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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

A Review on Word Sense Disambiguation

Sunita Rawat

Assistant Professor, Dept. of I.S.E., TOCE, Visvesvaraya Technological University, Bangalore, India

ABSTRACT: Word sense disambiguation (WSD) is described as the job of searching the sense of a word in a situation. WSD is a core problem in many tasks related to language processing. It is aggravated by make use of in several critical utilization like Part-of-Speech tagging, Machine Translation, Information retrieval, etc. Different topics such as ambiguity, evaluation, scalability and diversity cause challenges to results of WSD. In this paper we have discussed about some issues related to WSD and some WSD methods like knowledge-based, supervised, unsupervised and semi supervised.

KEYWORDS: WSD, Supervised, Unsupervised, Semi supervised, Knowledge-based

I. INTRODUCTION

Words may have numerous meanings and various senses. This is known as Polysemy. For example: one can consider bank as a river shore or a financial organization. Every so often two totally dissimilar words are spelled similar. This is known as Homonymy. Difference among homonymy and polysemy is unclear forever. WSD is the issue of resolving in what sense a word having a numerous senses is used in a specified statement. Take another example; consider the word "leaves", with 2 different senses:
1. plural of leaf (related to plant or tree)
2. to go away from

"The kids love to play in the leaves" and "kids do not like when their mom leaves for work". As a person it is understandable that the initial sentence is used in sense 1 mentioned above, along with the next sentence is used in sense 2. Even though this looks understandable to a person, making algorithms to imitate this person's skill is a tricky job. In terms such as the word "leaves" mentioned in given example, as a minimum some senses are noticeably dissimilar. In contrast, the various senses can be strongly associated and in this kind of cases separation of words into senses becomes very hard. Checking with various dictionaries will discover a lot of different separations of words into senses [1]. Some researchers have used one way is to select an exact dictionary, and utilize its all available senses. Shallow approaches and deep approaches are the two key approaches to WSD.

Deep approaches assume access to an inclusive body of world knowledge. Knowledge such as "you should not burn dry leaves instead bury it to create manure, except not for go away from somebody" and "kids use to get nervous if mom leaves them, except not for tree leaves" is then used to decide word sense. In practice all these approaches are not that much doing well, primarily there was no access to this kind of body of knowledge, excluding in incredibly restricted areas. However if this type of knowledge did subsist, it would be superior compare to the shallow approaches. It only takes the nearby words, by means of information such as "whether 'leaves' has words 'plant' or 'tree' close to, it almost certainly is in the first sense; but 'leaves' has the words 'going' close to, it is most likely in the leaving sense" [1]. This method, whereas theoretically is not that much strong as deep approaches, provides better outcome actually, because of partial world knowledge.

II. RELATED WORK

In WSD was initially invented as a different computational job for the duration (1940s) of machine translation. Most of the work in approximating the degree of ambiguity in bilingual dictionaries and texts is done during 1950s, moreover pertaining simple statistical methods.

Zipf released his "Law of Meaning" (1949) that subsequent for the skewed allocation of words by various senses, namely, in a power-law association more regular words have more senses compare to less frequent words [6]; the relationship has been confirmed for the British National Corpus (Edmonds 2005). Conclusion of Kaplan (1950) was

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

that two words of situation on both side of a vague word corresponding to an entire sentence of situation in determining power [7]. Master of a word, and then chose the heading whose contained words were most prominent in the context.

Madhu et. al, (1965) considered sense frequencies of words in various areas – examining first the domain limits sense – after that Bayes formula used to decide the most possible sense specified a situation [8]. WSD was revived in the 1970s in AI study on complete natural language sympathetic. By this inspiration, Wilks (1975) developed first structure to clearly consider for WSD i. e. "preference semantics" [9,10]. Selectional limits and a frame-based lexical semantics used by the system to discover a reliable group of word senses. The thought of entity "word experts" developed over this time (Rieger et. al, 1979).

In the AI paradigm, "Proper" knowledge representation was essential. Information basis had to be done by hand, as a result the consequent knowledge gaining bottleneck unavoidably led to inadequate lexical coverage of thin domains and would not extend.

WordNet (Miller 1990) plays very important role in research as it was both hierarchically ordered and computationally accessible into word senses known as synsets. Now a day, English WordNet is the popular sense record in WSD research [11]. Corpus-based WSD in statistical Machine Translation were first used by Brown et al. (1991) at the Senseval-3 workshop [12].

III. Issues in WSD

Here we have discussed commonly used terms related to word sense disambiguation.

A. *Dictionaries*

Dictionaries describe the speech of words, meanings and other attributes statically. Whereas, Corpus dynamically presents the use of polysemous words in real text situation. Rule base was planned as stated in the knowledge of linguistics by linguists.

B. *WordNet*

WordNet domain is used for identifying the accurate sense of the word. A domain may include synsets of different syntactic categories. It groups senses of the same word into homogeneous clusters, with the effect of reducing word polysemy in WordNet. WordNet domain provides semantic domain as a natural way to establish semantic relations among word senses.

C. *Information Retrieval(IR)*

As proposed by [2] in information retrieval WSD assists in enhancing term indexing. [2] has shown that if the senses are included as index terms, word senses give better retrieval performance. Therefore, ranking of documents should not based only on words but it also depends on word senses, or grouping of words and word senses. For example: Using different indexes for keyword "Java" as "programming language", as "type of coffee", and as "location" will give better accuracy to IR system.

D. *Machine Translation(MT)*

WSD is essential for MT. It facilitates in understanding of source language and translating it into target language. Depending upon the usage context it also affects the lexical choice [1].

E. *Speech Processing*

Specifically Speech processing is processing of homophone words which are pronounced the same way but spelled differently. For example: "base" and "bass" or "sealing" and "ceiling".

F. *Text Processing*

When words are pronounced in different ways based on their meaning is called Text to Speech translation. For example: "lead" can be "in front of" or "type of metal".

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

IV. APPROACHES AND METHODS

For Word Sense Disambiguation (WSD) here we have discussed four approaches:

A. *knowledge-based methods*

Knowledge-based methods rely on information that can be extracted or inferred from a knowledge source, such as a dictionary, thesaurus or lexical database. These methods learn based on information from curetted and structured data whereas supervised and clustering methods learn from example instances. The advantage of the knowledge-based methods over the supervised and the clustering methods is that training data is not required for each word that needs to be disambiguated. This allows the system to disambiguate words in running text, referred to as all-words disambiguation. All-words disambiguation methods have an advantage over what is termed lexical-sample disambiguation methods because lexical-sample methods can only disambiguate words in which there exists an ample set of training data.

All-word disambiguation methods are scalable and can be used in real-world practical applications in which ambiguous words may not be known ahead of time and training data is difficult to obtain. The disadvantage to this method is that it is language and domain dependent because a knowledge source is required in the appropriate language and domain. Historically, it has also not obtained as high of disambiguation accuracy as supervised methods.

Fig 1. shows a general model of knowledge-based WSD methods. In this method, the evaluation program takes the test data as input. The instances in the test data may be assigned their appropriate concept for evaluation purposes. These concepts are removed and the data is then sent to the vector creation module [5]. A test vector is created for each instance in the test data using information from the knowledge source. This information is obtained through the knowledge source interface module.

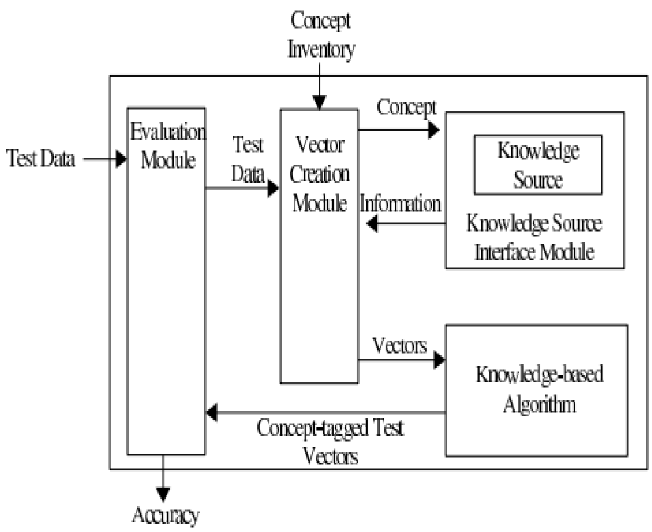

Fig 1. Knowledge-based WSD Method

The information obtained and knowledge source used varies, for example, [Mohammad and Hirst, 2006] use the category information from Macquarie's machine readable thesaurus, [Pedersen et al., 2005] use the semantic relatedness and similarity between the possible concepts and the words in the same situation as the target, and the test vectors are then sent to the knowledge-based algorithm, which uses the information in the vectors to determine the appropriate concept of the target word. There are several different types of knowledge-based methods, but they all rely on human curetted structured knowledge sources such as dictionaries, thesauri and/or lexical databases. The concept-tagged test data is then sent to the evaluation program and the accuracy of the method is returned. The remainder of this section describes two different knowledge-based algorithms that have been used in WSD: a similarity algorithm and a vector algorithm.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

B. *Supervised methods*

Supervised machine learning is the look for methods that cause from outwardly provided instances to create common hypotheses, using which then we can guess regarding future occurrences. However, the objective of supervised method is to make a brief model of the allocation of class labels. The resultant classifier is afterward used to allocate class labels to the testing examples where the values of the analyst features are recognized; excluding of the value of the class label is unidentified.

Collection of the dataset is the first step. If a necessary expert is accessible, after that s/he could suggest which fields (attributes, features) are the most informative [4]. If not, then the simplest method is that of “brute-force”, where it measures the whole thing available in the hope that the right features can be remote. However for induction, a dataset collected by the “brute-force” method is not directly suitable. It contains in most cases noise and missing feature values, thus needs important pre-processing (Zhang , 2002). The subsequent step is the data preparation and data preprocessing.

According to the situations, researchers have a plenty of approaches to decide from to tackle lost data. Hodge et al, have commenced a review of modern methods for outlier detection. And they have recognized the methods, pros and cons [12].

For sampling examples from a huge dataset lots of procedures are available. The process of recognizing and taking out immaterial and repeated features is known as feature subset selection. By using this dimensionality of the data get reduce and it also allows data mining algorithms to work quicker and more efficiently [4]. The reality that many features based on each other frequently excessively influences the correctness of supervised machine learning classification approaches. This difficulty can be overcome by making new features from the essential feature set. This method is known as feature construction transformation. Newly produced features like this might guide to the formation of more accurate and concise classifiers. Fig 2 shows the process of supervised machine learning.

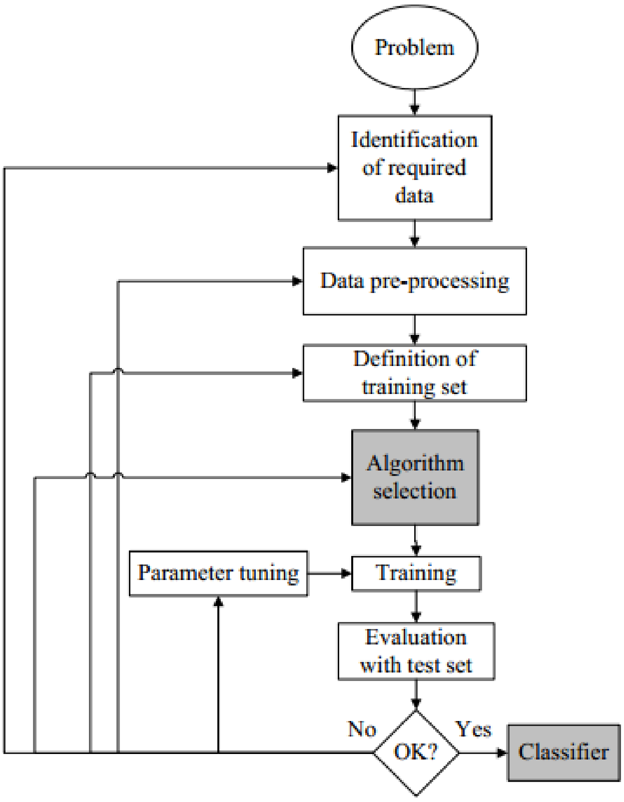

Fig 2. The process of supervised ML

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

C. *Semi-supervised methods*

Semi-supervised learning concerns with the learning of how human beings and computers obtain information in the presence of labeled as well as unlabeled data. As Semi–supervised approach shown improved performance in the presence of large volumes of data compared to the supervised as well as unsupervised learning, so preferred more [3]. Labels are difficult to achieve whereas unlabeled data are extra, hence semi-supervised learning is a decent sign to contract human effort and get better accurateness.

Semi-supervised learning is the method of discovering a improved classifier from labeled as well as unlabeled data. This method can give high concert of classification by using unlabeled data. It can give up an enhancement when unlabeled data can rebuild the best classification boundary. Some popular semi supervised learning models consist of mixture models, multiview learning, self-training, co-training and graph-based methods. The accomplishment of semi supervised learning based entirely on several basic hypotheses. Therefore the importance is on the hypotheses prepared by every one model.

D. *Unsupervised methods*

Unsupervised learning learns how machines can be trained to signify specific input patterns in a means that imitate the statistical structure of the inclusive collection of input examples. In dissimilarity with reinforcement learning or supervised learning, there are no overt final outputs or ecological assessments related with every one input [14]. Since unsupervised learning is probably to be greatly ordinary in the mind compare to supervised learning it is important [2].

The main thing is that unsupervised learning approaches have to work with the experiential patterns of input, which are frequently supposed to be self-determining samples from a basic unknown probability distribution, along with some unambiguous or implied a priori data as to what is significant. One main concept is that input, like the scene image, has distal self-determining reasons, for example items at specified positions light up by specific lighting. Because it is on those self-determining reasons that we usually should act, the most excellent depiction for an input is in their conditions. For unsupervised learning, two types of method have been recommended, one is density estimation techniques: clearly make statistical models of how fundamental sources could produce the input and other one is Feature extraction techniques: attempt to take out statistical regularities straightly from the inputs.

V. CONCLUSION

The Various approaches used for Word Sense Disambiguation are summarized in this paper. The hardness of WSD strictly depends on the granularity of the sense distinctions. We have surveyed various supervised, unsupervised, dictionary based and semi supervised approaches and came to know that depending on the problem or size of the dataset particular method is applicable.

REFERENCES

1. Esha Palta, 'Word Sense Disambiguation', First Stage Report Kanwal Rekhi School of Information Technology Indian Institute of Technology, Powai, Mumbai 2006-2007.
2. Peter Dayan, 'Unsupervised Learning', Appeared in Wilson, RA & Keil, F, editors. The MIT Encyclopedia of the Cognitive Sciences.
3. V. Jothi Prakash and L.M. Nithya, 'A Survey On Semi-Supervised Learning Techniques', International Journal of Computer Trends and Technology (IJCTT) – volume 8, number 1– Feb 2014.
4. S. B. Kotsiantis, 'Supervised Machine Learning: A Review of Classification Techniques', Informatica 31 (2007) 249-268.
5. Bridget Thomson McInnes, 'Supervised and Knowledge-based Methods for Disambiguating Terms in Biomedical Text using the UMLS and MetaMap', a thesis submitted to the faculty of the graduate school of the university of minnesota, September, 2009.
6. Zipf, George Kingsley, 'Human Behaviour and the Principle of Least Effort: An introduction to human ecology', Cambridge, MA: Addison-Wesley. Reprinted by New York: Hafner, 1972.
7. Edmonds, Philip 'Lexical disambiguation', The Elsevier Encyclopedia of Language and Linguistics, 2nd Ed., ed. by Keith Brown, 607–23, Oxford: Elsevier, 2005.
8. Madhu, Swaminathan & Dean W. Lytle, 'A figure of merit technique for the resolution of non-grammatical ambiguity' Mechanical translation, 8(2): 9–13, 1965.
9. Wilks, Yorick, 'Preference semantics: Formal Semantics of Natural Language', ed. by E. L. Keenan, III, 329–348. Cambridge, UK: Cambridge University Press, 1975.
10. Hirst, Graeme, 'Semantic Interpretation and the Resolution of Ambiguity', Cambridge, UK: Cambridge University Press, 1987.
11. Miller, George A., Special Issue, 'WordNet: An on-line lexical database', International Journal of Lexicography, 3(4), ed. 1990.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

12. Brown, Peter F., Stephen Della Pietra, Vincent J. Della Pietra & Robert L. Mercer, 'Word-sense disambiguation using statistical methods', Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL), Berkeley, California, 264–270, 1991.
13. Hetal Bhavsar and Amit Ganatra, 'A Comparative Study of Training Algorithms for Supervised Machine Learning', International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-2, Issue-4, September 2012.
14. Aditi Chawla, 'An Empirical Review on Unsupervised Clustering Algorithms In Multiple Domains', International Journal of Enterprise Computing and Business Systems, ISSN (Online) : 2230-8849, Volume 4, Issue 1, January - July 2014.