Named Entity Recognition for Chinese Social Media with Domain Adversarial Training and Language Modeling

Yong Xu[1] , Qi Lu[2] and Muhua Zhu[3]*

[1]School of Computer science, Fudan University, Shanghai, China.
[2]Meituan, Beijing, China. [3]Alibaba Group, Hangzhou, Zhejiang, China

Introduction

Motivations

1) Domains like social media lack large-scale labeled datasets, which renders NER for social media a challenging task. The text in social media is ungrammatical and contains nonstandard orthography and noises.
2) Domain adaptation approaches can learn from multiple datasets of diverse.
3) Language models can learn general representations from unlabeled in-domain data.

Contributions

1) We propose a novel neural model for domain adaptation of NER whose building blocks are domain adversarial training and language modeling. Experiments on Chinese social media data show the effectiveness of the model.
2) To experiment with a bigger Chinese NER corpus in social media, we expand a previously released dataset and render the new corpus publicly available for research in this direction.

Example

[陈雄/PER]围脖：请解释[北京/LOC]电工[老李/PER]
微博二传手瘦驼tuodi1968

Chenxiong Weibo: please explain Beijing electrician
laoli Weibo passer Thin Camel toudi1968

Figure 1:An example of Weibo posts annotated with named entities (highlighted in blue).

Approach

Overview

1) The Char embedding is used to represent words in corpus.
2) BiLSTM on the left is used as common representations between domains and the one on the right is used to learn private representations.
3) A CRF network on the basis of the two BiLSTM representations is used for the purpose of recognizing named entities.
4) Bidirectional language models are effective as an auxiliary objective for sequence labeling.
5) Domain-adversarial training is used to encourage the common BiLSTM to be domain-agnostic.
6) Loss of three components are combined together as the overall loss.

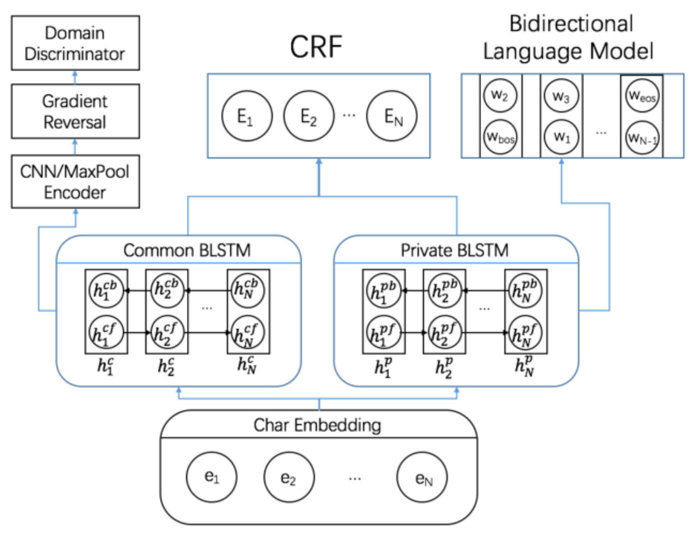

Figure 2: Architecture of our neural network model for NER domain adaptation.

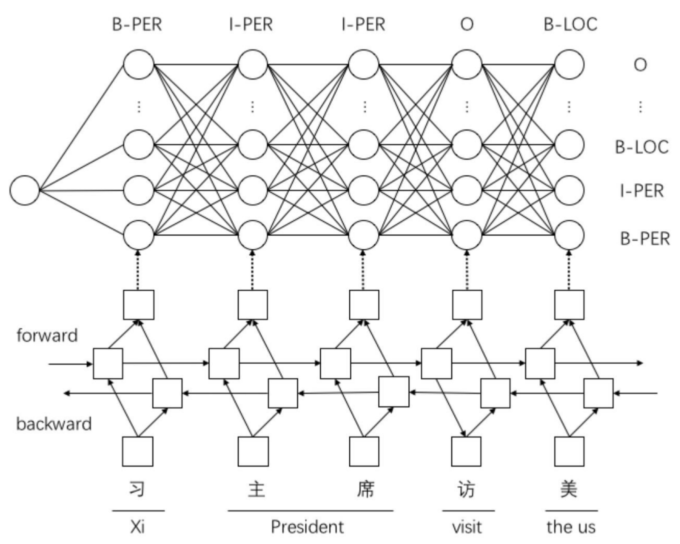

Figure 3: Illustration of bidirectional LSTM-CRF.

$$Loss = L_{CRF} + \lambda_1 L_{DA} + \lambda_2 L_{LM}$$

Domain-Adversarial Training

1) The first layer is a convolutional neural network and BiLSTM outputs.
2) The vectors from the previous layer are concatenated and forwarded to the domain discriminator through the gradient reversal layer.
3) The discriminator is implemented as a single fully-connected neural network.
4) The loss function of the domain classifier is formulated as

$$L_{DA} = -\sum_{i=1}^{S} d_i log(\hat{d}_i)$$

Training strategy

1) For BiLSTM-CRF, labeled in-domain and out-of-domain data are used together to optimize its objective. We assign labeled in-domain data to the positive class and labeled out-of-domain data to the negative class.
2) Language models are trained on unlabeled in-domain data.
3) To keep gradient from language model consistent with the other two components. We first train language models with a relatively big learning rate. Then we learn the three objectives with a smaller learning rate.

Evaluation

Dataset

- **MSR:** the labeled out-of-domain corpus, proposed by the sixth SIGHAN Workshop on Chinese language processing. Three NER types: **PER**, **LOC** and **ORG.**
- **Weibo:** the labeled in-domain corpus, released by Peng and Dredze (2015). Four NER types: **PER**, **LOC, ORG** and **GPE. GPE** type is changed to **ORG** to make sure that entity types of both domain are the same.

Data Partitions	#Sent*	#Char	#Entity
MSR Train	46,364	2,169,879	74,703
MSR Test	4,365	172,601	6,181
Weibo Train	2,000	119,714	8,092
Weibo Dev	890	52,719	698
Weibo Test	1,000	50,336	744
Weibo Unlabeled	1,000,000 Weibo posts		

Table 1: Statistics of the out-of-domain and in-domain data.∗We regard each Weibo post as a sentence.

Baseline Models

Models	Description
BiLSTM-CRF-OOD	a BiLSTM-CRF model trained on labeled out-of-domain training data (MSR training data)
BiLSTM-CRF-ID	a BiLSTM-CRF model trained on labeled in-domain training data (Weibo training data)
BiLSTM-CRF-Merge	a BiLSTM-CRF model trained on the combination of labeled out-of-domain and in-domain training data

Results

1. *+DA denotes the model which consists of BiLSTM-CRF and domain adversarial training.
2. *+DA+LM refers to the model with domain adversarial training and language models being combined with BiLSTM-CRF.

Systems†	Precision	Recall	F1
∗-OOD	30.3	34.1	32.1
∗-ID	46.5	37.6	41.6
∗-Merge	47.4	42.3	44.7
∗+DA	50.0	41.4	45.3
∗+DA+LM	55.9	46.2	50.6

Table 2: Results of the baseline systems and our models on the in-domain test set. Here the symbol ∗ refers to Bi-LSTM CRF.

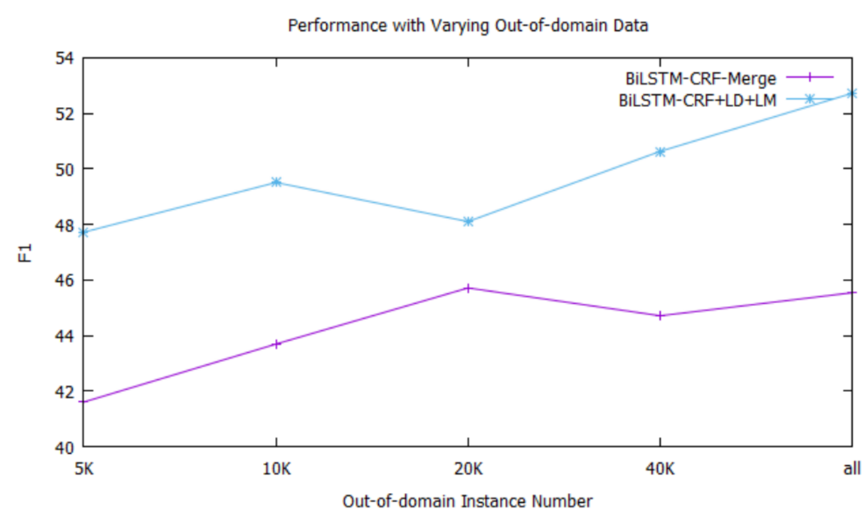

Figure 4: Varying sizes of labeled out-of-domain training data

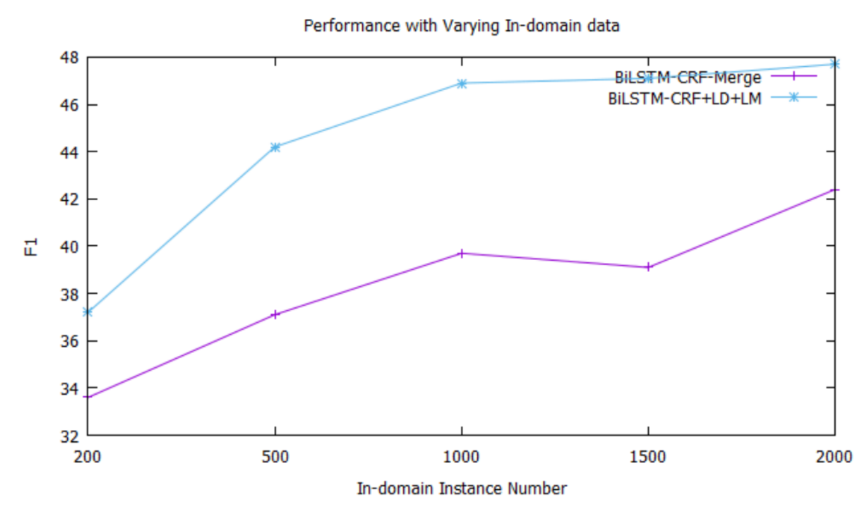

Figure 5: Varying sizes of labeled in-domain training data

Error analysis

System	CO	BC	CR	NC	CA	All
Base	73	58	14	162	36	343
+DA	62	51	11	136	37	297
+DA+LM	70	36	7	129	31	273

Table 3: Effect of our models changing the numbers of errors of each error type: CONTAIN(CO), BE-CONTAIN(BC), CROSS(CR), NO-CROSS(NC), CATEGORY(CA)

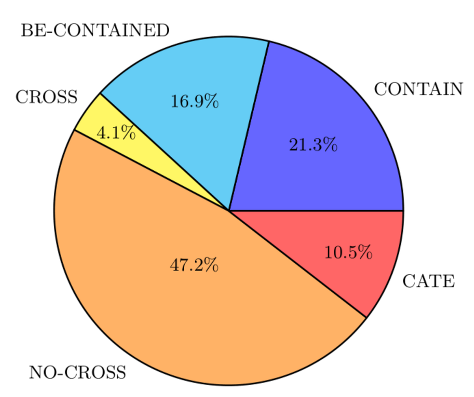

Figure 6: Error types and theirs distribution by BiLSTM-CRF-Merge on the Weibo test.

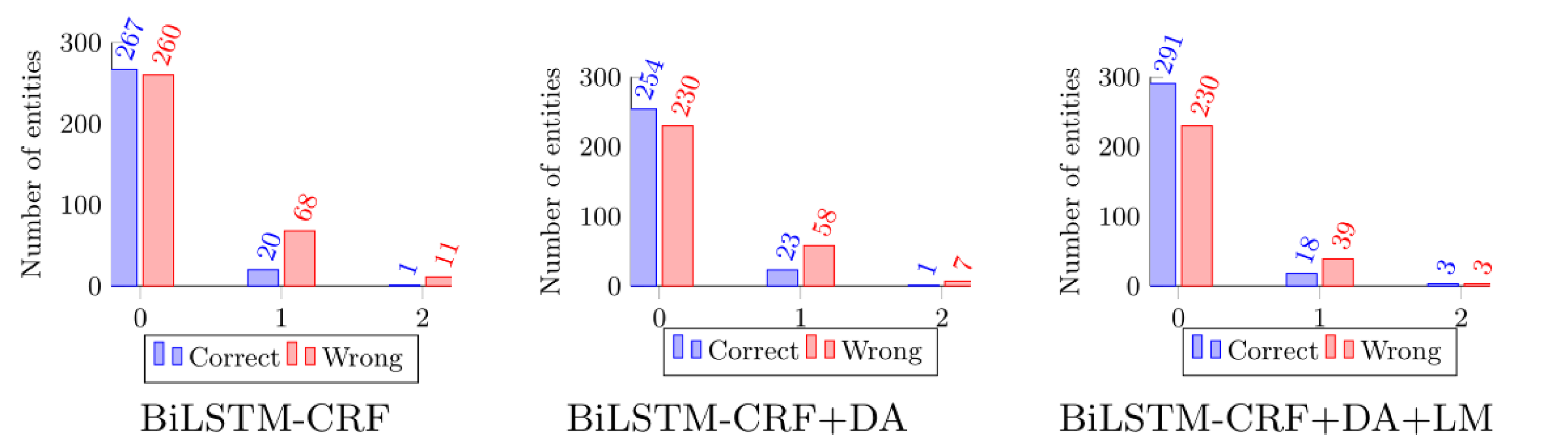

Figure 7: Correct and wrong prediction numbers in different entity lengths of different model. The x coordinates 0, 1, 2 refer to the length ranges of [1,4], [5,8], and [9,12], respectively.

Conclusion

1) We proposed a novel neural network model for domain adaptation of named entity recognition in Chinese social media.
2) The proposed model can learn from labeled out-of-domain data, labeled in-domain data, and unlabeled in-domain data.
3) Results showed that the proposed approach could improve over the baselines significantly.