

# Word Sense Disambiguation: A Survey

*Purabi Kalita, Anup Kumar Barman*

Department of Information Technology  
Gauhati University  
Guwahati, Assam, India  
*Purabikalita3@gmail.com*  
Department of Information Technology  
Gauhati University  
Guwahati, Assam, India  
*Anupbarman.gu@gmail.com*

**Abstract**— *Word sense disambiguation is the name of the process of assigning proper sense to a word having the potentiality of multiple senses in an automatic computational platform such as computational linguistics. It is one of the well discussed problems of natural language processing. The process involves a lot of difficulty due to lack of reasoning power and commonsense of computers. So human has to formalize some algorithms to computerize the process of sensing the appropriate sense of a word within a given context. This is the main aim of WSD approaches. During this survey, we have basically discussed about different approaches and also implementations of these approaches in different languages like Chinese, Japanese and also in Indian languages like Hindi, Nepali and Tamil.*

**Keywords:** WSD, Ambiguities, Knowledge based, Supervised, Unsupervised.

## 1. Introduction

Word Sense Disambiguation is an open problem in Natural Language Processing, which governs the process of identifying, which sense of a word is used in context when the word has multiple senses. Word sense ambiguity is prevalent in all natural languages. In context of WSD, ambiguity can be of two types, Homonymy and Polysemy. Homonymy is a group of words sharing same spellings and pronounces with different senses [7]. Polysemy is the special case of homonymy where multiple senses of the word are related to each other. For instance the Assamese word “কলা” means ‘arts’, ‘deaf’ and ‘black color’.

WSD task has two variants: "lexical sample" and "all words" task. The former comprises disambiguating the occurrences of a small sample of target words which were previously selected, while in the latter all the words in a piece of running text need to be disambiguated. The latter is more realistic form of evaluation, but the corpus is more expensive.

WSD is an AI complete problem as its solution seeks for the solution of general AI problems like common sense representation [6]. It involves much world knowledge.

A rich variety of techniques have been researched, from dictionary-based methods that use the knowledge from lexical resources, to supervised machine learning methods in which a classifier is trained for each word on a corpus of manually sense-annotated examples, to completely unsupervised methods that cluster occurrences of words, thereby inducing word senses. Among these, supervised learning approaches have been the most successful algorithms to date. [1, 2, 3, 4, 5]

## 2. Role of WSD

Main role of WSD is to disambiguate some text contents by assigning the appropriate sense to the target word of that context. It is important as most of the NLP applications require WSD. For example while doing some machine translation using computerized system; it is important to be sure that the computer takes the proper sense of the ambiguous word. কলা পোছাক পৰিহিত লৰাটোৱে কলা শাখাত পঢ়ে। It will be translated as either “the boy wearing black cloth studies in Arts section” or “the boy wearing arts studies in black section”. In information retrieval, to search the exact keyword one need to emit all the occurrences with.

## 3. Study of WSD

Starting of WSD approaches dates back to 1940s making it one of the oldest computational linguistic approach. 1980s became the turning point for WSD with the availability of large scale lexical resources and corpora. 1990 was the year of major development for WSD with wordnet available and sensval began. Wordnet makes it easy to access structured word senses as synsets and sensval helps in performing the comparison among different WSD approaches evaluated.

### 3.1 Study of WSD approaches

There are two main approaches in WSD: Deep and Shallow approaches. Deep approaches deal with world knowledge. Practically these are difficult to implement. Shallow approaches don't try to understand the text. They are dependent on some rules either given by human or automatically generated by the computer. Practically it gives superior result due to lack of resonance power of computer. So basically in computational linguistics, we deal with shallow approaches. There are different classifications of shallow approaches [8, 9, 13, 15].

### 3.1.1 Knowledge Based Approaches

Knowledge based approach works with the extensive use of knowledge sources like dictionaries, thesauri etc. It started in 1980s. These approaches use either grammar rules or hand coded rules for disambiguation. This method has comparatively lower performance than the supervised one [8, 9 and 17].

#### A) Overlapped Based Approach

- Contextual Overlapped Based Approach

Using knowledge based contextual overlap approach [10]; an ambiguous word is assigned with a more potential sense using three contexts:

1. Neighboring words of the target word
2. Wordnet and
3. The overlap between the above two pieces of information.

Here a bag of words called context bag C containing the context words from current as well as next and previous sentences is taken. Then another bag of words called sense bag S containing all different senses of the target word along with their related information like synonym, antonym, hyponymy, meronymy and their glosses is taken. After that, the overlap is taken doing intersection between the given two bags and find out the winner sense for the target word. The approach has been made efficient by using longer and multiple glosses by inducing proper nouns in wordnet, doing exhaustive preprocessing for glosses and preparing explanations for a sense using most probable senses of the word. Though implementing these ideas approach can be made enriched, the tasks become very tedious.

- LESK Approach

Traditional Lesk algorithm [19] compares the definitions of each sense  $s_i$  of a word W to the definition of every other word in the text context (words present in the phrase where W occurs). The sense whose definition or gloss shares the largest number of common words with the glosses of other words available in the context of the target word is assigned to the target word. Original Lesk works on short phrases and uses glosses found in traditional dictionaries. The disadvantage of traditional Lesk is that it is dependent on exact wordings in the target word context and hence provides poor performance. Banerjee and Pedersen [18] have proposed an adapted and modified version of original Lesk by using word net instead of using a standard lexical dictionary. This process extends the comparison to the definitions of words related to the words found in the context of the target word. The adapted algorithm compares glosses for synsets, hypernym, homonym, meronym between each pair of words in the considered window of context. For comparing two glosses, overlap between these two glosses is performed and then from each overlap a score is taken. A score is equal to the square of the words present in the overlap. After all the comparisons performed, all these scores are added to find the score for the candidate combination. The sense present in a candidate combination with the highest score is the winner sense assigned to the target word. For a window of context of N words has to do  $N*(N-1)/2$  comparisons. The process has attained an overall accuracy of 32% which is double of that of traditional one.

- Walker's Algorithm

It is a fully dictionary based algorithm. It works in steps as given below.

- a. To find out category given by the dictionary for each sense of the target word.
- b. To calculate score for each sense of the target word. One score is added while the category of the sense is similar to the category of the target word.
- c. The sense with the highest score gives the winning sense.

#### B) Selectional Restriction Based Approach

An approach to WSD using selectional restrictions restricts the occurrence of different senses of a word in a given context. Selectional restrictions are nothing but the concept of imposing constraints on the arguments by verb. It needs an exhaustive enumeration of argument structure of verbs. It also needs selectional preferences of arguments along with the explanation of properties of words. A measure of semantic relation is obtained for every pair of word  $w_1$  and  $w_2$  in a context  $y$  as  $(w_1, w_2, y)$ . The sense that violates this constraint is omitted. This approach has been used using Hindi wordnet [36]. Here using a small corpus of with co-occurrences and collocations, POS tagging has been done for that corpus to get a sense tagged corpus.

#### C) Conceptual Distance Based Approach

This approach provides a way to measure the distance among concepts by observing the relatedness of concepts. It is based on the idea of finding the relatedness of certain word sense with its context words to get the best sense out of others. It tries to give an efficient way of finding the closeness among words using a hierarchical net structure lexicon. Conceptual distance between two concepts is the length of the shortest path that connects the concepts in a hierarchical semantic net [38].

The conceptual distance measure is sensitive to following aspects.

- Length of the shortest path that connects senses.
- Depth in hierarchy
- Density of concepts in hierarchy
- Measure should be independent of number of concepts taken.

The result of the approach is promising considering the difficulty of the task and it does not need any discourse knowledge.

### 3.1.2 Supervised Approach

#### A) Naïve Bayes Approach

Naïve Bayes approach works by gathering information from the surrounding words of the target word. It is based on two assumptions: all considerable attributes (feature vectors) are independent of one another and words present in the phrase

do not maintain any order [28]. For a word  $W$ , with possible senses  $s_1, s_2, \dots, s_i$ , context of  $W$   $c_1, c_2, \dots$  and context features  $f_1, f_2, \dots$ , the process calculates  $\prod_{i=1}^n \prod_{k=1}^m p(s_k) p(c_i | s_k)$  for each sense of the target word and sense with the highest scores gives the result.

A naïve Bayes Classifier [11, 14] first takes the target word and then select the feature vectors such as neighboring words or POS etc. It works with the fact that probability of entire vector given the sense is the product of the probability of individual feature vectors given that sense. The approach naively presumes that feature vectors are independent of each other. The method then calculates the product of the prior probability of each sense and product of the probability of individual feature vectors given that sense to get a score. The sense with the highest score is the result [16].

## B) Association Rule Based Approach

An approach to word sense disambiguation using association rule (Kumar, Sharma and Niranjan [20]) is one kind of data mining association rule mining problem. Here, each context of the target word is considered as a transaction database as in data mining. Possible senses of the target word are taken as items. If frequency of occurrence of some items together is more then we can make an idea that these items must have some correlation. The process deals with the idea to find out the frequent item sets of the senses of the target word and its neighboring words by scanning the contexts of the target word considered as transactions and then the frequent items are used to generate association rules. Now, the sense deduced by most association rules is the winning sense assigned to the target word.

## C) Decision List Based Approach

A decision list can be best described as a classifier with if then else statement. This approach provides rules for assigning categories to the words in test data set. Feature vectors are selected from training data sets which are used to construct rules. In each obtained rule for each sense with a given sense a score is calculated as  $(F, S, \text{score})$  where  $F$ : feature vector;  $S$ : sense; and rules are then organized in decreasing order of the score value. Now for a given target word the winning sense is the one with the feature vector having highest score value [9].

### 3.1.3 Unsupervised Approach

Unsupervised approach removes the bottleneck problem of knowledge acquisition because it does not require any beforehand knowledge like lexical resources or training sets. Most Unsupervised methods are similarity and graph based. It deals with the fact that words having similar senses have similar neighboring words. This method processes by making clusters of word occurrences with target word as the vector of the cluster and assigning new occurrences to the available clusters according to their similarity in senses provided that a cluster contains words having similar senses [13, 17]. It is a highly robust and portable approach as it is independent of any resources.

## A) Graph Based

At first a graph is constructed with possible senses as nodes of the graph and edges as relations between senses. Then the graph is used for finding sense for the target word.

**HyperLex** is graph based WSD approach. This has been primarily made for information retrieval systems. Its aim was to identify the paragraph with the relevant sense of the target word. At first, all nouns and adjectives of the target word are considered to be taken as nodes. Verbs, determiners, any kind of menu, links etc. are removed. An edge is given between two nodes if the weight can be define as  $W_{a,b} = 1 - \max(p(a|b), p(b|a))$  where  $p(a|b) = \text{freq}(a,b) / \text{freq}(b)$  and  $p(b|a) = \text{freq}(a,b) / \text{freq}(a)$ . After creating the co-occurrence graph, strongly connected component and root hubs i.e. the most nodes connected to strongly connected components are obtained and then the target word is added to the graph such that each node has a connection to the target node. Each strongly connected component represents a distinct sense for the target word. A spanning tree out of that graph is found and every sub tree has been assigned a score. The component with highest score is the winner sense.

**PageRank** is an unsupervised knowledge based approach because it does not use any sense tagged corpora but uses wordnet to find out semantic similarity. The idea of this approach comes from the process of ranking webpages in search engines. Page rank is used to find the importance of each node in a graph. For each node the incoming votes (for connection from  $a$  to  $b$ ;  $b$  gets an incoming vote from  $a$ ) are scores for that node. All senses of the word are added to the graph with each node representing a sense. Edges are taken from semantic relations of words from wordnet. There is no edge between competing senses. After constructing the graph, pagerank is run and one rank for each node is obtained. The node with highest rank among its candidate synsets is assigned to the target word. This algorithm has given 45.11% accuracy while tested with SEMCOR [31].

## B) Semantic Analysis

This is an approach based on the clustering of semantically closed words using a statistical method called probabilistic latent semantic analysis and second order co-occurrence to generate efficient clusters (Gaurav S Tomar [12]). The correct sense for the target word can be obtained by comparing the similarity between the text contexts of target word and predefined clusters so that cluster with highest similarity gives appropriate sense. They have evaluated the method for English and Hindi language and obtained accuracy of 83% and 74% for English and Hindi respectively. The major advantage of this approach is that it is language independent due to use of statistical approach.

## C) Context Clustering

In this approach, different contexts of the target word are clustered using an agglomerative clustering method (McQuitty's Average link clustering). This is also called target word clustering approach. Each cluster starts with a single context and each context of the target word is represented as a feature vector which can be POS or any morphological characteristics. Similar clusters are merged together and thus merging clusters continues till specific number of clusters is found. The process does not require any kind of knowledge resources. Pedersen et al. [21, 22 and 23] has used this unsupervised WSD approach using context clustering.

## 3.2 Study of WSD for different languages

### 3.2.1 Indian Language

**Hindi Language:** Sinha et al. [24] were the first attempt to automatic WSD in Indian language. It has used the Hindi wordnet developed in IIT Bombay. Since then, there are a number of ongoing approaches for Hindi word sense disambiguation. The first attempt was to compare the context of the word with the contexts constructed from the wordnet. The output was a particular synset designating the word [7].

Another approach for Hindi language [25] has attempted to compare the different senses of the target word in the sentence with the word present in the wordnet synset and related parts of speech information of these words.

Sinha and Mihalea [34] proposed a n approach for Hindi word disambiguation using graph based approach. The method has provided a state of the art unsupervised WSD way with the right combination of similarity metrics and graph centrality algorithm.

Reddy et al. [35] proposed another n supervised approach for Hindi WSD based on ontological categories defined in Hindi wordnet. It provides the potentiality of applying in real time interactive system with the use of two unsupervised approaches.

**Nepali Language:** In an approach to Nepali word sense disambiguation [26] an attempt has been done to use knowledge sources using overlapped based and conceptual distance semantic graph based method. It has been found that the former one is less effective than the later one as the former one suffers from sparse overlap problem. In case of disambiguating nouns, overlapped based approach works well. But while in case of adjectives, semantic graph is more useful.

**Tamil Language:** In an approach to Tamil word sense disambiguation, [27] morphological inferences on morphologically rich Indian languages have been discussed. The basic idea deals with the fact that a morphological analyzer decreases different forms of an ambiguous word to their root. Here a supervised

### 3.2.2 Foreign Language

**Japanese Language:** In earlier days of WSD in Japanese, all sense tagged corpora were from newspaper articles. Construction of a balanced Japanese corpus is a national project being carried on. In Japanese WSD tasks, 3 types of information are considered: Sense inventory, Training data and test data [30].

In SemEval2 Japanese WSD tasks, 10 participating organizations worked on [30].

An approach to Japanese WSD is performed using clustering technique using 4 vectors: Collocation vector, Context vector, Association vector and Topic vector. Then clustering is done 4 times for 4 vectors using agglomerative algorithm. Among them the best one is selected for getting the appropriate sense [32].

Another approach to Japanese WSD is based on classification. A classifier is developed using Support Vector machine. The documents used here are from four different genres such as webpages, books, newspaper and whitepaper but training data set is made up from only the latter 3 genres. This method is a genre intrinsic feature based method with a view to excluding features from different genres [32].

**Chinese Language:** In a paper for Chinese WSD [29], has provided two supervised approaches based on context expansion considering knowledge from synonyms of context words. First method considers synonyms as feature vector to be used for Bayesian method assuming that representing context words by their synonyms will provide more information. In second method, a pseudo training data is made by making a new context by replacing actual context words by their synonyms. Here, both actual and pseudo training data are used for servicing the model giving more accuracy. It is found that the second method has increased the efficiency of existing method by 2.21%.

Zhou [33] has proposed a WSD approach based on Lexical Semantic Ontology (LSO) for ambiguous verb in Chinese language. LSO is a lexical resource for the purpose of semantic computation. It integrates information from several resources like lexicons and annotated corpora etc. LSO organizes verb into certain categories according to their meaning. A naïve Bayes classifier has been used for the task. The approach is totally based on the information obtained from LSO.

Another WSD system using unsupervised approach for Chinese language based on bilingual web search takes part in SensEval 2007. This is the first approach in Chinese language that uses bilingual search. They use google.co and baidu.com as search engines for the experiment. First of all, for a given Chinese ambiguous word within a Chinese sentence, the Chinese context is found and then using a Chinese English lexical dictionary, English contexts are also found and then the two contexts are given as query to the



search engines. After that by simply doing the association measurements, ambiguity is removed and translation between Chinese and English is operated [37].

#### 4. Conclusion

Dealing with WSD is a difficult task as it works with artificial intelligence problems like commonsense representation, knowledge acquisition of a computerized system. Though WSD is one of the most basic and essential component of NLP (without WSD all other applications of NLP remain incomplete), it is amongst the most difficult model to implement [8].

From the survey we have obtained that knowledge based approaches are mostly available but all of these suffer from the resource bottleneck problem. Also machine learning approaches are more capable for giving appropriate sense.

Also while doing survey for Indian languages, we have come across the idea that morphology affects sense disambiguation in Indian language [27].

Till date, no large scale, broad coverage accurate WSD has been built.

#### References

- [1] Roberto Navigli, *Word Sense Disambiguation*, ACM Computer surveys, vol. 41, no. 2, article 10, February 2009.
- [2] Rada Mihalcea, "Using Wikipedia for Automatic Word Sense Disambiguation", [online]. Available: [http://promethee.philo.ulg.ac.be/engdep1/download/bacIII/ACM\\_Survey\\_2009\\_Navigli.pdf](http://promethee.philo.ulg.ac.be/engdep1/download/bacIII/ACM_Survey_2009_Navigli.pdf).
- [3] <http://en.wikipedia.org/wiki/word-sense-disambiguation>.
- [4] Young Keok Lee and Hwe Tou Ng, *An Empirical Evaluation of Knowledge Sources and Learning Algorithms for Word Sense Disambiguation*, Proceedings of conference on Empirical Methods in Natural Language Processing, Philadelphia, pp. 41-48, July 2002.
- [5] David Yarosky, *Word Sense Disambiguation Using Statistical Models Roget's Categories Trained on Large Corpora*, Processing of COLING-92, Nantes, pp. 23-28, Aug. 1992.
- [6] Gerard Ascudero Bakx, "Machine Learning Techniques for Word Sense Disambiguation", Universitat Politècnica de Catalunya, Barcelona, May 22, 2006.
- [7] Preeti Yadav, Md. Shahid Husain, "Study of Hindi Word Sense Disambiguation Based on Hindi Wordnet", IJRASET ISeSN: 2321-9653 Vol. 2 Issue V, May 2014.
- [8] Rohit Giyanani, "A Survey on Word Sense Disambiguation", IOSR-JCE, e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 14, Issue 6 (Sep. - Oct. 2013), PP 30-33.
- [9] P. Borah, G talukdar and A Baruah, "Approaches for Word Sense Disambiguation-A Survey", IJRTE, ISSN: 2277-3878, Volume-3, Issue-1, March 2014.
- [10] A Chakrabarty, B S Purkayastha and Lavvyva Gavsinde, "Knowledge-Based Contextual Overlapped Keen Ideas for Word Sense disambiguation Using Wordnet".
- [11] Daniel Jurafsky, James H Martin, "Naïve Bayes Classifier Approach to Word sense Disambiguation", Seminar in Methodology and statistics, June 3, 2009.
- [12] Gaurav S Tomar, M Singh, S Rai, A kumar, R Sanyal and S Sanyal, "Probabilistic Latent Semantic Analysis for Unsupervised Word Sense Disambiguation", IJCSI, ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784, Vol. 10, Issue 5, No 2, September 2013.
- [13] Önder Eker, "DEVELOPING METHODS FOR WORD SENSE DISAMBIGUATION", Boğaziçi University, 2007.
- [14] T Pedersen, "A Simple Approach to Building Ensembles of Naïve Bayes classifiers for Word Sense Disambiguation", University of Minnesota Duluth.
- [15] R Mante, M Kshirsagar, Dr. Prasanta Chatur, "A Review of Literature on Word Sense Disambiguation", IJCSIT, Vol. 5, No.2, 2014.
- [16] P F Brown, S A Della Pietra, V J Della Pietra, R L Mercer, "Word Sense Disambiguation Using Statistical Method", IBM Thomas J Watson Research Centre.
- [17] X Zhou, H Han, "Survey of Word Sense Disambiguation Approaches", American Association for Artificial Intelligence, 2005.
- [18] S Banerjee, T Padersen, "An Adapted Lesk Algorithm for Word Sense Disambiguation Using Wordnet", University of Minnesota, Duluth.
- [19] M Lesk, "Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to tell a pine cone from an ice cream cone", Proceedings of SIGDOC, 1986.
- [20] S Kumar, N Sharma, Dr. S Niranjan, "Word Sense Disambiguation using Association Rule: A Survey", IJCTEE, ISSN 2249-6343, Vol.2 Issue 2, 27 March, 2012.
- [21] T Pedersen, R F Bruce, "Distinguishing word senses in untagged text", CoRR, cmp-lg/9706008.
- [22] D Singh Chaplot, Dr. Pushpak Bhattacharyya, "Literature Survey on Unsupervised Word Sense Disambiguation", IIT Bombay, May 7, 2014.
- [23] D Singh Chaplot, Dr. Pushpak Bhattacharyya, "Unsupervised Word Sense Disambiguation", IIT Bombay, Nov 25, 2013.

- [24] M Sinha, M Reddy, Dr. P Bhattacharyya, P Pandey, L Kashyap, "Hindi Word Sense Disambiguation", IIT Bombay, Department of Computer of Computer science and Engineering, Mumbai, 2008.
- [25] Rohan, "Word Sense Disambiguation for Hindi Language", Thapar University Patiyala, CSE Department, India, 2007.
- [26] A Roy, S Sarkar, B S Purkayastha, "Knowledge Based Approach to Nepali Word Sense Disambiguation", IJNLIC, vol.3, no.3, 2014.
- [27] B Sankaran, K V Shanker, "Influence of Morphology in Word Sense Disambiguation for Tamil", AU-KBC Research Centre.
- [28] Dempster, A., Laird, N., Rubin, D., "Maximum Likelihood from incomplete data via EM algorithm", J.Roy.Stac.Soc, B 39 (1), 1-38, 1977.
- [29] Z Yang, H Huang, "Chinese Word Sense Disambiguation based on Context Expansion", Proceedings of COLING 2012, Mumbai, pp. 1401-1408, December, 2012.
- [30] M Okumora, K Shirai, K Komiya, H Yokono, "SemEval 2010 Task: Japanese WSD", Proceedings of 5<sup>th</sup> International Workshop on Semantic Evaluation, ACL 2010, Sweden, pp. 69-74, July 2010.
- [31] R Mihalcea, P Trarao, E Figa, "PageRank on Semantic Network, with application to word sense disambiguation", Proceedings of COLIN, Geneva, 2004.
- [32] K Shirai, M Nakamura, "JAIST: Clustering and Classification based approaches for Japanese WSD", Proceedings of fifth International Workshop on Semantic Evaluation, Sweden, 2010.
- [33] K Zhou, "Chinese Word Sense Disambiguation based on Lexical Semantic Ontology", Journal of Chinese Language and Computing 18(1) 13-23.
- [34] Ravi Sinha and Rada Mihalcea, "Unsupervised Graph based Word Sense disambiguation Using Measures of Word Semantic Similarity," IEEE International Conference on Semantic Computing, pp. 363 – 369, Sept. 2007.
- [35] Siva Reddy, Abhilash Inumella, Rajeev Sangal, Soma Paul, "All Words Unsupervised Semantic Category Labeling for Hindi", Proceedings of the International Conference RANLP, Borovets, Bulgaria, pages 365–369, September 2009.
- [36] P Bala, "Word Sense Disambiguation using Selectional Restrictions", International Journal of Scientific and Research Publications, ISSN 2250-3153 Vol.3, Issue 4, April 2013.
- [37] P Liu, T Zhau, M Yang, "HIT-WSD: Using Search Engine for Chinese-English Lexical Task", Proceedings of 4<sup>th</sup> International Workshop on Semantic Evaluation, Prauge, June, 2007.
- [38] E Agiree, G Rigau, "A roposal for Word Sense Disambiguation using Conceptual Distance", cmp -lg/ 9510003, October, 1995.