Zhen-zhen LI, Da-wei FENG, Dong-sheng LI, Xi-cheng LU, 2020. Learning to select pseudo labels: a semi-supervised method for named entity recognition. *Frontiers of Information Technology & Electronic Engineering*, 21(6):903-916. https://doi.org/10.1631/FITEE.1800743

Learning to select pseudo labels: a semi-supervised method for named entity recognition

Key words: Named entity recognition; Unlabeled data; Deep learning; Semi-supervised method

Corresponding author: Dong-sheng LI
E-mail: dsli@nudt.edu.cn
ORCID: https://orcid.org/0000-0001-9743-2034

Motivation

- Deep learning models have achieved state-of-the-art performance on named entity recognition (NER) tasks, but they rely on high-quality labeled data, which is scarce in some domains.
- Semi-supervised methods that assign pseudo labels to unlabeled data can relieve the lack of labeled data.
- Previous approaches relying on heuristic rules or statistical features to improve the quality of pseudo labels cannot be applied to other domains; hence, we aim to learn a model to evaluate the pseudo labels without relying on any external knowledge.

Main idea

- A neural NER model is constructed according to the application domain and optimized with existing labeled data, which is used to annotate the unlabeled data.
- We learn a Judge model to evaluate the reliability of pseudo labels, which is based on machine learning algorithms.
- We propose the strategies to filter out possible wrong pseudo labels and create new labeled data.
- We propose the framework to improve the neural NER model by iterating between creating new labeled data and re-training the NER model with the enlarged training set.

Method

- We construct the transformed data for the Judge model with features from the pre-trained neural NER model and label information from NER training data.
- The algorithms to train the Judge model are explored among the softmax-based, SVMs, and feed-forward neural network model. We choose the simple softmax-threshold method as the final default Judge model, which shows the best generalization ability.
- We use the learned Judge model to filter out probably wrong self-labeled named entities, so as to create reliable labeled data from unlabeled data to improve the NER model.

Major results

1. Results on the CoNLL-2003 NER task

Table 4 Test set F_1 comparison with previous neural network models on the CoNLL-2003 NER task

Model	F_1-score (%)
Collobert et al. (2011)♣	89.59
Huang et al. (2015)	90.10
Lample et al. (2016)	90.94
Ma and Hovy (2016)	91.21
Chiu and Nichols (2016)♣ ‡	91.62
Peters et al. (2017)‡	91.93
Rei (2017)	86.26
Ghaddar and Langlais (2018)♣	91.73
Our semi-supervised model	91.73
Our semi-supervised model‡	92.04

NER: named entity recognition. ♣ indicates models that use additional supervised information (gazetteers or Wikipedia); ‡ indicates models trained on both the training and development sets

Major results (Cont'd)

Comparison of semi-supervised methods for the CoNLL-2003 NER task

Table 5 Improvements in test set F_1 in CoNLL-2003 NER with semi-supervised methods

Model	F_1-score (%)		△ (%)
	Baseline NER model	Semi-supervised method	
Rei (2017)	86.00	86.26	+0.26
Peters et al. (2017)‡	90.87	91.93	+1.06
Ours	90.79	91.73	+0.94
Ours‡	91.12	92.04	+0.92

NER: named entity recognition. ‡ indicates models trained on both the training and development sets; △ indicates the improvement

Major results (Cont'd)

2. Comparison of neural models for the OntoNotes 5.0 NER task

Table 6 Test set F_1 comparison with previous neural models on the OntoNotes 5.0 NER task

Model	F_1-score (%)
Chiu and Nichols (2016)♣	86.28
Strubell et al. (2017)	86.99
Li et al. (2017)	87.21
Ghaddar and Langlais (2018)♣	87.95
Our baseline NER model	87.18
Our semi-supervised model	87.80

NER: named entity recognition. ♣ indicates models that use additional supervised information (gazetteers or Wikipedia)

Major results (Cont'd)

3. Performance improvements with unlabeled data for CNER-2017

Table 8 Improvements in test set F_1 in CNER-2017 when using unlabeled data

Model	F_1-score (%)		△ (%)
	w/o	w/	
Xia and Wang (2017)[1]	90.17	91.14	+0.97
Xia and Wang (2017)[2]	91.02	91.08	+0.06
Hu et al. (2017)[1]	88.47	88.67	+0.20
Hu et al. (2017)[2]	88.47	88.83	+0.36
Ours	87.91	88.56	+0.65

Xia and Wang (2017)[1,2] represent single model and voted model, respectively; Hu et al. (2017)[1] represents adding self-taught learning only; Hu et al. (2017)[2] represents adding both self-taught learning and active learning. △ indicates the improvement

Conclusions

- We generalize the effective neural architecture for NER tasks.
- Based on the learned neural NER model, we learn a softmax-based Judge model to select pseudo labels for unlabeled data. Then we create reliable labeled data with the learned Judge model.
- Our semi-supervised approach effectively improves the neural NER model.