OD-LSTM: An Efficient Network For Named Entity Recognition

Junxiang Ge, Qi Zhang, Haizhou Zhao
AI Research Department, Sogou
Beijing, China

Abstract

With the risen use of AI application, we need to understand some specific type of message from user's query. Named Entity Recognition(NER) is a basic task for that use. Gazzete is very useful when doing NER, howerver, most of the current method use iit in word embedding format, which will bring out the case, we need to retrain model when the word in gazzete is greatly changed. In this paper, we propose a conception, called Outer Dictionary(OD) LSTM, which means to use the outer type info of gazzete for each word, without using the specific word. Experiments shows it useful when doing NER task, and acheive a great improvement on task with which gazzete is heavily depended on.

Introduction

Named Entity Recognition (NER) is a base task when doing many language tasks. Usually, we need to extract mainly information from users' sentence to help us understand their meanings, and then use the extracted information to do further process. For example, in weather domain, we may need to extract **date and position** to help us understand when and where the user is asked towards the weather; in music type service, we may need to track **singer, song, style** entities to suit user's command; for many commom use, we need **person, location, organization** message to understand a sentence.

To accomplish this task, many method has been explored. Convolution Neural Network (CNN) is used in Image Classification (Krizhevsky, Sutskever and Geoffrey 2012), and then proved also powerful in Text Classification (Kim 2014). CNN is also useful in NER tasks, whether to capture character-level features (;Ma and Hovy 2016), or to do word-level sequence labelling (Strubell et al. 2017).

Related Work

pass

Outer Dictionary LSTM

Character-Base LSTM

pass

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Tag Embedding Model

pass

Experiments

pass

Conclusion

pass

Acknowledgments

Especially grateful to Can Cui for denoting the music ner data, to Jindou Wu for advice on data processing of CCKS 2018 music dataset.

References

Krizhevsky, A.; Sutskever, I.; and Geoffrey E. H. 2012. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of Advances in Neural Information Processing Systems 25 (NIPS 2012).

Kim, Y. 2014. Convolutional Neural Networks for Sentence Classification. arXiv:1408.5882.

Sak, H.; Senior, A.; and Beaufays, F. 2014. Long Short-Term Memory Based Recurrent Neural Network Architectures for Large Vocabulary Speech Recognition. arXiv:1402.1128.

Collobert, R.; Weston, J.; Bottou, L.; Karlen, M.; Kavukcuoglu, K.; and Kuksa, P. 2011. Natural Language Processing (Almost) from Scratch. Journal of Machine Learning Research 12 (2011) 2493-2537.

Santos. C. N.; Xiang, B.; and Zhou, B. 2015. Classifying Relations by Ranking with Convolutional Neural Networks. arXiv:1504.06580.

Strubell, E.; Verga, P.; Belanger, D.; and McCallum, A. 2017. Fast and Accurate Entity Recognition with Iterated Dilated Convolutions. arXiv:1702.02098.

Peters, M. E.; Ammar, W.; Bhagavatula, C.; and Power. R. 2017. Semi-supervised sequence tagging with bidirectional language models. arXiv:1705.00108.

Peters, M. E.; Neumann, M.; Iyyer, M.; Gardner, M.; Clark, C.; Lee, K.; and Zettlemoyer, L. 2018. Deep contextualized word representations. arXiv:1802.05365.

Ma, X.; and Hovy, E. 2016. End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF. arXiv:1603.01354.

Sang, E. F.; and Meulder, F. D. 2003. Introduction to the CoNLL-2003 Shared Task: Language-Independent Named Entity Recognition. arXiv:cs/0306050.

Chiu, J. P.C.; and Nichols, E. 2016. Named Entity Recognition with Bidirectional LSTM-CNNs. Transactions of the Association for Computational Linguistics, Volume 4, 2016, p.357-370.

Huang, Z.; Xu, W.; and Yu, K. 2015. Bidirectional LSTM-CRF Models for Sequence Tagging. arXiv:1508.01991.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; and Polosukhin, I. 2017. Attention Is All You Need. arXiv:1706.03762.

Peng, N.; and Dredze, M. 2016. Improving Named Entity Recognition for Chinese Social Media with Word Segmentation Representation Learning. arXiv:1603.00786.

Chen, Q.; Zhuo, Z.; and Wang, W. 2019. BERT for Joint Intent Classification and Slot Filling. arXiv:1902.10909.

Socher, R.; Manning, C. D.; and Andrew Y. Ng. 2010. Learning Continuous Phrase Representations and Syntactic Parsing with Recursive Neural Networks. In Proceedings of the NIPS-2010 Deep Learning and Unsupervised Feature Learning Workshop.

Finkel, J. R.; Grenager, T.; and Manning, C. 2005. Incorporating non-local information into information extraction systems by Gibbs sampling. In Proceedings of the 43nd Annual Meeting of the Association for Computational Linguistics (ACL 2005), pp. 363-370.

Peng, N.; and Dredze, M. 2015. Named Entity Recognition for Chinese Social Media with Jointly Trained Embeddings. In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP).

Mikolov, T.; Chen, K.; Corrado, G.; and Dean, J. 2013. Efficient Estimation of Word Representations in Vector Space. In Proceedings of Workshop at ICLR.

Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.; and Dean, J. 2013. Distributed Representations of Words and Phrases and their Compositionality. In Proceedings of NIPS.

Mikolov, T.; Yih, W.; and Zweig, G. 2013. Linguistic Regularities in Continuous Space Word Representations. In Proceedings of NAACL HLT.

Pennington, J.; Socher, R.; and Manning, C.D. 2014. GloVe: Global Vectors for Word Representation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 15321543.

Mikolov, T.; Grave, E.; Bojanowski, P.; Puhrsch, C.; and Joulin, A. 2017. Advances in Pre-Training Distributed Word Representations. arXiv:1712.09405.