An Empirical Study of Multi-domain and Multi-task Learning in Chinese Named Entity Recognition

Yun Hu[1,2(✉)], Mingxue Liao[2], Pin Lv[2], and Changwen Zheng[2]

[1] University of Chinese Academy of Sciences, Beijing, China
[2] Institute of Software, Chinese Academy of Sciences, Beijing, China
{huyun2016,mingxue,lvpin,changwen}@iscas.ac.cn

Abstract. Named entity recognition (NER) often suffers from lack of annotation data. Multi-domain and multi-task learning solve this problem in some degree. However, previous multi-domain and multi-task learning are often studied in English. In the other part, multi-domain and multi-task learning are often researched independently. In this manuscript, we first summarize the previous works of multi-domain and multi-task learning in NER. Then, we introduce the multi-domain and multi-task learning in Chinese NER. Finally, we explore the universal models between multi-domain and multi-task learning. Experiments show that the universal models can be used in Chinese NER and outperform the baseline model.

Keywords: Chinese named entity recognition · Multi-domain learning · Multi-task learning

1 Introduction

Name entity recognition is a fundamental Natural Language Processing task. The NER system labels each word in sentences with predefined types, such as Person (PER), Location (LOC), Organization (ORG) and so on. The results of NER can be used in many downstream NLP tasks, such as question answering [27] and relation extraction [1]. The neural network methods [4,10] are used to realize the NER system recently. Large annotated data is required in neural network methods. However, the annotated data is usually scarce.

In order to improve the performance of NER system in low resource, multi-domain and multi-task methods are often used [2,7,18,26]. Multi-domain learning tries to transfer information from the source domain to the target domain [7]. Multi-task learning tries to transfer information from the source task to the target task [2].

There existing some challenges in previous works. First, most of the previous models only test in English. Can the models work well in Chinese NER? For example, the tasks are part-of-speech (POS) tagging and named entity recognition in English

The work is supported by both National scientific and Technological Innovation Zero (No. 17-H863-01-ZT-005-005-01) and State's Key Project of Research and Development Plan (No. 2016QY03D0505).

I. V. Tetko et al. (Eds.): ICANN 2019, LNCS 11728, pp. 743–754, 2019.
https://doi.org/10.1007/978-3-030-30484-3_58

multi-task learning [4], however, the tasks are Chinese Word Segmentation (CWS) and named entity recognition in Chinese multi-task learning [17]. Second, previous works often consider the multi-domain methods and multi-task methods separately. For example, Cao et al. only consider the multi-task learning [2]. Can the multi-task models be directly used in multi-domain, vice versa?

In this manuscript, we do an empirical study in Chinese NER. First, we summarize the previous multi-domain and multi-task models according to the model architecture. The neural network methods are considered in this manuscript. Second, we suppose that the multi-domain and multi-task learning methods are independent of languages. We use Chinese social media domain as the target domain and Chinese news domain as the source domain. The Chinese NER task is the target task and CWS task is the source task. These domains and tasks are similar and the information can be transferred. Third, we suppose that the methods used in multi-domain and multi-task are similar. The methods come from transfer learning. The methods used in multi-domain can be directly used in multi-task, vice versa. In other words, the model architecture is not required to be changed when the model is used in multi-domain or in multi-task, and only the data is required to be changed. Three types of universal model architectures are demonstrated: SHA (share model), FEAT (feature used model) and ADV (adversarial network model). Experiments show that the universal models are useful in Chinese NER and outperform the baseline model.

Specifically, we make contributions as follows:

- We summarize the previous multi-domain and multi-task models in NER.
- We explore the performance of the multi-domain and multi-task methods in Chinese NER task.
- We demonstrate three types of universal model architectures in multi-domain and multi-task learning.

2 Overview

2.1 Previous Summaries of NER

The existing surveys mainly focus on summarizing the methods used in named entity recognition, including supervised, semi-supervised, and unsupervised methods [14,20]. Yadav et al. provided recent advances in NER from deep learning models [25]. The transfer learning surveys mainly focus on general methods in multi-domain and multi-task learning [15]. Tan et al. presented a survey of deep transfer learning [21]. Compared with the previous summaries of NER, we focus on multi-domain and multi-task learning in Chinese NER.

2.2 Domains and Tasks in Multi-domain and Multi-task NER

Previous works show that multi-domain and multi-task learning improve the performance of English NER [4,12]. In multi-domain and multi-task learning, the target domains and tasks are often similar to the source domains and tasks. In English, the target domain is often twitter domain and the source domain is news domain. The target task is NER and the source task is chunk or POS. In this manuscript, the source domain

is Chinese news domain, and the target domain is Chinese weibo domain. The source task is CWS, and the target task is NER. An example is shown in Fig. 1. We suppose that Chinese weibo NER is similar to Chinese news NER. The weibo NER and news NER are the same task and use different domain data. Some tokens and labels are the same in two domains. For example, the "洛阳机场" is labeled as "LOC" in both news and weibo domains. We suppose that CWS task is similar to the NER. The CWS and NER all belong to sequence labeling task. CWS tries to find the word boundary. For example, "洛阳机场" is an independent word. NER tries to find the word boundary and types. For example, "洛阳机场" is seen as an independent word and the entity type of "洛阳机场" is "LOC".

嘿嘿,坐标洛阳,想看岳云鹏说的洛阳机场天价面
Heihei, in Luoyang, want to see the high-price noodles which Yue Yunpeng ate before in Luoyang Airport

副省长张大伟一行来到洛阳,视察洛阳机场航站楼扩建工程
Vice Governor Zhang Dawei arrived in Luoyang and checked up on Terminal Extension Project of Luoyang Airport

嘿 嘿 ， 坐 标 洛 阳 ， 想 看 岳 云 鹏 说 的
O O O O O B-GPE I-GPE O O B-PER I-PER I-PER O O

洛 阳 机 场 天 价 面
B-LOC I-LOC I-LOC I-LOC O O O

嘿嘿 ， 坐标 洛阳 ， 想看 岳云鹏 说的 洛阳机场 天价面
B E S B E B E S B E B I E B E B I I E B I E

副 省 长 张 大 伟 一 行 来 到 洛 阳 ，
O O O B-PER I-PER I-PER O O O O B-LOC B-LOC O

视 察 洛 阳 机 场 航 站 楼 扩 建 工 程
O O B-LOC I-LOC I-LOC I-LOC O O O O O O O

Fig. 1. The first block is the Chinese-English translation pair for understanding. The second block is from weibo NER. The third block is from weibo CWS. The fourth block is from news NER.

2.3 Methods in Multi-domain and Multi-task NER

A list of neural multi-domain and multi-task learning models are shown in Table 1. The multi-domain and multi-task models are divided into four types: SHA, FEAT, ADV and BV (variant of base model). The SHA model is prevalent in previous works [4,8,18,26]. The multi-task learning of English named entity recognition was first proposed by Collobert et al. using neural network model [4]. Lee et al. trained the model using the source data and retrained the model using the target data [8]. Yang et al. explored the transferring module in multi-domain and multi-task separately [26]. Peng and Dredze used domain projection and specific task Conditional Random Fields (CRF) combining the multi-domain and multi-task [18]. The FEAT model was first proposed by [17] used for multi-task in NER. Cao et al. used the adversarial network to integrate the task-shared word boundary information into Chinese NER task [2]. The BV models are

variant of base model. The BV models can not be directly used between multi-domain and multi-task.

Table 1. A summary of the multi-domain and multi-task learning.

	Multi-domain	Multi-task	Model
Yang et al. [26]	English	English	SHA
Peng and Dredze [18]	Chinese	Chinese	SHA
Lee et al. [8]	English	–	SHA
Collobert et al. [4]	–	English	SHA
Peng and Dredze [17]	–	Chinese	FEAT
Cao et al. [2]	–	Chinese	ADV
Peng and Dredze [16]	–	Chinese	BV
Changpinyo et al. [3]	–	English	BV
He and Sun [7]	Chinese	–	BV
Wang et al. [22]	Chinese	–	BV
Lin et al. [12]	English	–	BV

3 Model

3.1 Module

All the models are composed by some basic modules. We discuss the basic modules first. Four types of modules are considered: Character embedding, Bi-LSTM, CRF and Classifier.

Character Embedding. Character embedding is the first step of neural network models in Chinese NER. Character embedding is similar to the word embedding in English. For example, "洛" is a character and is mapped to a low dimension vector in Character embedding layer. Pre-trained character embedding is often used to utilize the information from the large unannotated data. For a sequence of character $c = \{x_1, x_2, ..., x_n\}$, we obtain $x = \{e_{x_1}, e_{x_2}, ..., e_{x_n}\}$ though looking up pre-trained character embedding.

Bi-LSTM. Bi-LSTM is used to extract the features from the sentence. The Bi-LSTM concatenates the forward LSTM output and backward LSTM output as the final output and can capture the information of a character from right context and left context [6].

A common LSTM unit is composed of a cell, an input gate, an output gate and a forget gate. The implementations of LSTM are as follows:

$$i_t = \sigma(W_i h_{t-1} + U_i e_{x_t} + b_i) \tag{1}$$
$$f_t = \sigma(W_f h_{t-1} + U_f e_{x_t} + b_f) \tag{2}$$
$$\tilde{c_t} = tanh(W_c h_{t-1} + U_c e_{x_t} + b_c) \tag{3}$$
$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c_t} \tag{4}$$
$$o_t = \sigma(W_o h_{t-1} + U_o e_{x_t} + b_o) \tag{5}$$
$$h_t = o_t \odot tanh(c_t) \tag{6}$$

where e_{x_t} is the input vector at time t, h_t is the output of LSTM model, σ is the element-wise sigmoid function, and $\odot$ is the element-wise product.

CRF. The CRF is used to predict the label sequence $y = \{y_1, y_2, ..., y_n\}$. The CRF uses the feature extracted by the Bi-LSTM and considers the neighborhood information in a sequence to make prediction. We define the source s of the sentence when X is used as the input sequence list and y is used as the output NER tag list:

$$s(X, y) = \sum_{i=0}^{n} A_{y_i, y_{i+1}} + \sum_{i=1}^{n} P_{h_i, y_i} \tag{7}$$

where $A_{y_i, y_{i+1}}$ describes the cost from tag y_i transferring to y_{i+1}, and P represents the probability from h_i predicting the tag y_i. The probability of tag sequences can be represented as:

$$P(y|X) = \frac{e^{s(X,Y)}}{\sum_{\tilde{y} \in Y_{all}} e^{s(X,\tilde{y})}} \tag{8}$$

where $\tilde{y}$ is the possible NER tags and Y_{all} is all the possible NER tags. When the model is trained, we maximize the log-probability of the correct sequence:

$$logP(y|X) = s(X, y) - log(\sum_{\tilde{y}} e^{s(X,\tilde{y})}) \tag{9}$$

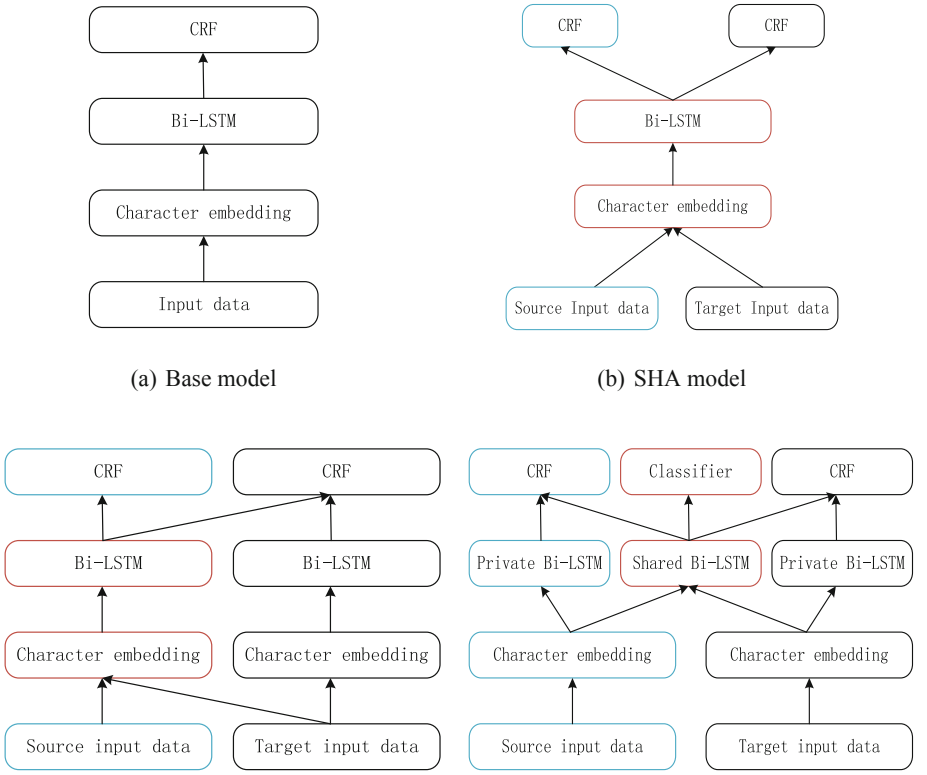
When the model is tested, we can obtain the best NER tag sequence y^* by:

$$y^* = \underset{\tilde{y} \in Y_{all}}{\text{argmax}}\, s(X, \tilde{y}) \tag{10}$$

Classifier. For the multi-domain models, the classifier discriminates the sentence from news domain or weibo domain. For the multi-task models, the classifier discriminates the sentence from NER or CWS. The classifier contains maxpooling and softmax.

$$h = Maxpooling(H) \tag{11}$$
$$D(h, \theta_d) = softmax(W_d h + b_d) \tag{12}$$

where H is the feature representation of the sentences and θ_d is the parameters in softmax, including W_d and b_d.



(a) Base model (b) SHA model

(c) FEAT model (d) ADV model

Fig. 2. Four types of architectures in multi-domain and multi-task learning. The blue block is source part, the black block is target part, and the red block is the share part. (Color figure online)

3.2 Base Model

We use LSTM-CRF model as the base model which is widely used for a single domain and task NER [5,10]. The architecture of the model is shown in Fig. 2(a). The model contains three parts. The character embedding part is used to utilize the word level features from the large unannotated data. The Bi-LSTM is used to extract the sentence level features. The features are fed into CRF to predict the sequence labels. Compared with the multi-domain and multi-task model, the LSTM-CRF model is a single task and domain model, which only uses weibo NER dataset as input.

3.3 Multi-domain and Multi-task Model

Multi-domain and multi-task methods are shown to improve the single task and domain NER performance. However, most of the models only focus on English NER.

In the other part, most of the models are only used in one specific situation: in multi-domain learning or in multi-task learning. In this manuscript, we do an empirical study in 3 types of Chinese named entity recognition models: SHA, FEAT and ADV. For the specificity of the BV model, we will do the experiments in the future. In the multi-domain and multi-task model, the models use two types of data as inputs. For the multi-domain model, the weibo NER dataset and news NER dataset are used. For the multi-task model, the weibo NER dataset and weibo CWS dataset are used.

SHA. The SHA model shares the feature extractor between different tasks or domains [4,8,18,26]. The feature extractor trained by source data can contain useful information for target task or domain. The architecture of the SHA model is shown in Fig. 2(b). The character embedding layer and the Bi-LSTM layer are shared. Different domains or tasks have specific CRFs. With different training methods, the SHA model can be divided into three different sub-models.

SHA-INIT. The training method is divided into two steps. First, the model uses the source data as the input to train the model until convergence. In the second step, target data is fed as the input of the model to continually train until convergence. Two steps use different CRFs. The parameters in character embedding layer and the Bi-LSTM layer are all updated in two steps.

SHA-CRF. The training method is the same as SHA-INIT except for the second step. In the second step, the parameters of character embedding and Bi-LSTM layer are frozen and we only update the parameters in CRF layer.

SHA-MUL. The model trains the source data and target data simultaneously. In one epoch, source data and target data are fed to the model alternatively. A hyper-parameter α can be used to control the size between source data and target data.

FEAT. The FEAT model supposes that the features extracted by the source part can be used as auxiliary information for the target part [17]. The architecture of the model is shown in Fig. 2(c). The target part uses the intermedia results of the source part. The Bi-LSTM output of the target part is concatenated with the output of the source part before being fed into CRF layer. Three different training methods lead to three sub-models.

FEAT-INIT. The model first uses the source data to train a base model. Then, the source part of the model is initialized by the pre-trained model. Finally, the target data is used to train the model. All parameters are updated in the model.

FEAT-CRF. The training method is the same as FEAT-INIT except for the final parameters updating step. In the FEAT-CRF model, the source part is initialized by the pre-trained model and then the parameters are frozen. The model only updates the target part parameters.

FEAT-MUL. The source part and the target part are trained alternatively. When the source data is used as input, the parameters in the source part are updated. When the target data is used as input, the parameters in the source part and target part are updated.

ADV. The ADV model uses the private feature extractor extracts the private information, and uses the shared feature extractor extract the shared information [2]. The architecture of the model is shown in Fig. 2(d). The model uses private character embedding,

Bi-LSTM and CRF to capture the different information between source domain and target domain, and uses shared Bi-LSTM to capture the common information between source domain and target domain. Character embedding tries to capture the word level representation. Bi-LSTM tries to extract the sentence level feature representation of the words. The classifier tries to guarantee that specific features of tasks do not exist in shared space. The source data and target data are fed to the model alternatively.

BV. The BV models are the models that can not find the universal models in multi-domain and multi-task learning [3,7,12,22]. For example, Wang et al. required the source domain and the target domain has the same label sets [22]. However, different tasks have different label sets. Lin et al. used a domain adaptation layer to reduce the disparity between different pre-trained character embeddings [12]. However, pre-trained embeddings are the same in different tasks.

4 Experiments and Results

4.1 Datasets

The Chinese weibo NER corpus is from [16]. The Chinese news NER corpus is from Sighan NER [11]. The Chinese weibo word segmentation corpus is from [19]. The sentence numbers of the different corpora are shown in Table 2.

Table 2. The details of the corpora.

	#Train	#Dev	#Test
Weibo NER	1350	270	270
News NER	16814	1868	4636
Weibo CWS	38086	3834	16673

4.2 Parameters Setting

The character embedding is initialized by pre-trained character embedding. The news embedding is pre-trained on Chinese Wikipedia data using word2vec [13]. The weibo embedding is pretrained on Chinese social media data using word2vec. The embedding dimension is 100. The LSTM dimension in both source and target part are 100. The optimization method we used is adam [9].

4.3 Results

The overview results of Chinese weibo NER are shown in Table 3. The results show that the multi-domain and multi-task learning can be used in Chinese named entity recognition. The FEAT and ADV models can always outperform the baseline model. The SHA-CRF model obtains the worst F1 score which is far worse than the baseline

model. The reason may be that the CRF is hard to process the features extracted by a source feature extractor. Meanwhile, the experiments show that multi-domain and multi-task learning can use the universal models. The same model architecture can directly be used in both multi-domain and multi-task learning.

Table 3. The overview results of multi-domain and multi-task learning in Chinese weibo NER. P represents precision, R represents recall, and F represents F1 score.

	Multi-domain			Multi-task		
	P	R	F	P	R	F
Base	56.07	44.50	49.62	56.07	44.50	49.62
SHA-INIT	58.06	41.28	48.25	62.00	42.66	50.54
SHA-CRF	65.52	17.43	27.53	69.49	18.81	29.61
SHA-MUL	61.64	46.39	52.94	60.65	48.45	53.87
FEAT-INIT	53.55	49.11	51.23	62.93	44.13	51.88
FEAT-CRF	59.64	45.41	51.56	62.24	46.77	53.41
FEAT-MUL	59.86	45.36	51.61	61.39	50.00	55.11
ADV	57.06	52.06	54.45	60.92	47.42	53.33

In Table 4, we make a survey in the performance of previous works in Chinese weibo NER. Compared with previous performances, the three types of models achieve competitive results.

Table 4. The performance of previous Chinese weibo NER models.

	P	R	F	Model
Peng and Dredze [17]	66.67	47.22	55.28	FEAT
Cao et al. [2]	59.51	50.00	54.34	ADV
Peng and Dredze [16]	74.78	39.81	51.96	BV
He and Sun [7]	52.94	51.18	52.05	BV

To show the generalization of the models in Chinese NER, the Ontonote NER dataset is also tested [23]. The broadcast news domain is used as the source domain, and the web text domain is used as the target domain. The Chinese weibo word segmentation is used as the source task, and the Chinese Ontonote web text named entity recognition is used as the target task. The broadcast news data contains 10083 sentences, and the web text contains 8405 sentences. In Table 5, the results show that the universal models can be used in different datasets.

5 Discussion

Experiments show that the multi-domain and multi-task learning can improve the performance of Chinese NER. More works can be done in the future.

Table 5. The overview results of multi-domain and multi-task learning in Ontonote dataset.

	Multi-domain			Multi-task		
	P	R	F	P	R	F
Base	48.23	44.60	46.34	48.23	44.60	46.34
SHA-INIT	48.39	43.45	45.79	47.84	44.26	45.98
SHA-CRF	36.34	25.17	29.74	30.31	19.43	23.68
SHA-MUL	59.78	55.24	57.42	62.93	44.85	52.37
FEAT-INIT	52.79	49.49	51.09	51.56	47.70	49.55
FEAT-CRF	51.90	51.15	51.52	52.45	47.75	49.99
FEAT-MUL	53.32	58.37	55.73	57.24	45.52	50.71
ADV	55.40	55.60	55.50	50.56	50.47	50.51

First, Chinese specific features can be considered in multi-domain and multi-task learning. In this manuscript, we only use the Chinese models which is similar to the English models. Some Chinese specific features are shown very helpful in Chinese NER, such as radical feature [5], glyph representation of Chinese character [24]. We will explore these features in multi-domain and multi-task learning.

Second, the BV model can be universal model through small changes. For example, some models require the same labels in source and target domain. The requirement can extend to that the source and target domain labels are related. The model architecture required small changes in the future.

Third, the multi-domain and multi-task have high similarity. Two works can be combined together. For example, Peng and Dredze used domain projection and specific task CRF combining the multi-domain and multi-task [18]. However, the Peng and Dredze only processed the situation that the domains have the same label sets. In the future, more general models could be considered.

6 Conclusion

In this manuscript, we focus on utilizing Chinese news domain information and Chinese word segmentation information to improve the performance of Chinese weibo named entity recognition by multi-domain and multi-task learning. Three types of universal model architectures are explored. Experiments show that the universal models outperform the baseline model.

References

1. Bunescu, R., Mooney, R.: A shortest path dependency kernel for relation extraction. In: Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (2005). http://aclweb.org/anthology/H05-1091
2. Cao, P., Chen, Y., Liu, K., Zhao, J., Liu, S.: Adversarial transfer learning for Chinese named entity recognition with self-attention mechanism. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 182–192. Association for Computational Linguistics (2018). http://aclweb.org/anthology/D18-1017
3. Changpinyo, S., Hu, H., Sha, F.: Multi-task learning for sequence tagging: an empirical study. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 2965–2977. Association for Computational Linguistics (2018). http://aclweb.org/anthology/C18-1251
4. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., Kuksa, P.: Natural language processing (almost) from scratch. J. Mach. Learn. Res. **12**(Aug), 2493–2537 (2011). http://dl.acm.org/citation.cfm?id=2078186
5. Dong, C., Zhang, J., Zong, C., Hattori, M., Di, H.: Character-based LSTM-CRF with radical-level features for chinese named entity recognition. In: Lin, C.-Y., Xue, N., Zhao, D., Huang, X., Feng, Y. (eds.) ICCPOL/NLPCC -2016. LNCS (LNAI), vol. 10102, pp. 239–250. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-50496-4_20
6. Graves, A., Schmidhuber, J.: Framewise phoneme classification with bidirectional LSTM networks. In: 2005 IEEE International Joint Conference on Neural Networks, IJCNN 2005, Proceedings, vol. 4, pp. 2047–2052 (2005). https://doi.org/10.1016/j.neunet.2005.06.042
7. He, H., Sun, X.: A unified model for cross-domain and semi-supervised named entity recognition in Chinese social media. In: AAAI, pp. 3216–3222 (2017). http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14484
8. Ji, Y.L., Dernoncourt, F., Szolovits, P.: Transfer learning for named-entity recognition with neural networks (2017). http://www.lrec-conf.org/proceedings/lrec2018/summaries/878.html
9. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. Comput. Sci. (2014). http://arxiv.org/abs/1412.6980
10. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., Dyer, C.: Neural architectures for named entity recognition. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 260–270. Association for Computational Linguistics (2016). https://doi.org/10.18653/v1/N16-1030
11. Levow, G.A.: The third international Chinese language processing bakeoff: word segmentation and named entity recognition. In: Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing, pp. 108–117. Association for Computational Linguistics (2006). http://www.aclweb.org/anthology/W06-0115
12. Lin, B.Y., Lu, W.: Neural adaptation layers for cross-domain named entity recognition. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2012–2022. Association for Computational Linguistics (2018). http://aclweb.org/anthology/D18-1226
13. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)
14. Nadeau, D., Sekine, S.: A survey of named entity recognition and classification. Lingvisticae Investigationes **30**(1), 3–26 (2007)

15. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE Trans. Knowl. Data Eng. **22**(10), 1345–1359 (2010)
16. Peng, N., Dredze, M.: Named entity recognition for Chinese social media with jointly trained embeddings. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing, pp. 548–554. Association for Computational Linguistics (2015). http://www.aclweb.org/anthology/D15-1064, https://doi.org/10.18653/v1/D15-1064
17. Peng, N., Dredze, M.: Improving named entity recognition for Chinese social media with word segmentation representation learning. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), pp. 149–155. Association for Computational Linguistics (2016). http://www.aclweb.org/anthology/P16-2025, https://doi.org/10.18653/v1/P16-2025
18. Peng, N., Dredze, M.: Multi-task domain adaptation for sequence tagging. In: Proceedings of the 2nd Workshop on Representation Learning for NLP, pp. 91–100. Association for Computational Linguistics (2017). http://aclweb.org/anthology/W17-2612
19. Qian, P., Qiu, X., Huang, X.: A new psychometric-inspired evaluation metric for Chinese word segmentation. In: Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 2185–2194. Association for Computational Linguistics (2016). http://www.aclweb.org/anthology/P16-1206, https://doi.org/10.18653/v1/P16-1206
20. Sharnagat, R.: Named entity recognition: a literature survey. Center For Indian Language Technology (2014)
21. Tan, C., Sun, F., Tao, K., Zhang, W., Chao, Y., Liu, C.: A survey on deep transfer learning, pp. 270–279 (2018). https://doi.org/10.1007/978-3-030-01424-7_27
22. Wang, Z., et al.: Label-aware double transfer learning for cross-specialty medical named entity recognition. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), pp. 1–15. Association for Computational Linguistics (2018). http://aclweb.org/anthology/N18-1001
23. Weischedel, R., et al.: Ontonotes release 5.0 ldc2013t19. Linguistic Data Consortium, Philadelphia (2013)
24. Wu, W., et al.: Glyce: Glyph-vectors for Chinese character representations. arXiv preprint arXiv:1901.10125 (2019)
25. Yadav, V., Bethard, S.: A survey on recent advances in named entity recognition from deep learning models. In: Proceedings of the 27th International Conference on Computational Linguistics, pp. 2145–2158. Association for Computational Linguistics (2018). http://aclweb.org/anthology/C18-1182
26. Yang, Z., Salakhutdinov, R., Cohen, W.W.: Transfer learning for sequence tagging with hierarchical recurrent networks (2017). https://openreview.net/forum?id=ByxpMd9lx
27. Yao, X., Van Durme, B.: Information extraction over structured data: Question answering with freebase. In: Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp. 956–966. Association for Computational Linguistics (2014). https://doi.org/10.3115/v1/P14-1090