# A REVIEW ON WORD SENSE DISAMBIGUATION

Priyanka[1], Amita Arora[2]
[1]Student, [2]Assistant Professor

**Abstract**
**Word sense disambiguation is a challenging task in natural language processing. Word sense processing can be defined as the process of finding the meaning of a word in the given context when it can have multiple meanings. It can be used in various applications like machine learning, text summarization, machine translations, information retrieval etc. In this paper we did survey of knowledge based, machine learning based and hybrid approaches of word sense disambiguation.**
**Keywords: Knowledge-based, Supervised, Unsupervised, Word sense disambiguation (WSD).**

## I. INTRODUCTION

Human languages are fairly ambiguous. There are various words, which have different meanings, depending upon the context in which those words appear. Most of the words in natural languages are polysemous, having multiple possible meanings or senses. Humans can understand the meaning of a word by simply looking at surrounding. Same idea is used by machines to identify correct sense of a word in a given context. WSD is the process of determining the sense of an ambiguous word in a particular context. Generally neighboring words are used as the context of a target word.

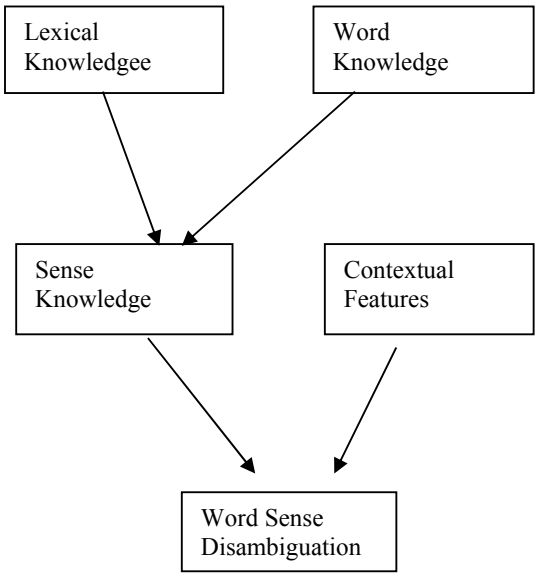


Fig 1 Conceptual Model of WSD

Example:

1. Oliver took felicity to a cafe on a date.
2. Laurel's favourite fruit to eat is a date.
3. William's date of birth is June 13, 1994.

In first sentence, word date stands for 'a romantic meeting'. While in second and third sentences, meaning of date is 'a fruit 'and 'a day of month 'respectively.

## II. LITERATURE SURVEY

In 1940s WSD was first formulated as separate computational task during the early days of machine translation. This makes it one of the oldest problems in computational linguistics. In 1949, Warren Weaver first introduced the concept in computational context.

Till 1970s WSD was a subtask of semantic interpretation systems which were developed within the field of artificial intelligence.

However, since WSD systems were at the time largely rule-based and hand-coded they were prone to a knowledge acquisition bottleneck. In the 1990s, the statistical revolution swept through computational linguistics, and WSD became a paradigm problem on which to apply supervised machine learning techniques. Since then, supervised techniques have reached a plateau in accuracy, and so attention has shifted to coarser-grained senses, domain adaptation, semi-supervised and unsupervised corpus-based systems, combinations of different methods, and the return of knowledge-based systems via graph-based methods. Still, supervised systems continue to perform best.

## III. Approaches and Methods:

**Knowledge based approach:** This approach use lexical knowledge bases such as dictionary, thesauri, wordnet and extract knowledge from word definitions and relation among words and senses. They may use grammar rules and/or hand coded rules for disambiguation. Selectional preferences [4] capture information about the possible relations between word categories, and represent commonsense knowledge about classes of concepts. DRIVE-VEHICLE, PLANTTREES, are examples of such semantic constraints, which can be used to rule out in correct word meanings and select only those senses that are in consistency with Common sense rules **.** Predicates often have a preference for a particular argument. This inclination of predicates to select for a particular argument is known as selectional preference.

Overlap based approaches depend on the finding the overlap between features of the senses definitions of two or more target words (Lesk algorithm) [5]. The sense which has maximum overlap is selected as contextually appropriate sense. In lesk Algorithm**,** there are two data bags : - a) Sense Bag: It contains the words in the definition sense of context word. b) Context Bag: It contains the definition of each sense of each context word. Maximum overlap between senses of words determines the correct sense of word in that context. Walker's Algorithm is a Thesaurus based approach. For each sense of the target word, the thesaurus category to which that sense belongs is found. The score for each sense is calculated by using the context words. Add 1 to the score of the sense if thesaurus category of the word matches that of sense.Word Sense Disambiguation with Conceptual Density method [3] selects a sense based on the relatedness of that word-sense to the context. Relatedness is measured in terms of conceptual distance i.e. how close the concept represented by the target word and the concept represented by its context words is Conceptual Distance. A structural hierarchical semantic net is used for finding the conceptual distance. Smaller the conceptual distance, higher will be the conceptual density.

**Supervised methods:** Labelled corpus data is used. A corpus provides a set of samples that enables the system to develop some numerical models. Main idea of this method is that the context can provide enough evidence on its own to disambiguate words [10].

Naïve Bayesian Classifiers[7] is based on the application of Bayle's theorem in which joint probability of each sense of a word over the features defined in the given context is determined and the maximum value of joint probability is chosen for the correct sense of word the using the trained annotated corpora [2].

Decision Tree and Decision List Method [6] is a word specific classifier and a separate classifier needs to be trained for each word. This approach can be considered as weighted „yes" or „no" rules where the exceptional conditions appear at the root node of the list with high weight and the general condition of the list appear at bottom with low weights. A scoring function calculates the weight which describes the association between the condition and the particular class.

Support Vector Machine (SVM) Method [2] is based on the idea of learning linear hyper plane from the training set that separates positive samples from the negative samples

**Unsupervised Methods:** This is based on unlabeled corpus. The occurrence of specific word is divided into number of classes in order to decide whether the occurrence of word have same sense or not. Various methods used in unsupervised approach are: context clustering, word clustering and co-occurrence graph. In context clustering method, every occurrence of target word is represented as context vector in the corpus. These vectors are grouped into clusters for the identification of sense of the target word. In word clustering, words that are

semantically similar are clustered to form a specific meaning. Graph is created on the basis of grammatical relationship between words. In co-occurrence graph method, graph is created on the basis of grammatical relationship between words. Every word in the text is called a vertex and syntactic relationship is called an edge. Weights are assigned to the edge on the basis of relationships. An iterative algorithm is applied on the graph to find the word that have the highest degree node and at last minimum spanning tree is used to disambiguate the target word. Unsupervised methods have the potential to overcome the knowledge acquisition bottleneck. Based on the idea that the same sense of a word will have similar neighbouring words, they are able to induce word senses from input text by clustering word occurrences, and then classifying new occurrences into the induced clusters Recently, researchers have proposed several Graph based methods [9] in which a researcher builds a graph with senses as nodes, and relations among words and senses as edges, with the relations usually acquired from an LKB such as WordNet. Then, the researcher conducts a ranking algorithm over the graph, and assigns senses that are ranked the highest to the corresponding words. Researchers using these methods have experimented with different relations and ranking algorithms, such as the Personalized PageRank algorithm [8]. These approaches are based on the notion of a cooccurrence graph, that is, a graph G = (V, E) whose vertices V correspond to words in a text and edges E connect pairs of words which co occur in a syntactic relation, in the same paragraph, or in a larger context. Hyperlex, a cooccurrence graph is built such that nodes are words occurring in the paragraphs of a text corpus in which a target word occurs, and an edge between a pair of words is added to the graph if they cooccur in the same paragraph. Each edge is assigned a weight according to the relative cooccurrence frequency of the two words connected by the edge. Similarity-based algorithms assign a sense to an ambiguous word by comparing each of its senses with those of the words in the surrounding context. The sense whose definition has the highest similarity is assumed to be the correct one [9]. Problems with unsupervised approach are that the instances in training data may not be assigned the correct sense, clusters are heterogeneous. Number of cluster may differ from the number of senses of target word to be disambiguated.

**Hybrid Approaches** : They make use of corpus evidence as well as the semantic rules. Few examples of hybrid approaches are bootstrapping approach, Yarowsky algorithm (1995).

## IV. Conclusion

Knowledge Based approach requires exhaustive knowledge resources and suffer from sparsity .Its accuracy is approximately 55-60%. .Supervised learning method requires large annotated data due to which Knowledge acquisition problem is there. Supervised approach cannot be used for resource scarce languages. Accuracy of supervised approaches is approximately 60-65%. In unsupervised approach, the training data is totally unlabelled and to form cluster from the corpus is quite difficult. Performance of unsupervised approach is lower as compared with other approaches. So effort should be made to use classifier such as Hybrid approach with annotation up to certain level.

## V. REFERENCES

[1] Rohit Giyanan "A survey on word sense disambiguation"https://www.researchgate.net/.../272719121_A_Survey_on_Word_Sense_Disambiguatio

[2] David Yarowsky, "Word Sense Disambiguation using Statistical Models of Rogets categories trained on large corpora." dx.doi.org/10.3115/992133.992140

[3] Agirre, Eneko & German Rigau. 1996. "Word sense disambiguation using conceptual density", in Proceedings of the 16th International Conference on Computational Linguistics (COLING), Copenhagen, Denmark, 1996

[4] Agirre, Eneko & David Martínez. 2001. Learning class-to-class selectional preferences. Proceedings of the Conference on Natural Language Learning, Toulouse, France, 15–22.

[5] Satanjeev Banerjee, Ted Pedersen, "An adaptive Lesk Algorithm for Word Sense Disambiguation Using WordNet", Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, page no: 136-145, 2002

[6] Agirre, E. and Martinez, d. 2000. Exploring automatic word sense disambiguation with decision lists and the web. In Proceedings of the

18th International Conference on Computational Linguistics (COLING, Saarbr ucken, Germany). 11–19.

[7] Gerard Escudero, Llu´ıs M`arquez and German Rigau, "Naïve Bayes and Exemplar-based approaches to Word Sense Disambiguation Revisited", arXiv:CS/0007011v1, 2000.

[8] E. Agirre and A. Soroa, "Personalizing PageRank for Word Sense Disambiguation," Proc. 12th Conf. European Chapter of the Assoc. for Computational Linguistics (EACL 09), Assoc. for Computational Linguistics, 2009, pp. 33–41

[9] R. Navigli and M. Lapata, "An Experimental Study of Graph Connectivity for Unsupervised Word Sense Disambiguation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 32,no. 4, 2010, pp. 678–692.

[10] Mukti Desai, Mrs. Kiran Bhowmick, "Word Sense Disambiguation" www.ijesi.org/papers/Vol%202(10)/Version-2/A021020104.pdf

[11] Samhith, Arun tilak, Prof. G. Panda, "Sense Disambiguation using WordNet Lexical Categories"ieeexplore.ieee.org/document/7955725