

An Instance Transfer-Based Approach Using Enhanced Recurrent Neural Network for Domain Named Entity Recognition

CHUANBO LIU¹, CHAOJIE FAN^{ID}¹, ZHENGJU WANG¹, AND YUEQING SUN²

¹School of Mechanical and Electronic Engineering, Wuhan University of Technology, Wuhan 430070, China

²School of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China

Corresponding author: Chuanbo Liu (lchb72@whut.edu.cn)

ABSTRACT Recently, neural networks have shown promising results for named entity recognition (NER), which needs a number of labeled data to for model training. When meeting a new domain (target domain) for NER, there is no or a few labeled data, which makes domain NER much more difficult. As NER has been researched for a long time, some similar domain already has well labeled data (source domain). Therefore, in this paper, we focus on domain NER by studying how to utilize the labeled data from such similar source domain for the new target domain. We design a kernel function based instance transfer strategy by getting similar labeled sentences from a source domain. Moreover, we propose an enhanced recurrent neural network (ERNN) by adding an additional layer that combines the source domain labeled data into traditional RNN structure. Comprehensive experiments are conducted on two datasets. The comparison results among HMM, CRF and RNN show that RNN performs better than others. When there is no labeled data in domain target, compared to directly using the source domain labeled data without selecting transferred instances, our enhanced RNN approach gets improvement from 0.8052 to 0.9328 in terms of F1 measure.

INDEX TERMS Neural networks, NER, instance transfer strategy, source domain, ERNN.

I. INTRODUCTION

In recent years, Web data and knowledge management attracts the interests from industry and research fields. There are various promising applications, such as intelligent recommendation, machine Question & Answer, knowledge graph and so on. Named entity recognition is a fundamental and very important step in the automatic information extraction. A recognition method with high quality can directly improve the follow-up processing results of Web data management products. NER research shows great successes in various domains and becomes a hot topic [10], [26], such as social media [28], [33], language texts [13], biomedicine [2], [25], and so on. As we know, texts of different domains may vary from features, writing styles and structures. Domain NER meets a challenge that annotating data for new domains is labor intensive.

For domain NER, a new domain (target domain) has no or a few labeled data. However, it is natural to think that if some similar domain (source domain) with enough labeled data

already exists, it is possible to borrow some from this similar domain. Domain adaptation targets at transferring the source domain knowledge to the target domain [16], which is an effective way to solve the problem of labeling large amount of data on new corpus or domains. In other words, if a NER model is trained well for some fixed source domain, it is interesting to study how to deploy them across one or more different target domains.

In this paper, we focus on domain NER for the politics text domain in Chinese high schools to support an automatic question and answer system (Q & A) that will take the national college entrance examination (NCEE) in the future. However there is no public politics text corpus used by Chinese high school for NER task. Moreover, there is no labeled data. We notice that People's Daily corpus is free for public download and similar to politics text. Therefore, we propose an instance transfer based approach for domain NER with enhanced Recurrent Neural Network (RNN). First, we design an instance transfer strategy to extract similar sentences from the People's Daily corpus (source domain). Here, instances mean labeled data and politics text is our target domain. Then, recurrent neural network (RNN) model

The associate editor coordinating the review of this manuscript and approving it for publication was Huiqing Wen ^{ID}.

can be trained based on the transferred instances. Moreover, we improve the traditional RNN model by enhancing its activation function and structure. Finally, an instance transfer enhanced RNN (ERNN) model is proposed to do NER for politics text target domain, which is trained based on the transferred instances from similar source domain (People's daily corpus). Compared with the traditional RNN model, experimental results show that our ERNN with the instance transfer strategy can get improvement from 80.52% to 93.28% in terms of F1 measure.

In addition, we consider other situation where a small number of labeled data is available to further investigate the performance of our proposed approach. Since the politics text as target domain data is needed to obtain, we collect texts from high school books and relevant websites and manually label a very small set of them, making labeled data much less than the unlabeled one. Experimental results show that labeling data for target domain is quite useful, reaching at 92.13 in terms of F1 measure. However, our instance transfer based enhanced RNN can further improve the F1 value to 93.81. Finally, we adopt the co-training approach using our proposed ERNN model and CRF by taking advantage of large unannotated target domain data, which get the F1 value of 94.02.

The rest of the paper is organized as follows. In Section 2, we introduce related work on NER (especially on domain NER), recurrent neural network and transfer learning. Section 3 elaborates our approach including the instance transfer strategy, our proposed ERNN. Experiments and results are given in Section 4, then we make the conclusion and future work in Section 5.

II. RELATED WORK

As the virtual step in information extraction and a shared task of CoNLL-2003 [23], NER has been widely studied nowadays and most are about domain NER. Reference [35] employ conditional random fields (CRF) and structure support vector machines (SSVMs) to do chemical entity mention recognition in patients and chemical passage detection. Reference [6] propose novel active learning algorithms to significantly save annotation cost for clinical NER task. Besides, considering the condition that available labeled data is not at hand sometimes, [5] present a NER system for tagging fiction. To avoid the lack of annotated data, they bootstrap a model from term clusters and leverage multiple instances of the same name in a text instead. However it needs to pass through the corpus to build a feature vector corresponding to all the contexts in one document. Moreover, a large number of researches about domain NER are on biomedical fields [9], [20], and the typical methods are CRF [12], [24], and other supervised learning approaches [11].

As neural networks have shown promising results in domain NER [9], such as RNN, LSTM, and so on. Reference [27] use deep neural networks for referring to the real world (i.e. game states) to improve Japanese chess NER. Reference [34] combine the bidirectional long short-term

memory (LSTM) and CRF to automatically explore words and characters level features. However, transfer learning and semi-supervised learning, like co-training, which are effective ways for the situation when train data is much less than unannotated data.

When considering how to deal with the situation of lack of labeled data, [21] have categorized and reviewed the research progress on transfer learning for classification, regression, and clustering problems. Similar to the work of [22] in domain adaption, in this paper we study a instance transfer strategy, which is hot today but rarely used in NER [3], [7], to make use of out-of-domain data (source domain).

Co-training is also an effective way to use unlabeled data in supervised learning task, [14] use a bilingual co-training and maximum entropy model to carry out English and Chinese NER. Reference [19] modify the original co-training to cover knowledge from unlabeled data to recognize bio named entities in text. Besides, clustering should be another research direction for this problem [29], [30], [31]. Our interest is domain NER for Chinese high school news texts. In this paper we also put our proposed ERNN into a co-training approach to see how much improvement can be obtained through large amounts of unlabeled data.

The contributions and differences of our work from others are listed as follow:

- 1) We design a instance transfer strategy by selecting similar sentences from a source domain.
- 2) We add an additional layer to the traditional RNN structure by coupling knowledge from the transferred similar instances.
- 3) We conduct experiments on two datasets. No matter there is no labeled data or no in target domain, our proposed approach can improve the quality of NER.

III. OUR APPROACH

A. PROBLEM DESCRIPTION

Manually labeling data is time consuming in most machine learning methods. However, there are some other domain data with well-labeled data. As Pan and Yang [21] have declared, in such cases, knowledge transfer, if done successfully, would greatly improve the performance of learning without taking lots of efforts to label data. Our interest is Chinese high school politics text NER to support automatic Question and Answer application. There is no labeled data at hand for this new domain, but we find that some public NER data, like People's Daily corpus, is similar to it. The similar domain with well-labeled data is called source domain, while the new domain without enough labeled data is called target domain. In this paper, we propose an instance transfer enhanced RNN model for domain NER by leveraging labeled data from a source domain. In Section 3.2, we will introduce our instance transfer strategy that will select similar labeled data (sentences) from a source domain. In Section 3.3, we will discuss how to improve the traditional RNN model

by adding those selected data into the network structure of RNN.

B. OUR INSTANCE TRANSFER STRATEGY

We design an instance transfer strategy where the target domain is Chinese high school politics texts, and People's Daily is the source domain. Since texts from different domains may vary from features, writing styles and structures, it is not suitable to use all labeled data from the source domain. In this paper we first consider two functions to compute sentence similarity between source domain and target domain. The two functions are Gaussian radial basis function (RBF) and polynomial kernel function which are described as following:

- Gaussian RBF kernel:

$$K(\vec{x}, \vec{z}) = \exp\left(-\frac{(\|\vec{x} - \vec{z}\|)^2}{2\sigma^2}\right) \quad (1)$$

- polynomial kernel function:

$$K(\vec{x}, \vec{z}) = (\vec{x} \bullet \vec{z} + 1)^d \quad (2)$$

Here the \vec{x} stands for the sentence vector in source domain, while \vec{z} represents the mean of target domain corpus matrix. And we set $\sigma = 1$ and $d = 2$. We use the two functions to evaluate the sentence similarity between source domain and target domain and sort source domain sentences according to their similarity scores. Then two transfer strategies are given to select top similar sentences:

- 1) Directly regard the top n sentences as target domain training data.
- 2) Repeat the nth sentence k times in source domain where the value of k depends on n. The original source corpus has been enlarged.

C. OUR ERNN MODEL

Through our instance transfer strategy, target domain borrows labeled data from source domain and top similar sentences have been assigned more weights than others. Therefore, the next step of our approach is to enhance RNN model by utilizing those labeled data. RNN is designed to solve the problems involved in time and sequence [18]. Not only the output of moment t is influenced by that of previous moment, but also the nodes between hidden layer are all connected to each other. In this paper, we propose an enhanced RNN by modifying both its activation function and structure to do domain NER task.

1) ACTIVATION FUNCTION

As described in [8], a deep neural network is an ensemble of vectors and matrices that stand for bias and weight values, and nonlinear activation function. A suitable activation function means much for a deep neural network. Its changes can speed up model training [32] and enhance stability [15], etc. However, for a given domain, the best nonlinear function remains unknown, even though many rectifier-type nonlinear functions have been proposed as activation functions [8]. The

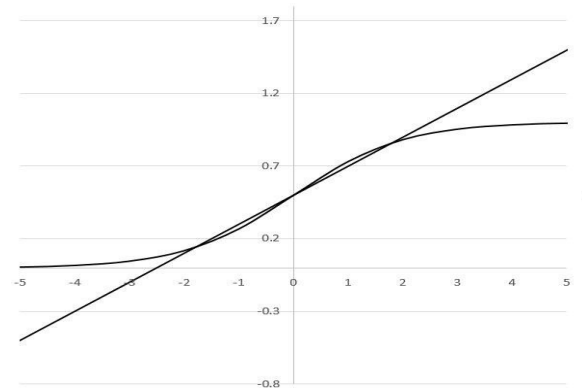


FIGURE 1. Sigmoid and its linear approximation function.

performance of a same activation function may be widely different when applying it to different tasks. For NER task, compared to other frequently used activation functions such as ReLu, PReLU, tanh and so on, the sigmoid function shows best in our preliminary experiments. Its definition is given in Eq. 3.

$$F(x) = \frac{1}{1 + e^{-x}} \quad (3)$$

However, the model training process involves the derivation, while the derivative of sigmoid function tends to zero when the argument x approaches infinity (which called the saturation phenomenon) both at its left and right. This may make it harder to train model. Even so, sigmoid is most similar to the reflex mechanism on the biological neuron level and the interval of its output is always between 0 and 1, which can represent the prediction probability of a label. In our experiments, its linear approximation function expressed by Eq. 4, shows a good performance and a is set to be 0.2 and b is 0.5 A more intuitive depiction is shown in Figure 1.

$$L(x) = a \times x + b \quad (4)$$

Focusing on our NER task, a new activation function is proposed by combining the two above functions (shown in Eq. 5). Here parameters α and β are coefficients and they are determined by experimental performances.

$$A(x) = \alpha \times F(x) + \beta \times L(x) \quad (5)$$

The new activation function has some advantages compared to its two original functions: sigmoid and linear approximation. It ameliorates the former's saturation phenomenon on the one hand and smooths the latter on the other hand. In our experiments, the results using the proposed activation function are better than that using sigmoid or its linear approximation.

2) MODEL STRUCTURE

In order to take full advantage of source domain data, we also improve RNN's structure and propose an instance transfer enhanced RNN (ERNN), which can input the source domain

TABLE 1. Sentences in ATIS.

Sentence	flight	from	memphis	to	tacoma
Label	O	O	B-fromloc.city_name	O	B-toloc.city_name
Sentence	limousine	service	at	logan	airport
Label	B-transport_type	O	O	B-toloc.airport_name	I-airport_name

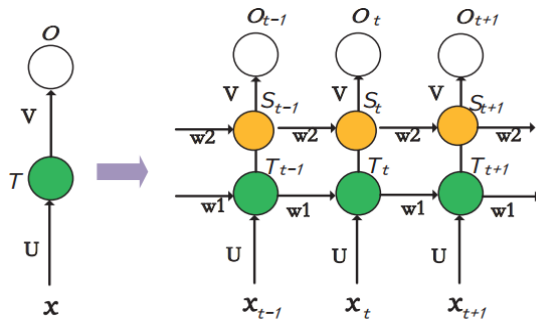


FIGURE 2. The RNN structure before/after modified.

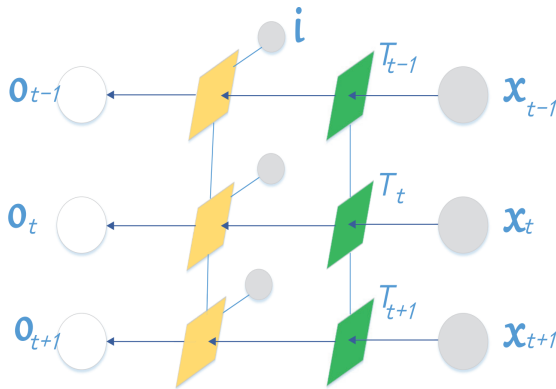


FIGURE 3. The overview of our ERNN.

sentences into RNN. Our ERNN is modified based on the Elmantype described in [17].

The new structure we modified is depicted as Figure 2 and Figure 3. In Figure 2, the left shows for the original structure and the right describes our modified recurrent neural network when unfolding its structure in time of the computation involved in its forward computation. T nodes represent the hidden layer, meanwhile S indicates the confluent layer. The source domain data combines with the output of the hidden layer T . The overview of the structure is shown in Figure 3 that depicts the domain source data input i compared to Figure 2. As shown in Figure 2, \vec{U} , \vec{V} , \vec{w} , \vec{w}_1 , \vec{w}_2 are weights and parameters. The output of the S_t is computed by Eq. 6. In output layer, we use softmax to get the prediction results i.e. $o(t)$, as shown in Eq. 7.

$$S_t = F(Tt \cdot \vec{W} + i \cdot \vec{w}_2 + \vec{b}_0) \quad (6)$$

$$O(t) = \text{softmax}(S_t \cdot \vec{V} + \vec{b}_1) \quad (7)$$

Here F is the activation function, i stands for the source domain data, as marked in Figure3. \vec{b}_0 and \vec{b}_1 are bias

vectors. In this way we achieve the reuse of the valuable source domain corpus.

IV. EXPERIMENTS

We conduct two sets of experiments on two totally different datasets for different purposes. One is NER experiments among three popular models, i.e., HMM, CRF and RNN, which aims to explore the classification quality with different training data sizes. The other is domain NER using our ERNN with instance transfer.

A. DATA DESCRIPTION

There are two datasets used in our experiments. For NER experiment among three popular models, we mainly focus the changes of the classification performance with different amount of training data. We conduct the experiment on a standard corpus named ATIS [17]. It contains 127 classes and uses the in/out/begin (IOB) representation. A sentence in ATIS can be expressed as in Table 1.

The other dataset is collected from Chinese high school politics textbooks and websites, which is the target domain in this paper. The source domain corpus is People’s Daily 1998 that is open online. In addition, due to lack of label information, we set 11 classes for target domain based on People’s Daily 1998, including person names, regular time words, single time words such as those means “night”, proprietary organization and company name. For example, Chinese characters are “Xinhua News Agency” and “the United Front Department”, and special region name, noun and etc.

B. DATA PRE-PROCESSING

Specifically, for NER experiment, ATIS contains 3983 training sentences and 893 testing sentences. In order to find out how it works when the labeled data is less than the unlabeled one, we reorganized the training set. Based on ATIS original train/test proportion (3983/893), we randomly select 20%, 40%, 60%, 80% and 100% of total train data (3983 sentences) as our training sets to train two traditional statistical models (Hidden Markov Model and CRF) and RNN model and one RNN model. For domain NER experiment, we first pick the most common words from target domain data as our dictionary. By counting the times a word has appeared in this corpus, we select top 13,450 words as our dictionary and other words are marked as ‘unknown’ in a sentence. After this, we clean the data based on a rule that is if a sentence meets any of the following two conditions, it would be regarded as noisy data and abandoned..

- The length is less than 3

TABLE 2. Sentences in ATIS.

Corpus-Domain	Labeled Data	Train Set	Test Set	Unlabeled
Source	4818	3855	963	0
Target	3043(manually labeled)	2035	1008	13549

- Too many ‘unknown’ tags(more than 50%)

After preprocessing, the People’s Daily sentences have been reduced from 19484 to 4818, meanwhile the number of politics text sentences is also reduced from 82700 to 15584. Since there is no annotated data of the latter, we manually mark 3043 sentences. Table 2 shows the statistics of the labeled and unlabeled data size of both datasets.

For People’s Daily, we use 80%/20%of the total 4818 sentences as our train/test data sets. For our target domain data in severe lack of tagged data, we use manually labeled 2035 sentences and tag other 1008 sentences from the large unannotated set as the test set.

C. EVALUATION AND SETUP

In this paper, all the experimental results were evaluated by the popular evaluation measures, i.e., precision (P), recall (R) and F1-score. In addition, for supervised learning experiment, we used K-fold cross validations. K is different based on the different size of train set. For example, when using 20% of the original train set, K is 5, and when using 40% and 60%, K is 3.

We also do some preliminary experiments to set the ERNN parameters. For example, the word embedding we use is trained by target domain texts via word2vec, a Google open source project which brings about 14.1% performance improvement. Besides, we set the ERNN context window size equal to 1, which performs better than those 5, 7 and 9 (probably because there are too many nonentity labels in a sentence). For instance transfer, since the RBF similarity function achieves a better result, we use it to enlarge the source domain corpus.

D. NER EXPERIMENTAL RESULTS OF THREE POPULAR NER MODELS

In this part, we compare the NER performance of a current popular deep learning model, i.e., RNN with two traditional models, i.e., HMM and CRF. And we also mainly explore the NER performance with different training data sizes. We conduct several experiments using ATIS’s total training data’s 20%, 40%, 60%, 80% and 100% to train the three popular NER models. The results are shown in Figure 4, Figure 5 and Figure 6. FromFigure4, Figure5 and Figure6 we can see how the three models behave when being trained on different data sizes. Overall, the P, R and F1 of these three models are gradually rising when using more training data. However, the performances are not growing linearly when the data size gets larger. While using more and more data to train a model,

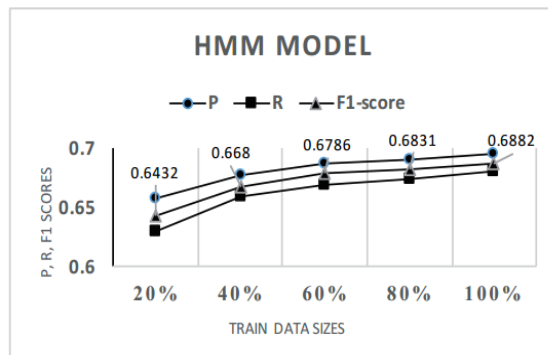


FIGURE 4. Results of the HMM model.

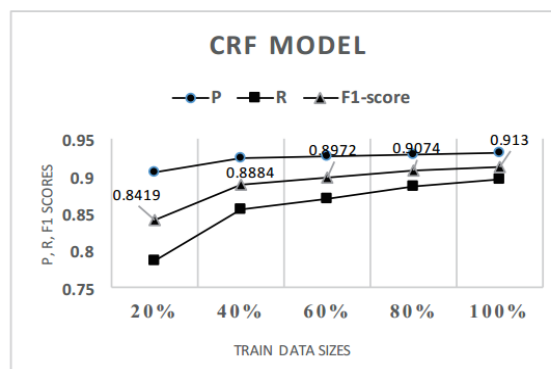


FIGURE 5. Results of the CRF model.

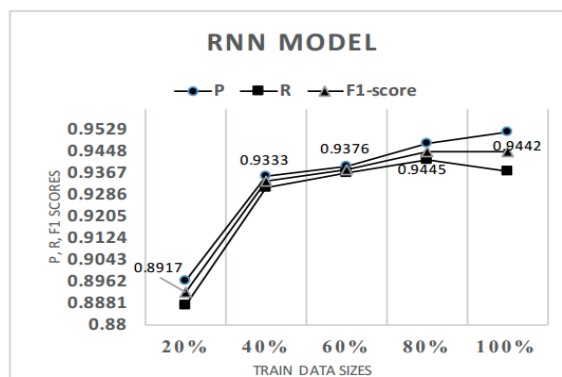


FIGURE 6. Results of RNN model.

the degree of improvement becomes smaller. Actually, when the training data size increases from 20% to 40%, the three models generally achieve a highest improvement (about 4% improvement in terms of F1 measure). After that, increasing the training data size brings little benefits.

We have done experiments on BiLSTM-CRF. When using all training data, its F1 score is 90.28. Because this article mainly discusses the impact of transfer learning on NER performance, the following experiments discuss the two basic models of CRF and RNN.

This indicates that when meeting a new domain for NER, labeled data will actually improve performance. But

manually labeling data is labor intensive. Putting more labor on labeling data could not bring us larger improvement considering the time and the labor we need. Therefore, in this paper, we propose an instance transfer strategy by transferring labeled data from other well labeled domain, which is a effective way to solve the problem of lack of labeled data in a new domain.

E. DOMAIN NER EXPERIMENTAL RESULTS

1) INSTANCE TRANSFER

We have designed the two transfer learning strategies introduced in Section 3.2. One is directly using the RBF top n (here we set $n=800$) similar sentences from source domain as new additional target domain training data. The other is increasing the source corpus by Eq. 8, i.e. the n th most similar sentences repeat k times. Our preliminary results show that the latter is better. Therefore, due to page limit, we just discuss the latter in this paper. Finally, 41500 sentences are obtained as labeled data. We still keep the 80%/20% train/test proportion when using the enlarged training data to train our ERNN model. The parameters are given in Equation 8.

$$\begin{cases} k = 80 & \text{if } 1 \leq or \leq n \leq or \leq 250 \\ k = 50 & \text{if } 250 \leq n \leq or \leq 500 \\ k = 30 & \text{if } 500 \leq n \leq or \leq 800 \\ k = 1 & \text{if } 800 \leq n \end{cases} \quad (8)$$

Since in the above experiment, RNN show better performance than HMM and CRF, RNN related baselines and our approach are listed as follows (Here ‘IT’ means ‘Instance Transfer’). As discussed in Section 2, most of current researches that study deep neural network for NER work on trying different variants of deep learning models by assuming that a number of training data is ready. However, we consider two situations. One is that there is no labeled data in target domain, i.e., RNN_D_IT and ERNN_IT. The other is there is a few labeled data in target domain, i.e., RNN_L, RNN_L_D_IT and ERNN_L_IT. Our problem context is quite different from most of current work.

1. RNN_D_IT: train a traditional RNN model (left part in Figure 2) directly with source domain. This training process is intuitive and any NER model can be applied like this.

2. ERNN_IT: train our ERNN model (Figure 3) with selected instances from source domain by our instance transfer strategy (Eq. 8).

3. RNN_L: train a traditional RNN model (left part in Figure 2) with a few labeled data from target domain. This is a traditional training process for supervised learning and more results are discussed in Section 4.4.

4. RNN_L_D_IT: train a traditional RNN model (left part in Figure 2) with a few labeled data from target domain and all instances in source domain. The problem context is the same as that of [22], but RNN is used in our paper instead of CRF.

TABLE 3. P, R and F1 scores without labeled data.

Experiments	P(%)	R(%)	F1(%)
RNN_D_IT	76.11	85.48	80.52
ERNN_IT	92.67	93.90	93.28

TABLE 4. P, R and F1 scores with a few labeled data.

Experiments	P(%)	R(%)	F1(%)
RNN_L	91.84	92.42	92.13
RNN_L_D_T	94.57	91.60	93.06
ERNN_L_IT	93.16	94.47	93.81

5. ERNN_L_IT: train our ERNN model (Figure 3) with a few labeled data from target domain and transferred instances in source domain by our instance transfer strategy (Eq. 8).

For the situation that there is no labeled data in target domain, the experimental results are reported in Table 3. The results of RNN_D_IT tell us that directly using all the labeled data from source domain directly is not much satisfactory. With the help of our instance transfer strategy, similar sentences are selected and repeatedly used as training data. The F1 measure of our ERNN_IT reaches at 93.28%, with the improvement about 15.84%.

In addition, the other situation for domain NER is considered where a few labeled data is available. Therefore, experiments are conducted by adding our manually labeled data into RNN_D_IT and ERNN_IT. Experimental results are show in Table 4. Traditional RNN model gets the F1 value of 92.13, which further indicates that manually labeled data for target domain can largely improve the quality of domain NER. Under this situation, our ERNN_L_IT can still obtain the F1 value of 93.81, and it benefits from our designed instance transfer strategy and adding a particular layer to RNN. The F1 value of RNN_L_D_T is 93.06, which says that source domain can help target domain NER. Transfer strategy should be studied. The results in Table 3 and Table 4 tell us that transfer learning is an effective way to leverage the labeled data from source domain corpus.

2) CO-TRAINING

The experiments in instance transfer part aim to utilize the source domain data to raise the recognition performances. Along with instance transfer learning methods, we also consider taking advantages of unannotated data since we have much of them at hand. Specifically, we conduct experiments by co-training our ERNN and a statistical probability model i.e. CRF. Co-training is kind of a semi-supervised strategy that is firstly proposed by [4] in 1998 and the conditional independence of the data views is declared as a required criterion. However, [1] shows that the independence assumption can be relaxed, which means co-training is still effective under a weaker independence assumption. In this paper, we explore the effect of co-training with our ERNN and adopt it to leverage the large unannotated in-domain data.

TABLE 5. P, R and F1 score of ERNN and CRF in each iteration of co-training.

CRF (%)			ERNN (%)		
P	R	F1	P	R	F1
94.77	84.57	89.38	93.16	94.47	93.81
95.47	84.95	89.90	92.50	94.96	93.71
95.57	86.25	90.67	93.84	93.39	93.61
96.18	86.49	91.08	93.92	93.88	93.90
96.02	86.86	91.21	93.04	94.81	93.92
96.70	86.57	91.36	95.04	93.01	94.02
96.61	87.12	91.62	94.23	93.75	93.99
96.89	87.31	91.85	94.51	93.17	93.83
96.65	87.77	92.00	94.31	93.70	94.00
96.92	88.09	92.30	95.16	91.68	93.39

The initial training data size is 2035 sentences here, while the unannotated data is 12541 sentences. We select top 800 high confidence level sentences in each iteration and after iterated 10 times, all 12541 sentences get labeled. Table 5 shows the results of each model in each iteration.

For the co-training experiments, large unannotated data also brings rise to the quality of NER, but not as much as that of instance transfer experiments. Due to the CRF's low starting point, the ERNN F1 score goes down after reaching 94.02 in terms of F1 measure.

V. CONCLUSION AND FUTURE WORK

Domain NER is important and useful in various applications. In this paper, we study instance transfer and RNN to improve the quality of domain NER. We leverage the source domain data by proposing an instance transfer enhanced RNN called ERNN. In addition, we adopt co-training strategy to leverage the large unannotated in-domain data to further improve the recognition performances. In the future, we are going to make a further exploration to automatic QA system, relationship extraction between entities.

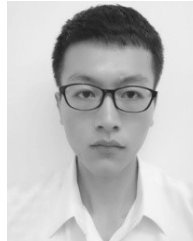
REFERENCES

- [1] S. P. Abney, "Bootstrapping," in *Proc. 40th Annu. Meeting Assoc. Comput. Linguistics*, Philadelphia, PA, USA, Jul. 2002, pp. 360–367.
- [2] M. Amith, Y. Zhang, H. Xu, and C. Tao, "Knowledge-based approach for named entity recognition in biomedical literature: A use case in biomedical software identification," in *Proc. 30th Int. Conf. Ind. Eng. Other Appl. Appl. Intell. Syst. (IEA/AIE)*, Arras, France, Jun. 2017, pp. 386–395.
- [3] A. Arnold, R. Nallapati, and W. W. Cohen, "Exploiting feature hierarchy for transfer learning in named entity recognition," in *Proc. 46th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, Columbus, OH, USA, Jun. 2008, pp. 245–253.
- [4] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory (COLT)*, Madison, WI, USA, Jul. 1998, pp. 92–100.
- [5] J. Brooke, A. Hammond, and T. Baldwin, "Bootstrapped text-level named entity recognition for literature," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics*, Berlin, Germany, vol. 2, Aug. 2016, pp. 344–350.
- [6] Y. Chen, T. A. Lasko, Q. Mei, J. C. Denny, and H. Xu, "A study of active learning methods for named entity recognition in clinical text," *J. Biomed. Informat.*, vol. 58, pp. 11–18, Dec. 2015.
- [7] Y. K. Chen, "A preliminary study of coupling transfer learning with active learning for clinical named entity recognition between two institutions," in *Proc. Amer. Med. Inform. Assoc. Annu. Symp. (AMIA)*, Washington, DC, USA, Nov. 2014, pp. 15–19.
- [8] H. Chung, S. J. Lee, and J. G. Park, "Deep neural network using trainable activation functions," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Vancouver, BC, Canada, Jul. 2016, pp. 348–352.
- [9] G. Crichton, S. Pyysalo, B. Chiu, and A. Korhonen, "A neural network multi-task learning approach to biomedical named entity recognition," *BMC Bioinf.*, vol. 18, no. 1, pp. 368:1-368:14, Aug. 2017.
- [10] A. Eiselt and A. Figueroa, "A two-step named entity recognizer for open-domain search queries," in *Proc. 6th Int. Joint Conf. Natural Lang. Process. (IJCNLP)*, Nagoya, Japan, Oct. 2013, pp. 829–833.
- [11] D. Jain, "Supervised named entity recognition for clinical data," in *Proc. Conf. Labs Eval. Forum*, Toulouse, France, Sep. 2015, pp. 8–11.
- [12] C. Jochim and L. Deleris, "Named entity recognition in the medical domain with constrained CRF models," in *Proc. 15th Conf. Eur. Chapter Assoc. Comput. Linguistics*, Valencia, Spain, vol. 1, Apr. 2017, pp. 839–849.
- [13] W. Karara and T. Slimani, "A new approach for arabic named entity recognition," *Int. Arab J. Inf. Technol.*, vol. 14, no. 3, pp. 332–338, 2017.
- [14] Y. Li, H. Huang, X. Zhao, and S. Shi, "Named entity recognition based on bilingual co-training," in *Proc. 14th Workshop Chin. Lexical Semantics (CLSW)*, Zhengzhou, China, May 2013, pp. 480–489.
- [15] S. S. Liew, M. Khalil-Hani, and R. Bakhteri, "Bounded activation functions for enhanced training stability of deep neural networks on visual pattern recognition problems," *Neurocomputing*, vol. 216, pp. 718–734, Dec. 2016.
- [16] G. Liu, Y. Yan, R. Subramanian, J. Song, G. Lu, and N. Sebe, "Active domain adaptation with noisy labels for multimedia analysis," *World Wide Web*, vol. 19, no. 2, pp. 199–215, May 2015.
- [17] G. Mesnil, X. He, L. Deng, and Y. Bengio, "Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding," in *Proc. 14th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Lyon, France, Aug. 2013, pp. 3771–3775.
- [18] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur, "Recurrent neural network based language model," in *Proc. 11th Annu. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Chiba, Japan, Sep. 2010, pp. 1045–1048.
- [19] T. Munkhdalai, M. Li, T. Kim, O.-E. Namsrai, S.-P. Jeong, J. Shin, and K. H. Ryu, "Bio named entity recognition based on co-training algorithm," in *Proc. 26th Int. Conf. Adv. Inf. Netw. Appl. Workshops*, Fukuoka, Japan, Mar. 2012, pp. 857–862.
- [20] G. Murugesan, S. Abdulkadhar, B. Bhasuran, and J. Natarajan, "BCC-NER: Bidirectional, contextual clues named entity tagger for gene/protein mention recognition," *EURASIP J. Bioinf. Syst. Biol.*, vol. 2017, no. 1, p. 7, May 2017.
- [21] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.
- [22] L. Qu, G. Ferraro, L. Zhou, W. Hou, and T. Baldwin, "Named entity recognition for novel types by transfer learning," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, Austin, TX, USA, Nov. 2016, pp. 899–905.
- [23] E. F. Sang and F. De Meulder, "Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition," in *Proc. 7th Conf. Natural Lang. Learn. (CoNLL)*, Edmonton, AB, Canada, May/June 2003, pp. 142–147.
- [24] G. A. Şeker and G. Eryiğit, "Extending a CRF-based named entity recognition model for Turkish well formed text and user generated content 1," *Semantic Web*, vol. 8, no. 5, pp. 625–642, Apr. 2017.
- [25] D. Song, S. Li, L. Jin, and D. Huang, "Biomedical named entity recognition based on recurrent neural networks with different extended methods," *Int. J. Data Mining Bioinf.*, vol. 16, no. 1, pp. 17–31, 2016.
- [26] H. Sun, R. Grishman, and Y. Wang, "Domain adaptation with active learning for named entity recognition," in *Proc. 2nd Int. Conf. (ICCCS)*, Nanjing, China, Jul. 2016, pp. 611–622.
- [27] S. Tomori, T. Ninomiya, and S. Mori, "Domain specific named entity recognition referring to the real world by deep neural networks," in *Proc. 54th Annu. Meeting Assoc. Comput. Linguistics. (ACL)*, Berlin, Germany, vol. 2, Aug. 2016, pp. 236–242.
- [28] K. N. Vavliakis, A. L. Symeonidis, and P. A. Mitkas, "Event identification in Web social media through named entity recognition and topic modeling," *Data Knowl. Eng.*, vol. 88, pp. 1–24, Nov. 2013.
- [29] Y. Wang, X. Lin, L. Wu, W. Zhang, Q. Zhang, and X. Huang, "Robust subspace clustering for multi-view data by exploiting correlation consensus," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3939–3949, Nov. 2015.
- [30] Y. Wang and L. Wu, "Beyond low-rank representations: Orthogonal clustering basis reconstruction with optimized graph structure for multi-view spectral clustering," *Neural Netw.*, vol. 103, pp. 1–8, Jul. 2018.
- [31] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, Oct. 2018.

- [32] A. Xing, Q. Zhao, and Y. Yan, "Speeding up deep neural networks in speech recognition with piecewise quantized sigmoidal activation function," *IEICE Trans. Inf. Syst.*, vol. E99.D, no. 10, pp. 2558–2561, 2016.
- [33] Y. Yao and A. Sun, "Mobile phone name extraction from Internet forums: A semi-supervised approach," *World Wide Web*, vol. 19, no. 5, pp. 783–805, Jul. 2015.
- [34] D. Zeng, C. Sun, L. Lin, and B. Liu, "LSTM-CRF for drug-named entity recognition," *Entropy*, vol. 19, no. 6, p. 283, Jun. 2017.
- [35] Y. Zhang, J. Xu, H. Chen, J. Wang, Y. Wu, M. Prakasam, and H. Xu, "Chemical named entity recognition in patents by domain knowledge and unsupervised feature learning," *Database*, vol. 2016, Apr. 2016, Art. no. baw049.
- [36] Z. Xie, Z. Zeng, G. Zhou, and W. Wang, "Topic enhanced deep structured semantic models for knowledge base question answering," *Sci. China Inf. Sci.*, vol. 60, no. 11, pp. 110103:1–110103:15, Sep. 2017.



CHAOJIE FAN was born in 1995. He received the B.E. degree from the Wuhan University of Technology, where he is currently pursuing the master's degree. His research interests include dynamic testing technology and signal processing.



ZHENGJU WANG was born in 1995. He received the B.E. degree from Liaoning Technical University. He is currently pursuing the master's degree with the Wuhan University of Technology. His research interests include optimization of engine heat dissipation performance and modern design theory.



CHUANBO LIU was born in 1972. He received the Ph.D. degree from the Wuhan University of Technology, China, in 2009. He is currently an Associate Professor with the Wuhan University of Technology. His research interests include human robot interaction, mechanical design and theory, and vehicle thermal management.



YUEQING SUN was born in 1990. She is currently pursuing the master's degree. Her research interests include deep learning and natural language processing.

...