Metonymy Detection with Deep Neural Networks

Taesun Whang, Daeun Cha, Heuiseok Lim

Korea University, Republic of Korea
{hts920928, chadaeun64, limhseok}@korea.ac.kr

Abstract. In this paper, we propose a deep neural networks based approach for metonymy resolution (MR). MR is a task to detect metonymic expression and clarify its metonymic reading. However, current named entity recognition (NER) taggers cannot describe exact reading of metonymic expression. We utilize long-short term memory (LSTM), bi-directional LSTM (Bi-LSTM), convolutional neural networks (CNN) and BERT for MR. We evaluate the models on three benchmark datasets. Compared with an earlier work, we analyze our result and discover that metonymic reading is strongly affected by adjacent words.

Keywords: Metonymy Resolution, Neural Networks, Classification

1 Introduction

Metonymy is cognitive process in which one conceptual entity, the vehicle, provides mental access to another conceptual entity, the target within the same idealized cognitive model [1]. MR has generated high interest in natural language processing (NLP), especially in NER tasks. In example (1),

(1) China says Taiwan spoils atmosphere for talks.

"China" and "Taiwan" do not refer to the countries but the government. Current NER taggers cannot handle metonymy, so there is no way of detecting the true entities in these cases.

Most dictionaries and corpora do not deal with information about metonymy. This is the main difficulty of MR. However, there has been some research about MR recently [2,3]. Due to the research and their contribution, some datasets with metonymy annotation are available now. In this paper we investigate the possibility of these datasets to resolve the metonymy problem. In particular, we adopt a deep learning based approach, specifically CNN [4], LSTM, and Bi- LSTM. We also use a fine tuning approach with BERT [5] which is state-of-the-art model for NLP tasks.

2 Our Approach

2.1 Recurrent Neural Networks

We create LSTM encoder to represent its contextual representations. The word embeddings of a given sentence, $\mathrm{S} = \{w_1, w_2, \ldots, w_k\}, (1 \leq \mathrm{k} \leq \mathrm{l})$, are fed into uni-LSTM or Bi-LSTM. In the case of Bi-LSTM, sentence representation is produced by concatenating last hidden states of forward LSTM and last hidden states of backward hidden states, while unidirectional LSTM only uses forward hidden states. Bi-LSTM sentence representations are defined by $h_t^w = \mathrm{BiLSTM}(\overrightarrow{h_{t-1}^w}, \overleftarrow{h_{t+1}^w}, \mathrm{w}_t\ )$. Each sequential representation is fed into 1-layer feed-forward neural network to get prediction score.

1.3 Convolutional Neural Networks

We create CNN encoder similar to the previous work [4]. A set of words in a given sentence denoted as $[w_i : w_{i+h-1}]$, where h is a convolution filter size, is multiplied with convolution filters. We apply a max-over-time pooling [6] and get max-pooled feature map $\mathrm{c} = [c_1, c_2, \ldots, c_{l-h+1}]$. Feature map is finally fed into fully connected layer.

2.3 BERT for Metonymy Resolution

BERT [5] language model is pre-trained on a large corpus (e.g., English Wikipedia) with two powerful unsupervised objectives, masked language model (MLM) and next sentence prediction (NSP). BERT can be utilized on various downstream tasks with fine-tuning approach. On the fine-tuning phase, position, segment and token embeddings are added and fed into the BERT layers. The BERT contextual representations of [CLS] token, $T_{[CLS]}$, is utilized to classify whether a given sentence is IsMetonymic or not. We fed $T_{[CLS]}$ to single-layer perceptron to compute the model prediction score following as, $\mathrm{score} = \sigma(W_{task}^T T_{[CLS]} + b)$, where W_{task} is a task-specific trainable parameter.

3 Results

All models are trained and evaluated on two datasets, ReLocaR [3] and SemEval2007, respectively. In Table 1, it is observed that only comparable performance is derived when the training set and test set are same even their tasks are similar. Among our experiments, BERT mostly showed the best performance, CNN also showed high performance for both datasets. Bi- LSTM showed better performance that vanilla LSTM model. None of neural network based models has achieved better results compared to the baseline, which utilized traditional machine learning technique such as SVM. We assume that sentence-level classification is not appropriate for MR, since metonymy is observed in a token-level rather than in a sentence-level.

Table 1. Evaluation results for each dataset (SemEval 2007, ReLocaR).

Model	SemEval2007	ReLocaR
Gritta et al. [3]	83.10	83.60
CNN	82.88	77.43
LSTM	76.58	74.54
Bi-LSTM	77.14	77.43
BERT	81.21	80.38

4 Conclusion

In this work, we conducted several experiments utilizing deep neural networks. While none of the models has achieved state-of-the-art result, we verified the possibility of neural networks approach for MR. Nevertheless, to improve the performance of MR, slightly different point of view for detecting metonymy is needed. For the future work, we handle MR as a sequence labeling task to give an attention on the metonymic word.

Acknowledgement

"This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2018-0-01405) supervised by the IITP(Institute for Information & communications Technology Planning & Evaluation)“

References

1. Klaus-Uwe Panther and G¨unter Radden. 1999. Metonymy in language and thought, volume 4. John Benjamins Publishing.
2. Vivi Nastase, Alex Judea, Katja Markert, and Michael Strube. 2012. Local and global context for supervised and unsupervised metonymy resolution. In EMNLP 2012.
3. Milan Gritta, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. 2017. Vancouver welcomes you! minimalist location metonymy resolution. In ACL 2017
4. Yoon Kim. 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.
5. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In NAACL 2019,
6. Ronan Collobert, Jason Weston, L´eon Bottou, Michael Karlen, Koray Kavukcuoglu, and Pavel Kuksa. 2011. Natural language processing (almost) from scratch. Journal of machine learning research, 12(Aug):2493–2537.