ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

A Survey on Word Sense Disambiguation Used In NLP

Swathy R

M. Tech Student, Dept. of I.T., Rajagiri School of Engineering, Ernakulam, India

ABSTRACT: The key idea of this survey paper is to show that Word Sense Disambiguation(WSD) when used in the Natural Language Techniques(NLP) techniques increases the efficiency in selecting the accurate keywords which are used as features in the classification techniques.The accuracy is increased when WSD is used along with the semantics of the word though WSD is considered as an open problem in NLP techniques

KEYWORDS: WSD,NLP

I. INTRODUCTION

Word Sense Disambiguation(WSD) is the ability to identify the correct sense of the word used in the sentence,when the word has multiple meaning. WSD is especially useful for machine translation. Examples include meaning of ``light'' in two different contexts as used below contexts as used below.

1. [I do like to eat something ``light (= about weight)''.]
2. [I find a switch for ``light (= about shine)''.]

These sense of ``light'' used here have the same spelling but are used in different context.

NLP involves the task of finding the semantic meaning of the text used in a content and thereby analysing the text information. To understand the opinion hidden inside the sentence NLP is mainly used.WSD when used along with NLP can be helpful in finding the correct sense of the word used in the context

WSD mainly depends on knowledge sources like corpora of texts which may be unlabeled, labeled or annotated with word senses, machine readable dictionaries, semantic networks etc because the framework of the procedure is that whenever a sentence is given, WSD makes use of more than one knowledge sources to attach the most exact senses with words in the context. Processing of WSD enables the selection of an optimal word sense. It is an important problem in NLP. Because, WSD is useful for machine translation and automated text summarization.

The objective of the system is to make the system understand the meaning of the words whose spelling is same but used in different context. The WSD task lends itself to different formalizations like the approach to the representation of a word sense . WSD heavily relies on knowledge. The procedure of any WSD system can be summarized as follows: given a set of words a technique is applied which makes use of one or more sources of knowledge to associate the most appropriate senses with words in context

II. LITERATURE SURVEY

The existing technique mainly uses Dictionary Based (DB) approach. The potential limitation with DB systems is that, the strength of classification depends only on the reference dictionary used. Also, most of these systems use Bag-of-Words concept which lacks domain/context based semantics . The main problem with bag-of-words approach is that, it considers only individual words and their frequencies during feature vector creation. When the feature vectors are selected based on keywords the classification will happen based on the keywords only. The relationship between the keywords will be missed out.These lexicon dictionaries either lack word senses or only contain words with their associated polarities rather than the words polarities with different senses.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 3, March 2017

Proposed System

Inorder to use the WSD system in NLP,the following system is built :

A. *Building Lexicon Dictionary*

WSD mainly relies on the lexicon dictionary for the different meaning of the word. If a word is not found in the dictionary, then this method will not be effective. A system is proposed where lexicon dictionary is built regarding all the word senses of the word

B. *Word Expansion*

In order to build lexicon dictionary, the word set is further expanded by obtaining the synonyms and antonyms from WordNet. Wordnet is a dictionary where the meaning of all the words are defined. So in order to build the lexicon dictionary,a connection with the wordnet is done to get all the word senses of the word.

Once the above steps are done.The next step is to get data from the user and processes data using WSD techniques toget feature vectors to be used in NLP.

III. CONCLUSION AND FUTURE WORK

The proposed systems shows that when WSD used in NLP increases the efficiency in selecting the keyword. Future work will be focused on the computational cost by optimization the feature vector size

REFERENCES

1. MonishaKanakaraj and Ram Mohana Reddy Guddeti ,``Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques",IEEE 9th International Conference on Semantic Computing
2. MonishaKanakaraj and Ram Mohana Reddy Guddeti ,``NLP Based Sentiment Analysis on Twitter Data Using Ensemble Classifiers", 3rd International Conference on Signal Processing, Communication and Networking
3. L. F. Coletta, N. F. F. d. Sommaggio Silva, E. R. Hruschka, and E.R.Hruschka,``Combining classification and clustering for tweetsentiment analysis,`` in intelligent Systems, 2014 Brazilian Conference on. IEEE,2014, pp. 210-215
4. A. Minanovic, H. Gabelica, and Z. Krstic, ``Big data and sentiment analysisusing knime: Online reviews vs. social media," in Information andCommunication Technology, Electronics and Microelectronics (MIPRO),2014 37th International Convention on. IEEE, 2014, pp. 1464–1468.
5. A. Porshnev, I. Redkin, and A. Shevchenko, ``Machine learning inprediction of stock market indicators based on historical data and datafrom twitter sentiment analysis," in Data Mining Workshops (ICDMW),2013 IEEE 13th International Conference on. IEEE, 2013, pp. 440–444.
6. P. T. Ngoc and M. Yoo, ``The lexicon-based sentiment analysis for fanpage ranking in facebook," in Information Networking (ICOIN), 2014International Conference on. IEEE, 2014, pp. 444–448
7. AgirrE, E., Ansa,O.,Martinez, D., Andhovy, E. 2001. Enriching WordNet concepts with topic signatures. InProceedings of the NAACL Workshop on WordNet and Other lexical Resources: Applications, Extensionsand Customizations (Pittsburg, PA). 23–28.
8. Alpaydin, E. 2004. Introduction to Machine Learning. MIT Press, Cambridge, MA.Statistical Post-Editing of a Rule-Based Machine Translation System, Proceedings of NAACL HLT 2009: Short Papers, pages 217–220, Boulder, Colorado, June 2009. c 2009 Association for Computational Linguistics, pp 217-224
9. Arabic/English Word Translation Disambiguation using Parallel Corpora and Matching Schemes, th EAMT conference, 22-23 September 2008, Hamburg, Germany, pp 6-11.
10. Gale, K. Church, and D. Yarowsky.: A Method for Disambiguating Word Senses in a Large Corpus. Computers and Humanities, vol. 26, pp. 415-439 .
11. Yarowsky.: Decision Lists for Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French. In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, pp. 88-95 .
12. T. Ng and H. B. Lee.: Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-based Approach. In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, pp. 40-47