A STUDY ON WORD SENSE DISAMBIGUATION APPROACHES

[1]Lavanya Shree. E,[2]Dr.(Mrs)V.Vidyapriya
[1]M.Phil Scholar,[2]Associate Professor
[1]PG and Research Department of Computer Science,
[1]Quaid-E-Millath Govt. College for Women (Autonomous), Chennai,India

Abstract :Word sense Disambiguation (WSD) is the capability to discover the meaning of words in context in a computational way. WSD is an important but challenging technique within the region of natural language processing (NLP). Human language is ambiguous, so several words are understood in multiple ways depending on the context within which they occur. Hundreds of WSD algorithms are available, however much less work has been carried out in regard of choosing the optimal WSD algorithms. This paper focuses on different approaches of WSD such as supervised, unsupervised, semi supervised and knowledge based approaches. This paper will provide the users with widespread knowledge for selecting WSD algorithms for their specific applications or for further variation. This paper focuses on different approaches for solving the ambiguity of words.

***IndexTerms* - Word Sense Disambiguation (WSD), Natural Language Processing (NLP).**

I. INTRODUCTION

WSD is used for identifying which sense of a word should be used in a sentence, when that particular word has multiple meanings. Our human brain is sort of good at word-sense disambiguation. The natural language is created like that it requires so much of it is a mirror image of that neurologic reality. In computer science and the information technology, it has been a long-term challenge to develop the ability in computers to do the natural language processing and machine learning. Disambiguation requires two strict inputs, a dictionary and a corpus. A dictionary is to specify the senses which are to be disambiguated and a corpus of language data to be disambiguated.

For Example, the word bass in the statement “I went fishing for a few ocean bass” and “The bass line of the song is simply too weak”. It is obvious that the primary sentence is victimising the word "bass (fish)", and also the second sentence, the word "bass (instrument)" is getting used within the second sentence. As the word bass has multiple meanings then it will become ambiguous at the time of translation. Developing algorithms to copy this human ability will usually be a tough task. The solution for this problem has impacts on other computer related writing, such as discourse, improving the search engines, resolution of anaphora , coherence, inference. One can discover the utilization of WSD in machine translation, lexicography, knowledge mining, knowledge acquisition, semantic interpretation, semantic web and information extraction.

II. WSD APPROACHES

Some of the commonly used approaches for Word Sense Disambiguation (WSD) are Dictionary and Knowledge Based Approach and Machine Learning Based Approach. In Dictionary based approach, the systems are trained to perform the task of word sense disambiguation whereas in Knowledge based approach requires external lexical resources such as Word Net, Dictionary, Thesaurus etc.

III. DICTIONARY AND KNOWLEDGE BASED METHOD

This method requires Word Net or thesaurus as their Knowledge base. The main difference between thesaurus and Word Net is that, thesaurus is like dictionary which does not provide relationship which simply gives word and its category. Whereas Word Net provides six different relationships.

3.1 Lesk Algorithm

The Lesk algorithm is the seminal dictionary based method. Lesk algorithm is very simple and an old approach but it has less accuracy. Lesk algorithm is based on the hypothesis that the words used together in the text are related to each other. Two or more words may be disambiguated by finding their pair of dictionary senses with their greatest word overlap in their dictionary definitions. For example: the disambiguated words is "pine cone", the appropriate senses for both will include the words evergreen and tree.

PINE –

1. An evergreen coniferous tree.
2. The wood of *pine* trees.

3. Pine is a softwood.

CONE –

1. A solid or hollow object which is in a circular base to a point.
2. The dry fruit of a conifer.
3. A fruit of evergreen tree.

The relationship between the words can be observed in the definition of the words and their senses. This algorithm has two bags of words. Semantic bag and context bag. Semantic bag will have all the meaning of the ambiguous words and context bag will contain contextual words. Each Semantic word is attached with all the contextual words. After the probability of co-occurrence, it will decide which pair is more appropriate and accordingly context will be decided and which will give the appropriate meaning of ambiguous words. This algorithm has accuracy of 47% on SemCor subset. The table which is given below provides the accuracy comparison among lesk variants: [1]

METHOD	ACCURACY
SensevalFirst	40.2%
SensevalSecond	29.3%
SensevalThird	24.7%
Original Lesk	18.3%

Table 3.1: Comparison with SENSEVAL-2

3.2 Walker's Approach

Walker's approach is a thesaurus based algorithm. This algorithm can be expressed as each word is assigned out to at least one category of subjects in the thesaurus. Different subjects are assigned out to different senses of the word. In thesaurus based algorithm each sense will have a score. If it lies within the same then the score will be given as 1 or else the score will be given as 0. After which the sum will be calculated, which sense has the highest score that will be chosen as the target word. The flaw of this algorithm is that it does not contain any relationships. To overcome this flaw, ontological information is required which is expensive.

3.3WordNet

WordNet is a large lexical database which contains words with their relationships among them. WordNet is a combination of dictionary and thesaurus. WordNet not just interlinks word forms or strings of letters but it specifies the senses of words whereas thesaurus will not follow any pattern other than meaning similarity. As a result of WordNet, the words that are close to one another are semantically disambiguated. There are three main things that resides in WordNet: Definition, Gloss and Relationship. A definition is a statement that gives the meaning of a word using other words. Gloss is a list of words that are arranged alphabetically related to a specific subject with a brief explanation. There are six main relationships: Hypernymy, Hyponymy, Meronymy, Holonymy, Synonymy and Antonymy. Hypernymy and Hyponymy is the subset superset relationship. Meronymy and Holonymy is a Part – Whole relation holds between synsets. Synonymy is the meaning of the respective word. Antonymy is the opposite of the respective word. For Example:

- Fruit is hypernymy of water melon.
- Lip is meronymy of face.
- Simple is synonymy of easy.
- Good is antonymy of bad.

IV. MACHINE LEARNING BASEED APPROACH

Machine learning based approaches is used to learn features and will assign sense to unseen examples. There are three types of Methods in machine learning based approaches: Supervised Method, Unsupervised Method and Semi Supervised or Minimally Supervised Method.

4.1 Supervised Method

Supervised methods are costly but it has high efficiency. Supervised method will always have two data sets: Training data and Testing Data. This method requires tagged corpora as training set. Some of the methods for supervised word sense disambiguation are: Decision list, Decision tree, Naive bayes, Neural networks.

Decision List

Decision list is a set of ordered rules which is used for categorizing test instances. It is a logarithm of fraction of sense of word, generally used as one sense per word. If the denominator is larger, then the log answer will be in minus or else the answer is positive. The advantage of decision list algorithm is that it is easy to implement. The disadvantage is that it is totally word specific, each word needs to be trained.

Decision Tree

A decision tree is like a flow chart structure which divides the training data into a recursive manner. It represents the rules for classifying a data in a tree structure. Each internal node will represent the test on the attribute. Each branch is the outcome of how the decision is being made. Each leaf node will give the outcome or prediction.

Naive Bayes

A Naive Bayes classifer is based on the Bayes theorem. It is a simple probabilistic classifier. Naive Bayes can be computed by frequent occurrence of words. The expression given below can be used to calculate the joint probability:

$$p(F_1; F_2; : : : ; F_n; S) = p(S)\prod_{i=1}^{n} pr(F_j jS)$$

Here,
F1; F2:::Fn are features.
S is classification variable.
(S) is previous probability of classification variable.

All the zero values will be smoothen out because it indicates that the feature words never have co-occurrence. [1]

Neural Networks

The process of Neural Network is based on computational model of connectionist approach. The input will include the features of the input and the target output. The training dataset will be divided into sets. The sets are divided in such a way that there is no overlapping. When the new input pairs are encountered, the weights are adjusted. The weights are adjusted to get the target output.

4.2 Unsupervised Approach

Unsupervised approach is based on the fact that the words which are having similar senses will always have similar surrounding words. Word sense is derived by forming the clusters of occurrences of words. The task is to classify new occurrence of words to derived clusters. This method will detect the clusters instead of assigning the sense label. This method is cheaper than supervised method. On the other hand, unsupervised methods are less accurate when compared to supervised method. Some of the methods for unsupervised word sense disambiguation are: Word Clustering and Context Clustering.

Word Clustering

In word clustering, the words which are having similar meaning will be assigned to the same cluster. There are two approaches for word clustering: Syntactical dependency and Clustering by Committee Algorithm.

- The similarity between the words is given by Syntactical Dependency. If w consist of words that are similar to wn then a tree will be formed. The tree which is formed will have only one node wn. A node wi will have a child node wn, when wi is the word that has more similar meaning to wn.

- Clustering by Committee Algorithm will represent each word as a feature vector. A similarity matrix Smn is constructed when the target words are encountered. Smn matrix is a matrix that is similar to two words wm and wn. The next step is to form a set of words w in a recursive manner. The algorithm will then tries to find out the words which are not similar to the words of any committee. The words that are not part of any committee will be reused again to form more committees. Atlast, all the target word will belong to w which is a member of committee. The member of committee depends in its similarity to the centroid of the committee. [2]

Context Clustering

Content Clustering is based on clustering technique. In this method the context vectors are created first, then they are grouped into clusters to identify the meaning of the words. It uses vector space as word space. A word in a corpus is denoted as a vector and the number of its occurrences within its context will be counted. After this, the co-occurrence matrix will be created and the similarity measures are also applied. Discrimination will be performed using any of the clustering technique.

4.3Semi Supervised or Minimally Supervised Approach

Due to lack of training data, many word sense disambiguation algorithms uses semi-supervised method. This method will allow both labelled data and unlabeled data. In this method, the information is presented as similar as supervised method but the given information is less when compared to the supervised method. The semi-supervised method is gaining popularity because of its ability to get the critic information with small amount of reference data, while unsupervised method is outperforming on large data sets. The yarowsky algorithm is an example of this minimally supervised method.

V. CONCLUSION

This paper summarized the various approaches that are used for word sense disambiguation (WSD). This paper primarily focuses on the Machine learning approaches and Dictionary based approach. From the above approaches, it concludes that supervised method is performing better than other methods but supervised approach requires large corpora. Whereas unsupervised method does not rely on large scale resource for disambiguation. On the other hand, knowledge based approach uses knowledge sources to decide on the sense of words.

References

[1] Parth J. Vasoya, and Tarjni Vyas. A survey on word sense disambiguation approaches, International journal of Trend in Research and Development Volume-1(1).

[2] Vimal Dixit, Kamlesh Dutta, and Pardeep Singh. 2015. Word Sense Disambiguation and Its Approaches.CPUH-Research Journal.

[3] Eneko Agirre, and Philip Edmonds. 2007. ”What is word sense?” WSD algorithm and applications. vol. 33, pp. 8.

[4] Miguel ngel Ros Gaona, Alexander Gelbukh, and Sivaji Bandyopadhyay. 2009. Eighth Mexican International Con-ference on Artificial Intelligence.Web-based Variant of the Lesk Approach to Word Sense Disambiguation.

[5] Agirre, E. and Edmonds P. 2006. Word Sense Disambiguation: Algorithms and Applications (Text, Speech and Language Technology). Springer-Verlag New York, Inc. Secaucus, NJ, USA.

[6] Navigli, R. 2009. Word Sense Disambiguation: A Survey. Universita di Roma La Sapienza, ACM Computing Surveys.

[7] Sreedhar, J. Viswanadha, S. Raju, A. Babu, V. Shaik, A. and Kumar P. 2012. Word Sense Disambiguation: An Empirical Survey, International Journal of Soft Computing and Engineering (IJSCE).

[8] Xiaohua Zhou, and Hyoil Han. 2005. Survey of Word Sense Disambiguation Approaches. American Association for Artificial Intelligence.