Few-shot NER methods can be broadly categorized into two paradigms: metric-based and prompt-based approaches. Metric-based methods focus on learning a feature space with strong generalization capability and classify spans based on nearest prototypes or neighboring samples[8,9,10,11,5]. Prompt-based methods leverage pre-trained language models via prompt engineering to better adapt to new entity types with minimal labeled data[12,13].

2.2. Decomposition Frameworks

Recent work decomposes NER into two subtasks: span detection and type classification, improving boundary detection and leveraging source domain knowledge[14,15,16,10,6]. This two-stage structure has shown superior performance in few-shot settings compared to end-to-end methods.

2.3. Self-Training

Self-training, a classic semi-supervised learning technique[17], has been successfully applied in vision[18,19] and NLP[20,21]. Recent work demonstrates that stronger data augmentation and larger unlabeled corpora can boost generalization. Our work is the first to integrate pseudo-label self-training into a decomposition-based few-shot NER framework, with targeted filtering mechanisms to address domain adaptation bias.

3. Problem Formulation

3.1. Named Entity Recognition Task

Given a token sequence $X = (x_1, x_2, \ldots, x_n)$ where $x_i \in \mathcal{V}$, the goal of NER is to predict entity spans $s = (b, e, c)$, where b and e are the start and end token positions and c is the entity type from a set C.

3.2. Cross-Domain Few-Shot Setting

We assume access to:

- A large fully annotated source domain dataset D_s.
- A few labeled target domain examples D_t^{ℓ} (k-shot, typically $k = 1$ or 5).
- A large set of unlabeled target domain data D_t^u.

The goal is to learn a model $\mathcal{M}_\phi$ that accurately detects and classifies named entities in a target domain test set D_t^{test}, under two main challenges:

- **Domain Shift**: $P_s(X, Y) \neq P_t(X, Y)$ due to differences in style and entity distribution.
- **Annotation Scarcity**: Only a very small number of labeled target examples are available.

To address these challenges, we propose a decomposition-based few-shot NER framework augmented with dual-stage pseudo-label filtering self-training. The model first detects spans in unlabeled data and then classifies them via a few-shot prototypical network, enhancing adaptation to the target domain.

4. The Proposed Approach

As illustrated in Figure 1, the proposed PFSTDNER follows the idea of existing cross-domain few-shot methods, decomposing the entity recognition process into two stages: entity span detection and entity type classification. The overall construction framework of PFSTDNER consists of three main components: (a) source domain training phase (Figure 1(a)), (b) target domain span detection phase (Figure 1(b)), and (c) target domain type classification phase (Figure 1(c)).

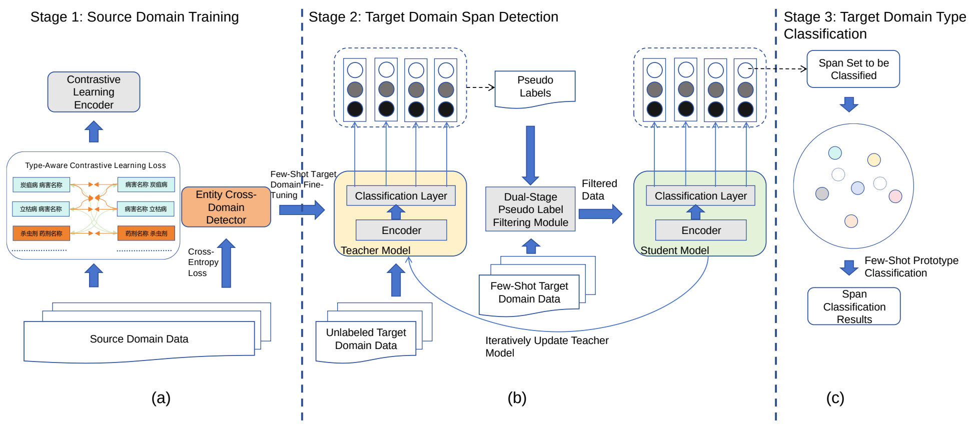

Figure 1: Framework of PFSTDNER Model Construction

In the source domain training phase, the model is trained on the abundant source domain data within a decomposition framework to learn an entity span detector and a contrastive learning encoder. In the target domain span detection phase, the model expands the guiding capacity of few-shot samples through a dual-stage pseudo-label filtering strategy. In the target domain type classification phase, the model assigns types to detected entity spans based on a prototypical network.

Compared with existing methods, the main characteristics of the proposed PFSTDNER are:

To address the limited guiding capability of few-shot samples, in the target domain span detection phase, we first introduce a traditional pseudo-label self-training approach to augment the training data. Furthermore, we innovatively design a dual-stage filtering mechanism within the pseudo-label self-training process. This mechanism more effectively utilizes the limited labeled target domain samples as a quality reference standard by computing matching scores between candidate spans and few-shot entity span words, establishing a pseudo-label filtering mechanism based on the minimum score criterion, thus further enhancing the guiding effect of few-shot samples. Additionally, we design an iterative pseudo-label self-training strategy, allowing the few labeled samples to continuously guide the model in gradually adapting to the target domain feature distribution.

4.1. Source Domain Training

The source domain training for the decomposed NER framework involves two stages: span detection and type classification. As shown in Figure 1(a), the objective is to use the source domain data to learn an entity span detector and an entity span classifier, providing two task-adapted models for subsequent few-shot learning in the target domain. For model selection, we adopt different base encoders depending on the language and domain: in the English public domain, we follow previous settings and use the bert-base-uncased model; in the Chinese agricultural pest and disease domain, we use the chinese-roberta-wwm-ext model as the base encoder.

We pre-train a span detector and contrastive learning encoder on D_s. The encoder learns type-aware representations [6].

4.2. Target Domain Span Detection

4.2.1. Few-Shot Fine-Tuning on Target Domain

As shown in Figure 1(b), after fine-tuning with a small number of labeled target domain samples, the model's adaptability to the target domain is initially improved. However, due to the limited number of labeled samples, the model's understanding of the target domain feature distribution remains incomplete, resulting in restricted feature-guiding capability. To alleviate this issue, we introduce a pseudo-label self-training approach in the target domain span detection phase to augment the training data and enhance the guidance capability of the few-shot samples.

4.2.2. Pseudo-Label Self-Training Framework

As illustrated in Figure 1(b), we adopt an iterative pseudo-label self-training framework, where the model fine-tuned on a few target domain samples generates pseudo-labels for the unlabeled data, thereby continuously optimizing model performance. In each iteration, the current model serves as the teacher to predict pseudo-labels for the unlabeled target domain data. A specific filtering strategy is applied to select high-quality pseudo-labels, and a new student model is then trained using the filtered pseudo-labels together with the few labeled samples. Upon completion of training, the student model replaces the teacher model and proceeds to the next iteration. The final trained student model is used to detect entity words in the target domain test sentences, where consecutive entity words are grouped as candidate spans, forming a candidate span set for subsequent type classification.

Formally, the iterative pseudo-label self-training framework is defined as follows:

Let:

- D_s: Labeled source domain data
- D_t^l: Labeled few-shot target domain data
- D_t^u: Unlabeled target domain data
- $\mathcal{M}_\phi$: Model with parameters ϕ

The iterative self-training process proceeds as:

Initialization: The initial model $\mathcal{M}_{\phi_0}$ is trained on D_s and fine-tuned on D_t^l to obtain $\mathcal{M}_{\phi_1}$.

Iteration: For each iteration k:

1. Use $\mathcal{M}_{\phi_k}$ to generate pseudo-labels for D_t^u
2. Apply the pseudo-label filtering strategy to select high-quality pseudo-labels
3. Train a new model $\mathcal{M}_{\phi_{k+1}}$ using D_t^l and the filtered pseudo-label data
4. Repeat the above steps until the preset number of iterations is reached

4.2.3. Pseudo-Label Filtering Strategy

In cross-domain few-shot learning scenarios, even after fine-tuning on a few target domain samples, models tend to recognize entity patterns from the source domain, which introduces source domain noise into pseudo-label self-training and hinders target domain adaptation. To address this, we design a dual-stage filtering mechanism to select high-quality pseudo-labels, consisting of token confidence filtering and span minimum score filtering.

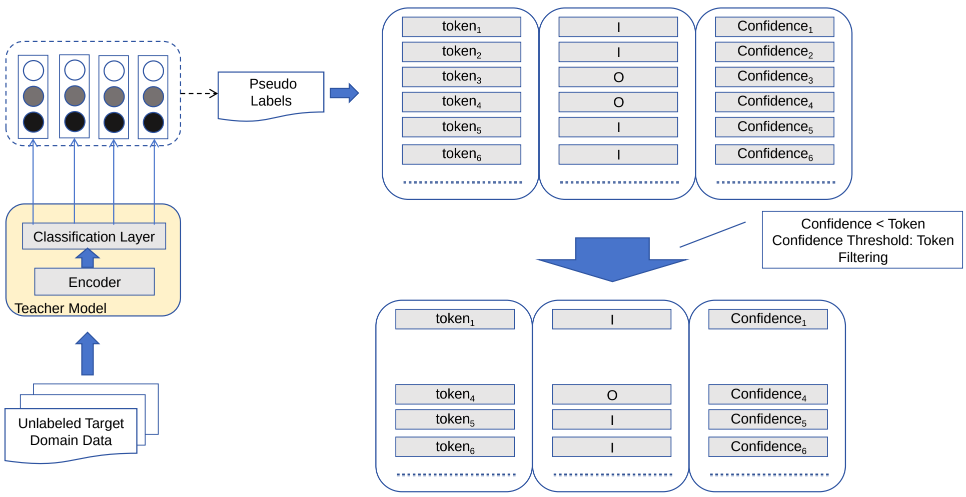

Figure 2: Token Confidence Filtering Module

First Stage: Token Confidence Filtering

In the first stage, we use token-level confidence scores for preliminary filtering. For each token, the model computes the probability distribution over possible labels, and the highest probability value is taken as the token's confidence. Tokens with confidence values exceeding a preset threshold are retained for further processing. This ensures that only tokens with highly confident predictions are preserved, filtering out uncertain ones.

Figure 2 shows the token confidence filtering module, which compares each token's predicted confidence against a preset threshold to select high-confidence tokens.

Formally, for each token t_i in the input sequence, the confidence is defined as:

$$C(t_i) = \max_{j \in \mathcal{Y}} P(y_i = j | t_i) \tag{1}$$

The token retention condition is:

$$Retain(t_i) = \mathbb{1}[C(t_i) > \tau] \tag{2}$$

where τ is the preset confidence threshold, and $\mathbb{1}[\cdot]$ denotes the indicator function.

After token confidence filtering, tokens labeled as entities are extracted and consecutive entity tokens are grouped into spans to form the candidate span set.

Second Stage: Span Minimum Score Filtering

In the second stage, we further filter the candidate spans composed of consecutive entity tokens from the first stage. Considering that single-stage confidence filtering may not sufficiently address semantic shifts caused by domain differences, we perform few-shot matching-based pseudo-label filtering during this phase.

Figure 3 illustrates the workflow of the span minimum matching score filtering module. This module leverages a small number of labeled target domain samples as a quality reference standard to further select high-quality pseudo-labels. Specifically, the module first uses the fine-tuned

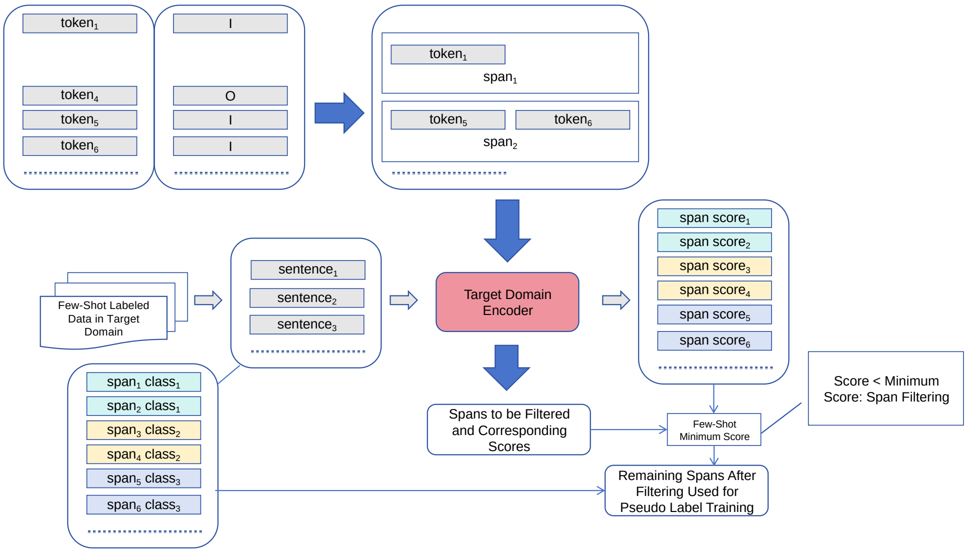

Figure 3: Span Minimum Matching Score Filtering Module

target domain encoder to process all entity spans and their corresponding ground-truth types from the few-shot samples, calculating their semantic matching scores.

Formally, the matching score for each entity span and its true label is calculated as:

$$score_{true}(e) = \mathrm{sim}(f_\theta(e), f_\theta(c_{true}(e))) \tag{3}$$

where e denotes the entity span, $c_{true}(e)$ is its corresponding true label, $f_\theta(\cdot)$ represents the target domain encoder, and $\mathrm{sim}(\cdot,\cdot)$ denotes the semantic similarity function calculated via dot product.

The minimum matching score among these spans is determined as the quality threshold:

$$\theta_{min} = \min_{e \in E} score_{true}(e) \tag{4}$$

For each candidate span $s \in \hat{S}_t^u$ obtained from the first stage, we compute its semantic similarity scores with all possible type labels. The predicted matching score is defined as:

$$score_{pred}(s) = \max_{c \in C} \; f_\theta(s)^\top f_\theta(c) \tag{5}$$

where $f_\theta(s)$ and $f_\theta(c)$ denote the encoded representations of the span s and the type label c, respectively, obtained from the target domain encoder. Here, the similarity function $\mathrm{sim}(\cdot,\cdot)$ is defined as the dot product between the two vector representations.

Spans are retained based on the following condition: a candidate span is preserved only if its matching score is greater than or equal to θ_{min}:

$$\hat{S}_t^{u,\text{filtered}} = \{s \in \hat{S}_t^u \mid score_{pred}(s) \geq \theta_{min}\} \tag{6}$$

where $\hat{S}_t^u$ is the set of candidate spans after the first stage, and $\hat{S}_t^{u,\text{filtered}}$ is the set of high-quality candidate spans after the second stage.

The final high-quality pseudo-label set is composed of the O-label tokens filtered in the first stage and the candidate entity spans filtered in the second stage:

$$\hat{S}_t^{\text{final}} = T_O \cup \hat{S}_t^{u,\text{filtered}} \tag{7}$$

where T_O denotes the set of high-confidence O-label tokens.

Finally, the filtered high-quality pseudo-label data, combined with the labeled few-shot target domain data, is used to train the student model:

$$\phi_{k+1} = \arg\min_{\phi} \mathcal{L}(\mathcal{M}_\phi; D_t^l \cup \hat{S}_t^{\text{final}}) \tag{8}$$

Through this dual-stage filtering mechanism and iterative self-training framework, the model can continuously enhance its adaptability to the target domain features, thereby improving the guidance effect of the few-shot samples.

4.3. Target Domain Type Classification

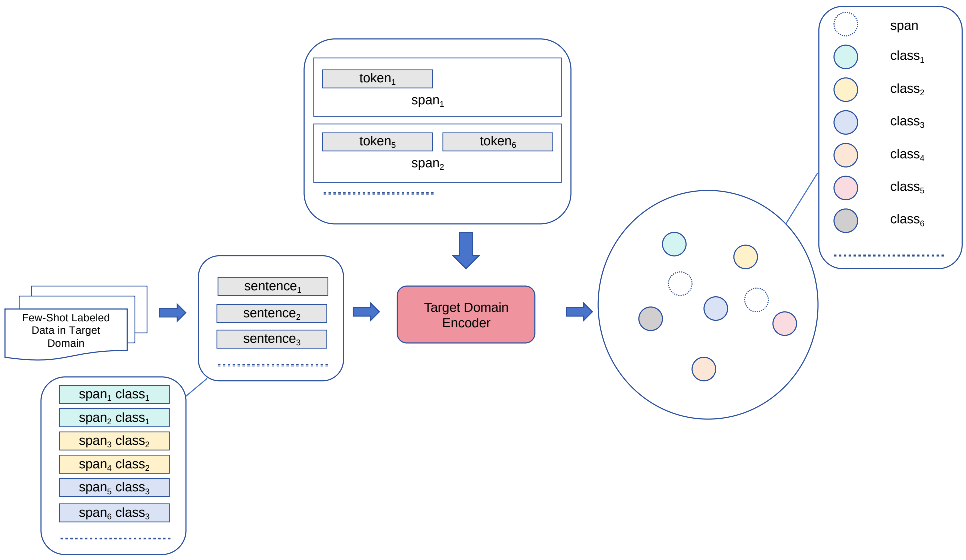

Figure 4: Few-Shot Prototypical Network Classification

As shown in Figure 4, PFSTDNER follows TadNER [6] to perform type classification using a prototypical network. Each entity type is represented by a prototype vector, constructed by aggregating span and label embeddings from few-shot samples.

During classification, candidate spans are matched against these prototypes based on similarity scores, and the type with the highest score is assigned. This approach enables stable and effective type prediction in few-shot settings.

5. Experiment

We evaluate the effectiveness of PFSTDNER in addressing the domain adaptation bias problem in cross-domain few-shot named entity recognition through three groups of experiments: (1) cross-domain few-shot experiments on English public datasets; (2) ablation experiments to assess the contributions of our proposed method; and (3) cross-domain few-shot experiments in the Chinese agricultural pest and disease subdomain.

The experiments aim to verify the effectiveness of the proposed method and the designed dual-stage pseudo-label filtering strategy in enhancing the guidance capability of few-shot samples and improving the model's adaptation to the target domain feature distribution.

5.1. Datasets and Experimental Setup

5.1.1. Cross-Domain Few-Shot Experimental

Following the setup proposed by Li et al.[6], we conduct domain transfer experiments using data collected from different textual domains. We use OntoNotes[22] as the source domain and evaluate few-shot performance on CoNLL[23], WNUT[24], and GUM[25] datasets.The baseline methods include one-stage methods (ProtoBERT[8], NNShot[11], StructShot[11], FSLS[26], CONTaiNER[5]) and two-stage methods (DecomposedMetaNER[10], TadNER[6]).

Additionally, to further validate the practical value of cross-domain few-shot methods, we conduct cross-domain few-shot experiments within Chinese agricultural pest and disease subdomains. Considering the scarcity of publicly available annotated data in the agricultural pest and disease domain, researchers often focus on a single crop, leading to practical cross-domain few-shot scenarios. We collect and clean unstructured textual data from the Internet to construct the Agricultural Pest and Disease Corpus (APDCorpus), sourced from online encyclopedic knowledge bases and agricultural protection information websites, covering crops such as cotton, rice, wheat, and corn. The corpus currently includes 65,292 sentences and approximately 2.08 million characters. We manually annotate the relevant wheat and cotton pest and disease texts from the corpus to construct datasets and additionally obtain a public dataset for rice pests and diseases. We use the wheat and rice datasets as source domains and the cotton dataset as the target domain for Chinese agricultural pest and disease subdomain transfer experiments. Dataset details are as follows:

- **Rice Pest and Disease Dataset**[27]: Focused on typical pest and disease issues in rice production, this dataset is collected from publicly available data curated by researchers. It includes annotations for seven types of entities such as rice disease names, rice pest names, and rice varieties, comprising approximately 60,000 characters and 2,064 sentence-level annotations.

- **Wheat Pest and Disease Dataset**: Focused on typical pest and disease issues in wheat production, this dataset is manually annotated from the APDCorpus using wheat-related pest and disease texts. It includes annotations for five types of entities: wheat disease names, wheat pest names, pesticides for wheat pest and disease control, affected regions, and wheat variety names, totaling approximately 170,000 characters and 3,610 sentence-level annotations.

- **Cotton Pest and Disease Dataset**: Focused on typical pest and disease issues in cotton production, this dataset is manually annotated from the APDCorpus using cotton-related pest and disease texts. Additionally, we incorporate text materials from a specialized book.[28] The final dataset includes annotations for five types of entities: cotton disease names, cotton pest names, pesticides for cotton pest and disease control, affected regions, and cotton variety names, totaling approximately 200,000 characters and 5,306 sentence-level annotations.

We adopt the k-shot sampling method proposed by Ding et al.[29] to sample 1-shot and 5-shot training sets for the few-shot pest and disease NER task. For each experimental setting, we sample five different training sets.

5.1.2. Experimental Parameter Setup

We focus primarily on the experimental parameter settings for the pseudo-label self-training phase. Table 1 details the parameters and their search spaces during pseudo-label training. Other hyperparameter settings are consistent with those used in TadNER[6] for reproducibility.

Table 1: Pseudo-Label Training Experimental Parameter Settings

Parameter Category	Parameter Name	Search Space
Optimizer Settings	Learning Rate Optimizer Type	1e-4, 1e-5, 3e-4, 3e-5 AdamW
Learning Rate Scheduler	Scheduler Type Warmup Steps Ratio	Linear Decay 0.0, 0.05, 0.1
Training Settings	Token Filtering Threshold Batch Size Pseudo-Label Iterations	0.9, 0.95, 0.99 8, 16, 32 1, 2, 3

5.2. Results Analysis

Table 2: Cross-Domain Few-Shot F1 Results on English Public Datasets, † Results reported from TadNER, * Our reproduction results.

Paradigms	Models	1-shot				5-shot			
		CoNLL	WNUT	GUM	Avg.	CoNLL	WNUT	GUM	Avg.
One-stage	ProtoBERT†	49.9±8.6	17.4±4.9	17.8±3.5	28.3	61.3±9.1	22.8±4.5	19.5±3.4	34.5
	NNShot†	61.2±10.4	22.7±7.4	10.5±2.9	31.5	74.1±2.3	27.3±5.4	15.9±1.8	39.1
	StructShot†	62.4±10.5	24.2±8.0	7.8±2.1	31.5	74.8±2.4	30.4±6.5	13.3±1.3	39.5
	FSLS†	50.9±6.5	14.3±5.5	12.6±2.8	25.9	63.9±3.3	24.0±3.2	18.8 ±2.2	35.6
	CONTaiNER†	61.2±10.7	27.5±1.9	18.5±4.9	35.7	75.8±2.7	32.5±3.8	25.2±2.7	44.5
Two-stage	DecomposedMetaNER†	61.2±9.2	27.7±5.3	20.3±4.2	36.4	75.2±5.8	29.8±3.9	33.5±2.4	46.2
	TadNER*	72.8±10.3	32.6±3.1	24.8±3.2	43.4	80.6±2.3	33.2±3.1	35.5±1.5	49.8
	PFSTDNER	**75.2±9.7**	**33.8±4.6**	**26.0±3.2**	**45**	**81.4±2.4**	**33.7±3.4**	**35.9±1.5**	**50.3**

Table 2 compares PFSTDNER with baseline methods on English cross-domain few-shot tasks. PFSTDNER achieves the best performance across CoNLL, WNUT, and GUM datasets. Under the 1-shot setting, PFSTDNER attains an average F1 of 45.0%, outperforming TadNER (43.4%) by 1.6 points. In 5-shot, PFSTDNER reaches 50.3%, surpassing TadNER (49.8%) by 0.5 points. All models perform best on CoNLL, likely due to its similarity to the source domain OntoNotes, while WNUT's domain shift challenges cross-domain generalization.

Table 3: Ablation Study Results for Span Detection F1 Scores in PFSTDNER

	CoNLL	WNUT	GUM
PFSTDNER	**84.6**	**51.7**	**53.1**
w/o token-filter	82.9	51.4	51.5
w/o score-filter	83.6	49.0	52.0
w/o token-score-filter	81.0	47.1	50.6
w/o self-training	80.5	47.3	49.0

To analyze module contributions, we conduct ablation studies (Table 3). Removing token or span filtering individually causes noticeable F1 drops, particularly on WNUT (up to 2.7 points), highlighting the necessity of both filtering stages. Eliminating both filters leads to the most severe degradation (up to 4.6 points). Removing self-training entirely further reduces F1 by over 4 points across datasets, confirming its critical role in mitigating domain adaptation bias.Notably, when applying self-training without filtering, performance slightly drops on WNUT, further emphasizing the necessity of proper filtering in cross-domain few-shot scenarios.

Table 4: Cross-Domain Few-Shot F1 Results on Chinese Agricultural Pest and Disease Subdomain

Models	1-shot	5-shot
TadNER	61.0	65.0
PFSTDNER	**62.8**	**66.5**

Table 4 presents our experimental results on the Chinese agricultural pest and disease subdomain cross-domain task. The results reveal:

PFSTDNER outperforms TadNER under both sample size conditions. Under the 1-shot setting, PFSTDNER achieves an F1 score of 62.8%, improving by 1.8 percentage points over TadNER's 61.0%. Under the 5-shot setting, PFSTDNER achieves 66.5%, 1.5 points higher than TadNER's 65.0%. This demonstrates that our designed dual-stage pseudo-label filtering self-training framework is also effective in Chinese agricultural pest and disease subdomain transfer tasks, validating the practical value of cross-domain few-shot methods.

Overall, the three groups of experiments lead to the following conclusions: (1) the dual-stage pseudo-label filtering strategy effectively identifies and removes pseudo-labels with source domain bias, helping the model better capture semantic boundaries and type characteristics of target domain entities; (2) the pseudo-label self-training framework expands the guiding capacity of few-shot samples and effectively mitigates domain adaptation bias in cross-domain few-shot NER.

6. Conclusion

To address the domain adaptation bias problem in cross-domain few-shot named entity recognition, we propose the PFSTDNER model, constructed based on a dual-stage pseudo-label filtering self-training decomposition framework. The proposed improvements, through token-level confidence filtering and few-shot minimum score filtering, effectively remove pseudo-labels with source domain bias, thereby enhancing the model's ability to adapt to target domain features. Experimental results demonstrate that PFSTDNER outperforms baseline methods across multiple cross-domain few-shot tasks and validate the effectiveness of the dual-stage filtering mechanism through ablation studies. Furthermore, the practical value of the proposed model is verified in cross-domain transfer experiments within the Chinese agricultural pest and disease subdomain.

7. Acknowledgments

This work was supported by the Major Science and Technology Projects in Xinjiang Uygur Autonomous Region (Grant No. 2022A02012-1).

References

[1] Z. Wang, Z. Zhao, Z. Chen, P. Ren, M. de Rijke, Z. Ren, Generalizing few-shot named entity recognizers to unseen domains with type-related features, in: Findings of the Association for Computational Linguistics: EMNLP 2023, 2023, pp. 2228–2240.

[2] T. T. H. Hanh, A. Doucet, N. Sidere, J. G. Moreno, S. Pollak, Named entity recognition architecture combining contextual and global features, in: International Conference on Asian Digital Libraries, Springer, 2021, pp. 264–276.

[3] J. Zheng, H. Chen, Q. Ma, Cross-domain named entity recognition via graph matching, in: Findings of the Association for Computational Linguistics: ACL 2022, 2022, pp. 2670–2680.

[4] Y. Chen, Y. Zheng, Z. Yang, Prompt-based metric learning for few-shot ner, in: Findings of the Association for Computational Linguistics: ACL 2023, 2023, pp. 7199–7212.

[5] S. S. S. Das, A. Katiyar, R. Passonneau, R. Zhang, CONTaiNER: Few-shot named entity recognition via contrastive learning, in: Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 6338–6353. doi:10.18653/v1/2022.acl-long.439.
URL `https://aclanthology.org/2022.acl-long.439`

[6] Y. Li, Y. Yu, T. Qian, Type-aware decomposed framework for few-shot named entity recognition, in: H. Bouamor, J. Pino, K. Bali (Eds.), Findings of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023, Association for Computational Linguistics, 2023, pp. 8911–8927.

[7] T. Ma, H. Jiang, Q. Wu, T. Zhao, C.-Y. Lin, Decomposed meta-learning for few-shot named entity recognition, in: Findings of the Association for Computational Linguistics: ACL 2022, 2022, pp. 1584–1596.

[8] J. Snell, K. Swersky, R. Zemel, Prototypical Networks for Few-shot Learning, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017.
URL `https://proceedings.neurips.cc/paper/2017/file/cb8da6767461f2812ae4290eac7cbc42-Paper.pdf`

[9] A. Fritzler, V. Logacheva, M. Kretov, Few-shot classification in named entity recognition task, in: Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing, SAC '19, Association for Computing Machinery, New York, NY, USA, 2019, p. 993–1000. doi:10.1145/3297280.3297378.
URL `https://doi.org/10.1145/3297280.3297378`

[10] T. Ma, H. Jiang, Q. Wu, T. Zhao, C.-Y. Lin, Decomposed meta-learning for few-shot named entity recognition, in: Findings of the Association for Computational Linguistics: ACL 2022, Association for Computational Linguistics, Dublin, Ireland, 2022, pp. 1584–1596. doi:10.18653/v1/2022.findings-acl.124.
URL `https://aclanthology.org/2022.findings-acl.124`

[11] Y. Yang, A. Katiyar, Simple and effective few-shot named entity recognition with structured nearest neighbor learning, in: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 6365–6375. doi:10.18653/v1/2020.emnlp-main.516.
URL `https://aclanthology.org/2020.emnlp-main.516`

[12] G. Jiang, Z. Luo, Y. Shi, D. Wang, J. Liang, D. Yang, Toner: Type-oriented named entity recognition with generative language model, in: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), 2024, pp. 16251–16262.

[13] D. Li, B. Hu, Q. Chen, Prompt-based text entailment for low-resource named entity recognition, in: N. Calzolari, C.-R. Huang, H. Kim, J. Pustejovsky, L. Wanner, K.-S. Choi, P.-M. Ryu, H.-H. Chen, L. Donatelli, H. Ji, S. Kurohashi, P. Paggio, N. Xue, S. Kim, Y. Hahm, Z. He, T. K. Lee, E. Santus, F. Bond, S.-H. Na (Eds.), Proceedings of the 29th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022, pp. 1896–1903.
URL `https://aclanthology.org/2022.coling-1.164/`

[14] Y. Shen, X. Ma, Z. Tan, S. Zhang, W. Wang, W. Lu, Locate and label: A two-stage identifier for nested named entity recognition, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), Association for Computational Linguistics, Online, 2021, pp. 2782–2794. doi:10.18653/v1/2021.acl-long.216.
URL `https://aclanthology.org/2021.acl-long.216`

[15] Y. Wang, H. Chu, C. Zhang, J. Gao, Learning from language description: Low-shot named entity recognition via decomposed framework, in: Findings of the Association for Computational Linguistics: EMNLP 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 1618–1630. doi:10.18653/v1/2021.findings-emnlp.139.
URL `https://aclanthology.org/2021.findings-emnlp.139`

[16] X. Zhang, B. Yu, Y. Wang, T. Liu, T. Su, H. Xu, Exploring modular task decomposition in cross-domain named entity recognition, in: Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2022, pp. 301–311.

[17] Probability of error of some adaptive pattern-recognition machines, IEEE Transactions on Information Theory 11 (3) (1965) 363–371.

[18] Q. Xie, M.-T. Luong, E. Hovy, Q. V. Le, Self-training with noisy student improves imagenet classification, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 10687–10698.

[19] B. Zoph, G. Ghiasi, T.-Y. Lin, Y. Cui, H. Liu, E. D. Cubuk, Q. Le, Rethinking pre-training and self-training, Advances in neural information processing systems 33 (2020) 3833–3845.

[20] S. Mukherjee, A. Awadallah, Uncertainty-aware self-training for few-shot text classification, Advances in Neural Information Processing Systems 33 (2020) 21199–21212.

[21] J. Du, É. Grave, B. Gunel, V. Chaudhary, O. Çelebi, M. Auli, V. Stoyanov, A. Conneau, Self-training improves pre-training for natural language understanding, in: Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2021, pp. 5408–5418.

[22] R. Weischedel, M. Palmer, M. Marcus, E. Hovy, S. Pradhan, L. Ramshaw, N. Xue, A. Taylor, J. Kaufman, M. Franchini, et al., Ontonotes release 5.0 ldc2013t19, Linguistic Data Consortium, Philadelphia, PA 23 (170) (2013) 20.

[23] E. F. Tjong Kim Sang, F. De Meulder, Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition, in: Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003, 2003, pp. 142–147.
URL `https://aclanthology.org/W03-0419`

[24] L. Derczynski, E. Nichols, M. van Erp, N. Limsopatham, Results of the WNUT2017 shared task on novel and emerging entity recognition, in: Proceedings of the 3rd Workshop on Noisy User-generated Text, Association for Computational Linguistics, Copenhagen, Denmark, 2017, pp. 140–147. doi:10.18653/v1/W17-4418.
URL `https://aclanthology.org/W17-4418`

[25] A. Zeldes, The gum corpus: Creating multilayer resources in the classroom, Language Resources and Evaluation 51 (3) (2017) 581–612.

[26] J. Ma, M. Ballesteros, S. Doss, R. Anubhai, S. Mallya, Y. Al-Onaizan, D. Roth, Label semantics for few shot named entity recognition, in: Findings of the Association for Computational Linguistics: ACL 2022, 2022, pp. 1956–1971.

[27] Y. Ren, J. Shen, S. Yuan, Named entity recognition of agricultural pests and diseases based on flat, in: Proceedings of the 22nd Chinese National Conference on Computational Linguistics (CCL 2023), 2023, pp. 410–419, (in Chinese).

[28] W. Zhao, H. S. (Eds.), Atlas of cotton pests and diseases, (in Chinese) (2017).

[29] N. Ding, G. Xu, Y. Chen, X. Wang, X. Han, P. Xie, H. Zheng, Z. Liu, Few-nerd: A few-shot named entity recognition dataset, in: Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), 2021, pp. 3198–3213.