Augmenting Named Entity Recognition with Commonsense Knowledge

Ghaith Dekhili, Ngoc Tan Le, Fatiha Sadat
University of Quebec in Montreal
201 President Kennedy avenue, H2X 3Y7 Montreal, Quebec, Canada
dekhili.ghaith, le.ngoc_tan, sadat.fatiha@uqam.ca

Abstract

Commonsense can be vital in some applications like Natural Language Understanding (NLU), where it is often required to resolve ambiguity arising from implicit knowledge and under-specification. In spite of the remarkable success of neural network approaches on a variety of Natural Language Processing tasks, many of them struggle to react effectively in cases that require commonsense knowledge. In the present research, we take advantage of the availability of the open multilingual knowledge graph ConceptNet, by using it as an additional external resource in Named Entity Recognition (NER). Our proposed architecture involves BiLSTM layers combined with a CRF layer that was augmented with some features such as pre-trained word embedding layers and dropout layers. Moreover, apart from using word representations, we used also character-based representation to capture the morphological and the orthographic information. Our experiments and evaluations showed an improvement in the overall performance with +2.86 in the F1-measure. Commonsense reasonnig has been employed in other studies and NLP tasks but to the best of our knowledge, there is no study relating the integration of a commonsense knowledge base in NER.

1 Introduction

NLP and Machine Learning (ML) communities have long been interested in developing models capable of commonsense reasoning. In addition to that, commonsense can be vital in some applications like NLU, where it is often required to resolve ambiguity arising from implicit knowledge and under-specification by taking word meaning and context into account (Havasi et al., 2010).

ConceptNet (Speer et al., 2016) is a knowledge graph designed to represent the general knowledge involved in understanding languages to improve natural language applications. When word embeddings extracted from ConceptNet, which represent relational knowledge (ConceptNet PPMI (Speer et al., 2016)), are combined with word embeddings acquired from distributional semantics such as Word2Vec (Mikolov et al., 2013), they provide applications with understanding that they would not acquire from distributional semantics alone, nor from narrower resources such as WordNet or DBPedia.

2 Methodology

The present research aims at integrating commonsense knowledge into a Named Entity Recognition (NER) task in order to learn more entities and improve the efficiency and effectiveness of the NER. This research relies on the work presented by Speer et al.(2016), who created a robust set of embeddings (ConceptNet Numberbatch (Speer et al., 2016)), that represents both ConceptNet and distributional word embeddings learned from text. This set of embeddings represents different domains and has complementary strengths. As in Lample et al. (2016), we used in our architecture a Bi-directionnal Long Short-Term Memory (BiLSTM) layers, combined with a Conditional Random Field (CRF) layer (Lafferty et al., 2001), augmented with some features such as pretrained word embedding layers and dropout layers. Apart from using word representations, we used also character-based representation to capture the morphological and the orthographic information. This study aims to compare word embeddings that represent only relational knowledge (ConceptNet PPMI) and their combination with word embeddings that represent only distributional semantics (word2vec and GloVe (Pennington et al., 2014)).

3 Experiments and Evaluations

In this section we describe the data we used in this research as well as evaluations and results.

3.1 Dataset and Training

Our experiments are based on the CoNLL-2003 datasets (Tjong Kim Sang and De Meulder , 2003). The dataset contains four types of named entities: location (LOC), person (PER), organization (ORG) and miscellaneous (MISC) which do not belong to the previous types of entities. We did not use language-specific knowledge or external resources such as gazetteers. Except for replacing every digit with a zero in the dataset, we neither perform any data preprocessing. In our experiments, we re-implemented the architecture inspired from the work of Lample et al. (2016).[1]

3.2 Evaluations and Results

According to our results, we notice on the one hand, that the use of ConceptNet PPMI embeddings improves the F1-score of two NE categories, LOC and PER. On the other hand, the increase of precision of these two entities brought an increase of the overall precision of all NE categories. In contrast, the use of ConceptNet PPMI embeddings did not improve the results of the ORG NE category, which is a common problem in the NER task as organizations can be expressed by acronyms, which are highly ambiguous. However if we analyze results obtained with the use of ConceptNet Numberbatch embeddings, we notice an increase in the precision of the ORG NE category. This can be explained by the use of GloVe and Word2Vec embeddings. Moreover, the precision of the PER category increased remarkably, which led to an improvement of the overall precision.

To conclude, ConceptNet embeddings (PPMI) combined with GloVe and Word2Vec performs better than using them independently. The use of this hybrid embeddings showed a significant improvement with +2.86 in the F1-measure of the overall NER system, as we passed from 84.46% to 87.32% . A combination of the strength of the three pre-trained embeddings, constructed from different domains and data, helped increase the performance of the overall system. Moreover, the small amount of data from which we extracted ConcepNet PPMI embeddings (21 million edges and over 8 million nodes, its English vocabulary contains approximately 1.5 million nodes) comparing to GloVe (840 billion words of the Common Crawl) and word2vec (100 billion words of Google News) can explain the improvement in the performance, especially when using all embeddings. Moreover, using ConceptNet Numberbatch embeddings improved remarkably the performance of each of the four entities taken separately.

4 Conclusion

Prior research on NER has focused on the effectiveness of the combination of BiLSTM and CRF models (Lample et al., 2016). However this research did not try to integrate a commonsense knowledge base to help solve frequent problems in NER such as the ambiguity. A key aspect in our proposed research is the use of the robust ConceptNet Numberbatch in NER. Our results showed an improvement for the LOC and PER NE categories. Using ConceptNet PPMI embeddings, without those of GloVe and Word2Vec, improved LOC's and PER's performance. Most notably, this is the first study, to the best of our knowledge, which evaluated the effectiveness of the use of relational knowledge embeddings combined to distributional semantics embeddings in an English NER task.

In future work, we aim to investigate the usefulness of adding additional layers such as CNN layer to enhance our model structure. Moreover, we would like to investigate the use of unsupervised learning in training a language model on a massive amount of unlabled data as done in Trinh and Le (2018) and deal with the challenge of implicit entities.

[1] `https://github.com/glample/tagger.`

References

Catherine Havasi and Robyn Speer and James Pustejovsky 2010. Coarse Word-Sense Disambiguation Using Common Sense. *AAAI Fall Symposium: Commonsense Knowledge*.

Erik F. Tjong Kim Sang and Fien De Meulder. 2003. Introduction to the CoNLL-2003 Shared Task: Language-independent Named Entity Recognition. *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003*, 142–147.

Guillaume Lample and Miguel Ballesteros and Sandeep Subramanian and Kazuya Kawakami and Chris Dyer. 2016. Neural Architectures for Named Entity Recognition. *Computing Research Repository (CoRR).*

Jeffrey Pennington and Richard Socher and Christopher D. Manning 2014. GloVe: Global vectors for word representation. *12th Proceedings of the Empiricial Methods in Natural Language Processing (EMNLP 2014)*,1532–1543.

John Lafferty and Andrew McCallum and Fernando Pereira 2001. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. *Proceedings of the Eighteenth International Conference on Machine Learning*, 282–289.

Robert Speer and Joshua Chin and Catherine Havasi. 2016. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. *Computing Research Repository (CoRR).*

Tomas Mikolov and Kai Chen and Greg Corrado and Jeffrey Dean 2013. Efficient Estimation of Word Representations in Vector Space. *Computing Research Repository (CoRR).*

Trieu H. Trinh and Quoc V. Le. 2018. A Simple Method for Commonsense Reasoning. *Computing Research Repository (CoRR).*