*IOSR Journal of Computer Engineering (IOSR-JCE)*
*e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 14, Issue 6 (Sep. - Oct. 2013), PP 30-33*
*www.iosrjournals.org*

# A Survey on Word Sense Disambiguation

Rohit Giyanani
*Thadomal Shahani College of Engineering, Mumbai University, India.*

***Abstract:*** *Ambiguity has been always interwoven with human language and its evolution. Some argue that ambiguity of the human languages is a byproduct of its complexity, with words that are frequently used in language often being assigned to more than one reference in the real world, thus resulting in ambiguity. One of the biggest challenges in Natural Language Processing is for the computer to comprehend in what sense a specific word is used. The process of identifying the correct sense of the word in a particular sentence is called Word Sense Disambiguation. WSD is simply the lexical and semantic analysis in order to determine the contextual meaning of a word. It is a complex problem as it involves drawing knowledge from various sources. Significant amount of effort has been put into resolving this problem in machine learning since its inception but the toil is still ongoing.*
***Keywords:*** word *sense disambiguation, natural language processing, polysemy.*

## I. Introduction

Ambiguity is a major part of any human language. Almost every word in natural languages is polysemous, that is, they have numerous meanings or sentences. [15] Humans understand these various meanings of words with relative ease in most cases as they have a vast experience and a deductive mind which allows them to deduce sense of words in a sentence almost seamlessly. Computers on the other hand don't seem to have the human's vast experience of world and language. [1]Since WSD is based on knowledge, it is one of the primary obstacles for automated systems for sense disambiguation and thus it makes WSD a fundamentally difficult problem to crack. [9] Thatare why ambiguity of words has always been an Achilles heel of computers since the beginning. The lexical and semantic analysis of words is necessary for computers to make sense of the words. This is known as word sense disambiguation. Word sense disambiguationis the problem of picking a sense for a word from a set of predefined possibilities.

Sense Inventory usually comes from a dictionary or thesaurus.Knowledge intensive methods, supervised learning, and (sometimes) bootstrapping approaches. [4]. A classic example of sense disambiguation is the following sentence:

The word fair has at least three distinct senses, namely:

1. The price was fair.
2. The book fair had interesting titles.
3. The girl's face was fair.

Humans being naturally intelligent and experienced species understand the meaning of the word 'fair' in each of these sentences. But the computer cannot decipher the sense in which it is used. The computer can take the word fair in the following ways: {an exhibition, reasonable, or light skin tone}.Thus it can misinterpret the sentences.

WSD has been described as an AI-complete problem. [3] The basic procedure of any WSD system can be surmised as follows: from a given set of words, a method is applied which uses one or more knowledge sources to apply a sense to the words which seems most appropriate. These knowledge sources are be huge collections of texts, either in structured form in machine readable format or unstructured annotations, but creation of these knowledge bases is extremely costly and is a continuous process. This is one of the biggest issues of WSD, since there is no limit to the polysemous words. The potential of WSD is huge since most traditional NLP tools lack the semantic knowledge

Required to fully understand the meaning of words in sentences as they are mainly concerned with the lexicosyntactic analysis of data. With the ever-increasing data on the World Wide Web, an automatic tool which will be able to understand the data and sort it is very important and this is where WSD comes into picture.

## II. Related Work

Word Sense Disambiguation is a task which requires the use of artificial intelligence to alleviate the ambiguity of natural languages. The most important step is choosing a suitable classification method. The major classification methods are described below.

### 2.1 Knowledge Based Approach

Knowledge based approach for WSD involves the use of dictionaries, thesauri, ontologism, etc to understand the sense of words in context. Even though these methods have a comparatively lower performance than some other forms of approaches, but they do have large-scale knowledge resources. Knowledge based approach involves using an external dictionary source like WordNet or some other machine language dictionary. Initial knowledge based approaches to WSD were dated back to the 1980s when experiments were piloted on very small domains. But the lack of large scale computational assets did not permit a proper evaluation and comparison in end to end applications. Knowledge based approach either uses grammar rules for disambiguation or use hand coded rules for disambiguation.

A type of knowledge based approach involves using Selectional preferences. It requires an exhaustive enumeration of argument-structure of verbs. [5] It also requires selectional preferences of arguments along with the explanation of properties of words so as to meeting the selectional preferencecan be determined.
Another type is Overlap based Approach which requires a Machine Readable Dictionary. Its basic functioning involves finding an overlap amongst the features of different senses of an ambiguous word and the features in its content. These features could be assigned weights. The sense with the maximum overlap is nominated as the contextually appropriate word.

### 2.2 Supervised Learning Approach

Supervised learning approach for WSD uses Machine learning techniques to set a classifier from manually sense-annotated data sets. The classification task for assigning the correct sense to each instance of that word is done by a word expert known as the classifier. The training set usually contains some examples with the target word manually tagged with a sense from the sense inventory of a dictionary.
Naive Bayes classifier applies the Bayes' theorem to form a simple probabilistic classifier. Relying on the calculation of conditional probability of each sense $S_i$ of a word w given the features $f_j$ in the context.
An ordered set of rules for classifying trial instances is called a decision list. In the case of WSD this list contains a list for assigning an appropriate sense to a target word. It is basically a list of weighted condition rules. Training sets are used to induce a set of features. Rules of the kind {Feature value, sense, score} are created and the rules are ordered based on their decreasing score, constituting the decision list.

A decision tree is a predictive model which is used to signify classification rules with the help of a tree structure which repeatedly partitions the training data set. Every internal node of the tree signifies a test on a feature value while each branch represents an output of the test. A final prediction is made when it reaches a terminal node.

Neural Network is a unified group of artificial neurons that uses a computational model for data on the basis of a connectionist approach. The input to the learning program is a pair of input including (input feature, desired response). Goal is to use the input features to divide the training contexts into non overlapping sets corresponding to the desired responses. Link weights are increasingly adjusted so that the output unit representing the wanted response has a larger activation than any other output unit.

### 2.3 Unsupervised Learning Approach

Unsupervised Learning approach has the potential to overthrow the knowledge acquisition bottleneck which is the huge amount of resources manually marked with word senses.This approach is based on the thinking that same sense of a particular word will have alike neighbor words. Clustering word occurrences and classifying new occurrences into induced clusters is how they induce word sense, that is, they are independent of training sets and do not require machine readable resources like dictionaries, etc. Unsupervised learning for WSD performs word sense discrimination as it divides the occurrence of a particular word into various classes by determining by occurrences whether they have the same sense or not. Hence these methods may or may not discover cluster equivalent to dictionary sense inventory and hence evaluation is more difficult.
An important type of unsupervised learning for WSD is context clustering. Every occurrence of a targeted word in a corpora is depicted as context vector. The vectors are then clustered into various groups, all of which identify a sense of the target word. Historically this kind of approach is based on the idea of word space which is a vector space whose dimensions are words.

A different approach to the induction of word senses consists of word clustering techniques, that is, methods which are aiming at clustering words that are semantically similar and therefore the depict a similar meaning. Word clustering is identifying the word W= ($w_1$, …, wk) which is similar to target word w0. The similarity is determined on the information of their features and syntactical dependencies. To discriminate between the senses a clustering algorithm is applied.

## III. Applications

Word sense disambiguation hasn't demonstrated any clear cut benefits in human language technology applications till date but this failure is more due to the current deficiencies in its accuracy, and it will start becoming a very resourceful tool. Here are a few examples of real world examples of WSD.

1. Machine translation: It is an extremely difficult task to automatically identify the precise translation of a word in context. WSD has been considered as a major task which needs to be solved to enable an accurate machine translation, this is because it is widely known that disambiguation of words in a sentence can help choose better candidates as depending on the context words can have totally different translations. For example, the English word plant can be translated in Italian as *piante, impianto, stabilimento, cacciare* etc. Even though WSD disambiguation is very difficult to implement and some other methods have been proposed it still is the best option.
2. Information Retrieval: Explicit semantics are not used to narrow down documents which are not relevant to the user by even the most advanced search engines. The performance issues and the large overhead that might result due to the huge knowledge base traversed is the major reason that WSD has not contributed significantly to information retrieval historically. But with better methods to implement WSD it could be used to accurately offer what the user requested, an accurate disambiguation of the document database along with the disambiguation of the queried words will facilitate the selection of only those documents which are actually required.
3. Content analysis: Analysis of text with respect to ideas, themes, tones, etc can benefit from WSD used in content analysis domain. Consider the example of Blogger, it contains so many blogs, and their number is increasing rapidly. Content analysis using WSD can help in classification of data with as per user requirements.
4. Semantic Web: Semantic Web is nothing but a collaborative movement by World Wide Web Consortium to encourage webpages to include semantic content into their web pages to convert the currently existing unstructured or semi structured documents into a web of data. Almost all the above mentioned techniques can be used to achieve this vision and thus WSD plays an important role in achieving it.

Apart from the above mentioned applications, WSD can be implemented in applications such as Word Processing, Information Extraction and Ontology etc.

## IV. Conclusion

WSD is a difficult task as it involves with full complexities of language and aspires to identify a semantic structure from unstructured sources of text. In this paper, we discussed different methods by which WSD can be implemented. Out of these supervised learning seems to have the upper hand when it comes to accuracy and overall performance. [13] But it is not a realistic assumption to rely on the availability of a large training corpora for various languages and domains. On the other hand knowledge based learning methods show better potential at least in the short to medium term as there exists a structured knowledge exists (like WorldNet) and it can be utilized for WSD. [6] It has huge future potential in many fields both as an individual application and also as a homogenous part of another NLP application or as a plugin.

Word Sense disambiguation despite being an essential component of natural language processing remains amongst the most difficult to model adequately. It is true that the vagueness and the sheer volume of interpretations possible in natural languages puts a barrier on the accuracy of a polysemous word, but it is true that much more should be improved.

## References


[1] Raviv, Ariel, Shaul Markovitch, and Sotirios-Efstathios Maneas. "Concept-Based Approach to Word-Sense Disambiguation." *AAAI*. 2012.

[2] Fernandez-Ordonez, Erwin, Rada Mihalcea, and Samer Hassan. "Unsupervised Word Sense Disambiguation with Multilingual Representations." *LREC*. 2012.

[3] Navigli, Roberto. "Word sense disambiguation: A survey." *ACM Computing Surveys (CSUR)* 41.2(2009): 10.

[4] Mihalcea, Rada, and Dan I. Moldovan. "A method for word sense disambiguation of unrestricted text." *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*. Association for Computational Linguistics, 1999.

[5] Thottempudi, Sree Ganesh. "WORD SENSE DISAMBIGUATION."

[6] Ramakrishnan, Ganesh, et al. "Soft word sense disambiguation." *Proceedings of GWC*. Vol. 4. 2004.

[7] Dandala, Bharath, Rada Mihalcea, and Razvan Bunescu. "Word Sense Disambiguation Using Wikipedia." *The People's Web Meets NLP*. Springer Berlin Heidelberg, 2013. 241- 262.

[8] Mihalcea, Rada - "Word Sense Disambiguation". (2010): 1027-1030.

[9] Satapathy, Shruti Ranjan. *Word Sense Disambiguation*. Diss. Indian Institute of Technology, 2013.

[10] Resnik, Philip, and David Yarowsky. "A perspective on word sense disambiguation methods and their evaluation." *Proceedings of the ACL SIGLEX workshop on taggingtext with lexical semantics: Why, what, and how*. 1997.

[11] Lee, Wei Jan, and Edwin Mit. "Word Sense Disambiguation by using domain knowledge." *SemanticTechnology and Information Retrieval (STAIR), 2011 International Conference of* IEEE, 2011.

[12] Charhate, Sayali, et al. "Adding intelligence to non-corpus based word sense disambiguation." *Hybrid Intelligent Systems (HIS), 2012 12th International Conference on*. IEEE, 2012.
[13] Chatterjee, Niladri, and Rohit Misra. "Word-Sense Disambiguation using maximum entropy model." *Methods and Models in Computer Science, 2009. ICM2CS 2009. Proceeding of International Conference on*. IEEE, 2009.
[14] Bruce, Rebecca, and Janyce Wiebe. "Word-sense disambiguation using decomposable models." *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1994.
[15] Kulkarni, M., & Sane, S. (2011, April). An ontology clarification tool for word sense disambiguation. In *Electronics Computer Technology (ICECT), 2011 3rd International Conference on* (Vol. 1, pp. 292-296). IEEE.