EMNLP 2023

Paper ID: 680

CONTINUAL NAMED ENTITY RECOGNITION WITHOUT CATASTROPHIC FORGETTING

Duzhen Zhang, Wei Cong, Jiahua Dong, Yahan Yu, Xiuyi Chen, Yonggang Zhang, Zhen Fang
University of Chinese Academy of Sciences, Beijing, P.R.China
bladedancer957@gmail.com, congwei45@gmail.com, dongjiahua1995@gmail.com

Abstraction

Continual Named Entity Recognition (CNER) is a burgeoning area, which involves updating an existing model by incorporating new entity types sequentially. Nevertheless, continual learning approaches are often severely afflicted by catastrophic forgetting. This issue is intensified in CNER due to the consolidation of old entity types from previous steps into the non-entity type at each step, leading to what is known as the semantic shift problem of the non-entity type. In this paper, we introduce a pooled feature distillation loss that skillfully navigates the trade-off between retaining knowledge of old entity types and acquiring new ones, thereby more effectively mitigating the problem of catastrophic forgetting. Additionally, we develop a confidence-based pseudo-labeling for the non-entity type, *i.e.*, predicting entity types using the old model to handle the semantic shift of the non-entity type. Following the pseudo-labeling process, we suggest an adaptive re-weighting type-balanced learning strategy to handle the issue of biased type distribution. We carried out comprehensive experiments on ten CNER settings using three different datasets. The results illustrate that our method significantly outperforms prior state-of-the-art approaches, registering an average improvement of 6.3% and 8.0% in Micro and Macro F1 scores, respectively. Our code is available at `https://github.com/BladeDancer957/CPFD`.

***Index Terms*—** Continual Learning, Named Entity Recognition

Model Overview

In this paper, we present a novel approach named CPFD, an acronym for Confidence-based pseudo-labeling and Pooled Features Distillation, shown in Figure 1, which utilizes the old model in two significant ways to address the aforementioned challenges inherent in CNER.

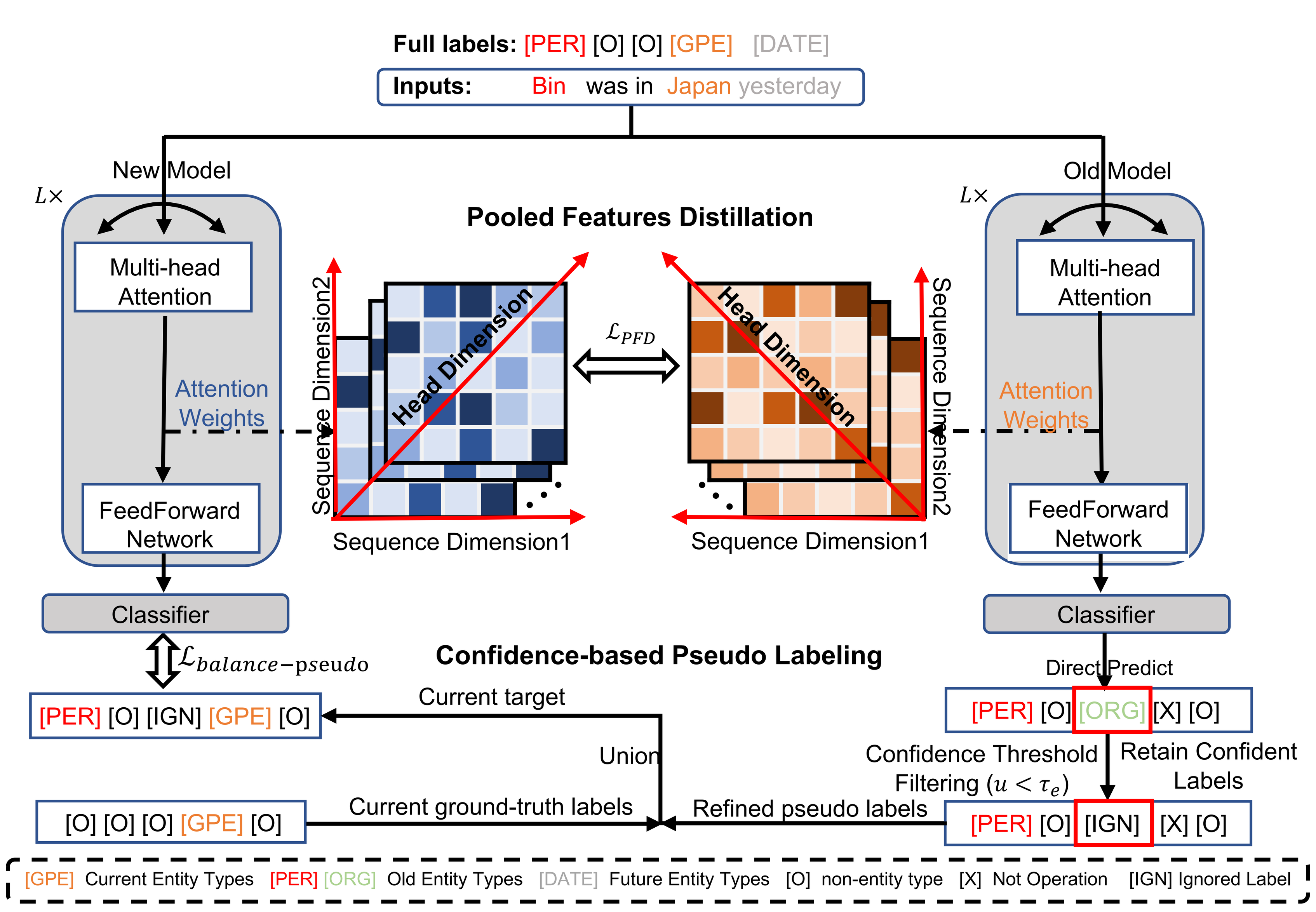

Figura 1: Our CPFD method aims to learn a NER model within a continual learning paradigm, where old entity types are collapsed into the non-entity type in the current step. We constitute a suitable balance between stability and plasticity by pooled features distillation loss to prevent catastrophic forgetting and generate high-quality pseudo-labels from old predictions by a confidence-based pseudo-labeling strategy to deal with the semantic shift problem.

Our contributions can be summarized as follows:

- We design a pooled features distillation loss to alleviate catastrophic forgetting by retaining linguistic knowledge and establishing a suitable balance between stability and plasticity.
- We develop a confidence-based pseudo-labeling strategy to better recognize previous entity types for the current non-entity type tokens and deal with the semantic shift problem. To cope with the imbalanced type distribution, we propose an adaptive re-weighting type-balanced learning strategy for CNER.
- Extensive results on ten CNER settings of three datasets indicate that our CPFD achieves remarkable improvements over the existing State-Of-The-Art (SOTA) approaches with an average gain of 6.3% and 8.0% in Micro and Macro F1 scores, respectively.

Main Results

As shown in Table 1, our CPFD method significantly surpasses the previous SOTA method, CFNER, yielding enhancements ranging from 1.74% to 14.96% in Mi-F1 and from 1.79% to 18.89% in Ma-F1 across eight CNER settings. As presented in Figure 2, our CPFD method outshines other CNER baseline methods in almost all step-wise comparisons under the eight settings. These results validate CPFD's superior performance in learning a robust CNER model, demonstrating enhanced resilience against catastrophic forgetting and semantic shift problems.

Tabela 1: Comparisons with baselines on I2B2 and OntoNotes5. The **red** denotes the highest result, and the **blue** denotes the second highest result. The marker † refers to significant test $p\text{-}value < 0.05$ comparing with CFNER. $*$ represents results from re-implementation. Other baseline results are cited from CFNER.

Dataset	Baseline	FG-1-PG-1		FG-2-PG-2		FG-8-PG-1		FG-8-PG-2	
		Mi-F1	Ma-F1	Mi-F1	Ma-F1	Mi-F1	Ma-F1	Mi-F1	Ma-F1
I2B2	FT	17.43±0.54	13.81±1.14	28.57±0.26	21.43±0.41	20.83±1.78	18.11±1.66	23.60±0.15	23.54±0.38
	PODNet	12.31±0.35	17.14±1.03	34.67±2.65	24.62±1.76	39.26±1.38	27.23±0.93	36.22±12.9	26.08±7.42
	LUCIR	43.86±2.43	31.31±1.62	64.32±0.76	43.53±0.59	57.86±0.87	33.04±0.39	68.54±0.27	46.94±0.63
	ST	31.98±2.12	14.76±1.31	55.44±4.78	33.38±3.13	49.51±1.35	23.77±1.01	48.94±6.78	29.00±3.04
	ExtendNER*	41.65±10.11	23.11±2.70	67.60±1.15	42.58±1.59	45.14±2.91	27.41±0.88	56.48±2.41	38.88±1.38
	ExtendNER	42.85±2.86	24.05±1.35	57.01±4.14	35.29±3.38	43.95±2.01	23.12±1.79	52.25±5.36	30.93±2.77
	CFNER*	**64.79±0.26**	**37.79±0.65**	**72.58±0.59**	**51.71±0.84**	56.66±3.22	36.84±1.35	**69.12±0.94**	**51.61±0.87**
	CFNER	62.73±3.62	36.26±2.24	71.98±0.50	49.09±1.38	**59.79±1.70**	**37.30±1.15**	69.07±0.89	51.09±1.05
	CPFD (Ours)	**74.19±0.95**†	**48.34±1.45**†	**78.19±0.58**†	**56.04±1.22**†	**74.75±1.35**†	**56.19±2.46**†	**81.05±0.87**†	**65.04±1.13**†
	Imp.	**⇑9.40**	**⇑10.55**	**⇑5.61**	**⇑4.33**	**⇑14.96**	**⇑18.89**	**⇑11.93**	**⇑13.43**
OntoNotes5	FT	15.27±0.26	10.85±1.11	25.85±0.11	20.55±0.24	17.63±0.57	12.23±1.08	29.81±0.12	20.05±0.16
	PODNet	9.06±0.56	8.36±0.57	19.04±1.08	16.93±0.85	29.00±0.86	20.54±0.91	37.38±0.26	25.85±0.29
	LUCIR	28.18±1.15	21.11±0.84	56.40±1.79	40.58±1.11	66.46±0.46	46.29±0.38	76.17±0.09	55.58±0.55
	ST	50.71±0.79	33.24±1.06	68.93±1.67	50.63±1.66	73.59±0.66	49.41±0.77	77.07±0.62	53.32±0.63
	ExtendNER*	51.36±0.77	33.38±0.98	63.03±9.39	47.64±5.15	73.65±0.19	50.55±0.56	77.86±0.10	55.21±0.51
	ExtendNER	50.53±0.86	32.84±0.84	67.61±1.53	49.26±1.49	73.12±0.93	49.55±0.90	76.85±0.77	54.37±0.57
	CFNER*	58.44±0.71	41.75±1.51	72.10±0.31	55.02±0.35	78.25±0.33	**58.64±0.42**	80.09±0.37	**61.06±0.37**
	CFNER	**58.94±0.57**	**42.22±1.10**	**72.59±0.48**	**55.96±0.69**	**78.92±0.58**	57.51±1.32	**80.68±0.25**	60.52±0.84
	CPFD (Ours)	**66.73±0.70**†	**54.12±0.30**†	**74.33±0.30**†	**57.75±0.35**†	**81.87±0.47**†	**65.52±1.05**†	**83.38±0.18**†	**66.27±0.75**†
	Imp.	**⇑7.79**	**⇑11.90**	**⇑1.74**	**⇑1.79**	**⇑2.95**	**⇑6.88**	**⇑2.70**	**⇑5.21**

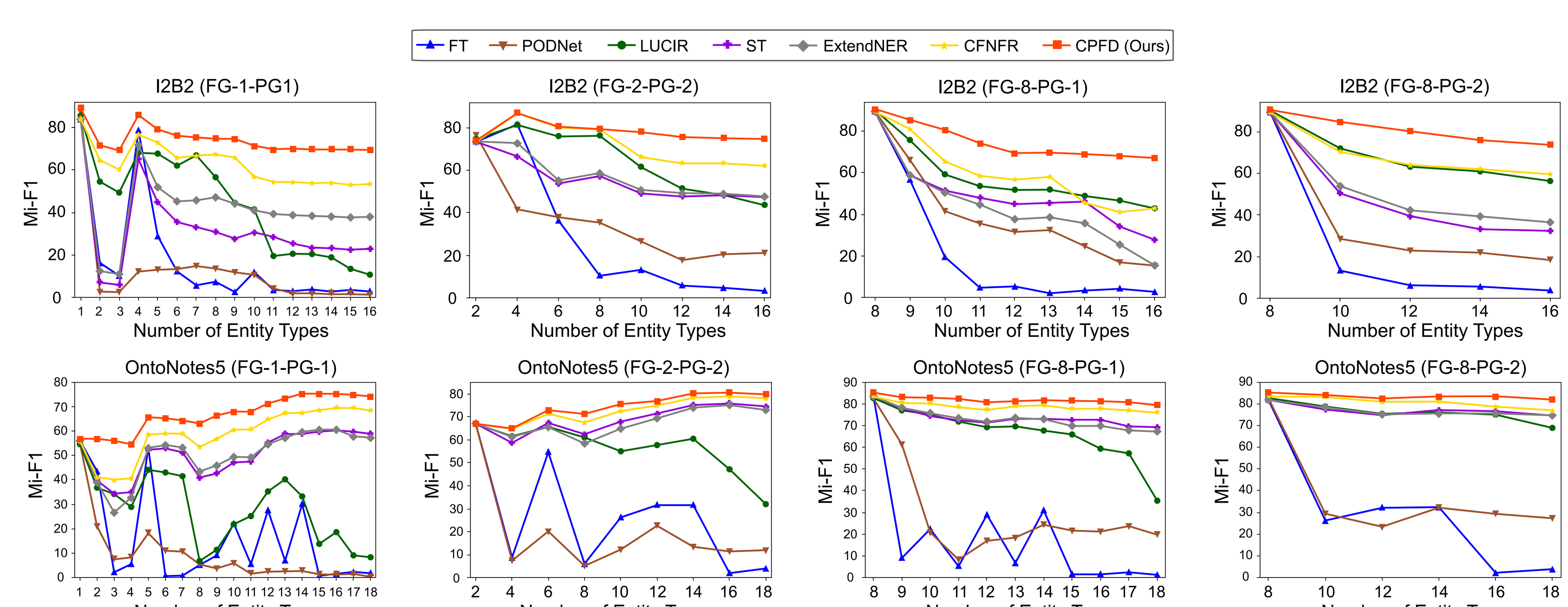

Figura 2: Comparison of the step-wise Mi-F1 on I2B2 and OntoNotes5. The result of baselines is directly cited from CFNER.

Ablation Study

This section examines the effectiveness of individual components in our CPFD method through ablation studies, the results of which are shown in Table 2. These results verify the importance of all components to address CNER collaboratively.

Tabela 2: The ablation study of our CPFD on I2B2 and OntoNotes5 under the setting FG-1-PG-1. When compared with Ours, all ablation variants severely degrade CNER performance. It verifies the importance of all components to address CNER collaboratively.

Methods	I2B2		OntoNotes5	
	Mi-F1	Ma-F1	Mi-F1	Ma-F1
CPFD (Ours)	**74.19±0.95**	**48.34±1.45**	**66.73±0.70**	**54.12±0.30**
w/ $\mathcal{L}_{\text{FD}}$	71.46±1.19	45.17±1.28	63.80±1.01	51.83±0.73
w/ $\mathcal{L}_{\text{PFD-lax}}$	70.22±0.90	43.89±1.10	62.32±0.53	50.12±0.70
w/o $\mathcal{L}_{\text{PFD}}$	68.66±0.88	42.28±0.79	60.80±0.86	48.94±1.38
w/o CPL	54.86±5.36	37.39±3.58	59.37±0.82	46.68±0.45
w/o ART	72.29±1.56	45.35±1.83	65.19±1.33	52.94±0.46