The 3rd International Conference on Interdisciplinary research on Computer science, Psychology, and Education (ICICPE' 2019)
December 17-19, 2019. Phu Quoc, Vietnam

Word Sense Disambiguation with Representations of Context-Sensitive Meaning using Knowledge Graphs

Dongsuk Oh , HeuiSeok Lim

Department of Computer Science and Engineering, Korea University, Seoul 02841, Korea
{inow3555, limhseok}@korea.ac.kr

Abstract. There are two methods of solving word sense disambiguation: knowledge-based method for solving problems using knowledge information and supervised learning method for solving problems using various machine learning models. Supervised learning methods show high performance but require large amounts of refined learning data. In contrast, knowledge-based methods do not require large amounts of refined learning data but cannot expect high performance. Recently, to solve this problem, information in the knowledge and refined learning data are learned in machine learning models to solve the method of word sense disambiguation. In this paper, embedding of syntax information and semantic relation graph information is expressed using GCN(Graph Convolutional Network) to make representation of words containing semantic information and contextual information. It showed higher performance than word representation.

Keywords: Graph Convolutional Network, Word Sense Disambiguation, Memory Network, WordNet, ConceptNet, Word Embedding

1 Introduction

Analyzing words that have more than one meaning in natural language processing is called word sense disambiguation. Just as humans use knowledge and information accumulated through many experiences to communicate, machines must understand sentences through such processes. There are two ways to study word sense disambiguation.

The first is a knowledge-based method for predicting the meaning of words in sentences using lexical knowledge. Knowledge-based methods include dictionary-based methods[1] and graph-based methods.[2,3,4] The dictionary-based method is a method of inferring meaning based on a description of a word defined in a dictionary, and the graph-based method is a method of inferring meaning based on the relationship between words.

The second is a supervised learning method that learns machine learning models and predicts the meaning of words using data labeled with the meaning of words in sentences. [5,6,7,8] The supervised learning method shows high performance because it uses machine learning, but has a disadvantage of constructing a large amount of learning data.

Humans use both of these methods to analyze the meaning of words, but machines do not. However, as deep learning models are being actively researched, models using two methods are being studied.

Luo et al. [8] solved the data shortage problem by using Gloss information of WordNet.[9] The Gloss information of the words Hypernym, Hyponym, and Synonym and the sentence information of the learning data are represented. These two representations are used to identify semantic relationships and to solve problems of word sense disambiguation. It is very important to accurately represent sentences even if they complement each other's information. For accurate representation, a lot of learning data is needed to see various pattern information. In this paper, the graph embedding proposed by Vashishth et al. [10] is used to have a semantic relationship with the syntax information in the representation of words. The word representation method proposed by Vashishth et al. [10] displays syntax graphs and thesaurus graphs through Graph Convolutional Network. The model using this representation showed higher performance than the previous model.

2 Method

In this paper, we use GCN(Graph Convolutional Network) to reflect phrase information and semantic information in the word representation. The syntax information was taken from the Stanford CoreNLP parser. WordNet and ConceptNet information was used to reflect semantic information.

The word sense disambiguation is solved by Luo et al. [8] and consists of four modules: Context, Gloss, Memory, and Scoring. All word vectors used the word representation results of GCN.

The Context module represents input sentence by concatenate vector values from forward and backward through Bi-LSTM. The Gloss module represents Gloss information of words through Bi-LSTM in the same way. In this paper, Gloss Expansion method is used, and all Gloss information of Hypernym and Hyponym with verbs and noun parts is represented through Bi-LSTM. Hypernym and Hyponym information is extracted by depth k through BFS (Breadth First Search) to represent Gloss information as Context module. Gloss information represented in this way is concatenated by inputting Hypernym information into forward LSTM and Hyponym information into backward LSTM. In the memory module, the vector results in the Context module and the vector results in the Gloss Expansion module are updated through Attention. In the Scoring module, the meaning of the word is selected using the vector result in the Context module and the last attention result in the Memory module.

3 Experiments

The test datasets used in the experiments were Senseval-2, Senseval-3 task 1, SemEval-07 task 17, SemEval-13 task 12, and SemEval-15 task 13. The training data was SemCor3.0. This data is based on the semantic tags provided by WordNet 3.0, with a total of 225,036 semantic labels in 352 documents. The performance is shown in Table 1.

Table 1. F1-Score for test set of fine-grained English all-words WSD.

System	All
MFS	65.5
Bi-LSTM(att+LEX+POS) [7]	69.9
IMS(emb) [6]	70.1
GAS(concat) [8]	70.6
GCN(Dependency, WordNet)[10] + GAS(concat)	70.8
GCN(Dependency, WordNet + ConceptNet) + GAS(concat)	71.0

4 Conclusion

Analyzing the meaning of words is one of the most important parts of natural language processing. Humans consider contextual information to accurately understand the meaning of words in a sentence, and then use grammatical and dictionary information together.
In the future, we will consider how to represent the word embedding through various dictionary information to solve the limited number of words and to represent the contextual vector value.

Acknowledgments. 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. M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone," Proceedings of the 5th annual inter national conference on Systems documentation, pp.24–26, 1986.
2. E. Agirre, O. L ́opez de Lacalle, and A. Soroa, "Random walks for knowledge-based word sense disambiguation," Computational Linguistics, Vol. 40, No. 1, pp. 57–84, 2014.
3. A. Moro, A. Raganato, and R. Navigli, "Entity linking meets word sense disambiguation: a unified approach," Transactions of the Association for Computational Linguistics, Vol. 2, pp. 231–244, 2014.
4. O. Dongsuk, S. Kwon, K. Kim, and Y. Ko, "Word sense disambiguation based on word similarity calculation using word vector representation from a knowledge-based graph," Proceedings of the 27th International Conference on Computational Linguistics, pp. 2704–2714, 2018.

5. D. Ustalov, D. Teslenko, A. Panchenko, M. Chernosku-tov, C. Biemann, and S. P. Ponzetto, "An unsupervised word sense disambiguation system for under-resourced languages," arXiv preprint arXiv:1804.10686, 2018.
6. I. Iacobacci, M. T. Pilehvar, and R. Navigli, "Embeddings for word sense disambiguation: An evaluation study," Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1:Long Papers), pp. 897–907, 2016.
7. A. Raganato, C. D. Bovi, and R. Navigli, "Neural sequence learning models for word sense disambiguation," Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 1156–1167,2017.
8. F. Luo, T. Liu, Q. Xia, B. Chang, and Z. Sui, "Incorporating glosses into neural word sense disambiguation," arXiv preprint arXiv:1805.08028, 2018.
9. G. A. Miller, "Wordnet: a lexical database for english," Communications of the ACM, Vol. 38, No. 11, pp. 39–41, 1995.
10. S. Vashishth, M. Bhandari, P. Yadav, P. Rai, C. Bhattacharyya, and P. Talukdar, "Incorporating syntactic and semantic information in word embeddings using graph convolutional networks," Proceedings of the 57thAnnual Meeting of the Association for Computational Linguistics, pp. 3308–3318, 2019.