The Various Approaches for Word Sense Disambiguation: A Survey

K. Gunavathi[1], S. Rajini[2]
[1]*PG Student, Dept. of CSE., Kumaraguru College of Technology, Coimbatore, India.*
[2]*Associate Professor, Dept. of CSE., Kumaraguru College of Technology, Coimbatore, India.*

Abstract- **Word sense disambiguation is an important problem of natural language processing and ontology in the field of computational linguistics. It is defined as identifying the exact sense of a word which is used in the sentence. It is used in various applications of natural language processing like machine translation, information extraction and text mining, information retrieval etc. This survey paper describes the various approaches adopted for word sense disambiguation such as dictionary and knowledge based approaches, supervised approaches, semi-supervised approaches and unsupervised approaches.**

Index Terms- **Word sense disambiguation, dictionary and knowledge based approaches, supervised approaches, semi-supervised approaches, unsupervised approaches**

I. INTRODUCTION

Word sense disambiguation is the process of identifying the exact sense of an ambiguous word which has multiple senses in a given sentence. For example, the word mouse represents a rodent and a computer device. Let's consider the two sentences 'the mouse eats cheese.' and 'I want to buy a new mouse for my computer.' In the first sentence, mouse represents a rodent but in the second sentence it refers to a computer device. Thus the same word has different senses depending on the context.

Word sense disambiguation has different approaches such as dictionary and knowledge based methods, supervised methods, semi-supervised methods and unsupervised methods.

II. LITERATURE SURVEY

In 1940s, WSD was first formulated as a distinct computational task during the early days of machine translation. In 1949, Warren Weaver first introduced the problem in a computational context.

In 1950s, Kaplan determined that two words of context of an ambiguous word was equivalent to a whole sentence of context in resolving power.

In 1960s, Bar-Hillel argued that WSD could not be solved by "electronic computer" because of the need in general to model all world knowledge. In 1965, Madhu and Lytle calculated sense frequencies of words in different domains and then applied Bayes formula to choose the most probable sense given a context.

In 1970s, WSD was a subtask of semantic interpretation systems developed within Artificial Intelligence (AI) research. In 1975, Wilks developed "preference semantics", to find a consistent set of word senses for the words in a given context using selectional restrictions and a frame-based lexical semantics.

In 1980s, large-scale lexical resources, such as the Oxford Advanced Learner's Dictionary of Current English (OALD), and corpora became available so hand-coding was replaced with knowledge automatically extracted from these resources.

In 1990s, the statistical revolution in natural language processing swept through computational linguistics, and WSD became a paradigm problem on which to apply machine learning techniques.

In 2000s, supervised techniques reach a mature in accuracy, and so the researchers has shifted to other domains such as coarse-grained senses, domain adaptation, semi-supervised based systems, unsupervised corpus-based systems, combinations of different methods, and dictionary and knowledge-based systems via graph-based methods.

III. WSD APPROACHES AND METHODS

A. DICTIONARY AND KNOWLEDGE BASED APPROACHES

Knowledge based methods use knowledge sources of Machine Readable Dictionaries (MRD) such as

WordNet, dictionary and thesaurus etc. to decide the senses of words in a given sentence. There are four knowledge based approaches as follows:

1) *The Lesk Algorithm*

It computes the overlaps between words that are the number of words in common between the definitions of senses. For example, consider the words pine and cone. The senses of the two words are as follows:

Four senses of pine:-

- An evergreen tree with long needle shaped leaves
- Pineapple
- Waste away through sorrow or illness
- Desire for something

Three senses of cone:-

- Circle shape which narrows to a point
- Solid or hollow shape
- Dry fruit of a pine or fir tree

In this example, the first definition of pine and third definition of cone have the largest overlap among all sense combinations with three words evergreen, tree and pine in common.

2) *Measure of semantic similarity computed over semantic network*

This includes the methods for finding the semantic distance between concepts. They are divided into two categories: local context and global context. It depends on the size of the context. Semantic similarity is the most powerful constraint used in automatic word sense disambiguation.

3) *Heuristic method*

Heuristic method consists of simple rules that can reliably assign sense to certain word categories that includes Most frequent sense (i.e., one meaning of word occurs more often than other meaning), One sense per collocation, states that a word tends to preserve its meaning across all its occurrences in a given discourse, and One sense per discourse, states that a word tends to preserve its meaning when used in same collocation.

4) *Selectional preferences*

Selectional preferences identify the information of the likely relations of word types, and denote common sense using the knowledge source. For example, modeling-dress and walk-shoes are the words with semantic relationship. In this approach improper word senses are omitted and only those senses of the word are selected which have harmony with common sense rules.

B. SUPERVISED APPROACHES

Supervised approaches are approaches that rely on sense tagged corpora for disambiguation. They consist of training phase and testing phase. In training phase, classifier builds using sense annotated training corpus from which syntactic and semantic features are extracted, and in the testing phase, classifier picks the best sense of a word on the basis of surrounding words. Supervised methods yield very high accuracy in the domain of the training corpus. But this accuracy comes at the cost of sense tagged corpora which is a costly resource in terms of the time and the manual efforts involved. Regarding automatic word sense disambiguation, supervised learning is one of the most successful approaches in recent years. The supervised approach to WSD uses semantically annotated corpora to train machine learning algorithm to decide which sense to choose in which context. The supervised approaches are as follows:

1) *Decision Tree*

Decision tree is one of the prominent methods for word sense disambiguation. It uses selective rules associated with each word sense in a tree structure. In this approach the system selects one or more rules which satisfies features and assign sense to the ambiguous word based on their prediction. The sense of a word is represented at a leaf node of the tree.

2) *Neural Networks*

Neural networks can be used to represent words as nodes and these words will activate the ideas to which they are semantically related. The goal of neural network is to partition the training context into non-overlapping sets. The input includes the input features and the target output. When the network encounters new input pairs, the weights are adjusted so that the output unit has the larger activation. The network can have weights both positive and negative corresponding to correct or wrong sense of the word.

3) *Naive Bayes*

Naive Bayes classifier is the classifier based on Bayes theorem and assumes that every feature is class

conditionally independent of every other feature. This approach classifies text documents using two parameters: the conditional probability of each sense of a word and the features in the context.

4) *Support Vector Machine*

Support Vector Machine uses the theory of structural risk minimization. The goal of this approach is to separate positive examples from negative examples with maximum margin. Here, margin is the distance of hyperplane to the nearest of the positive and negative examples and support vectors are the positive and negative examples which are closest to the hyperplane.

C. SEMI-SUPERVISED APPROACHES

Semi-supervised algorithms (or minimally supervised algorithms) make some assumptions about the language and discourse in order to minimize the restrictions to few languages because of their dependence on sense tagged corpora. The semi-supervised approaches are as follows:

1) *Bootstrapping*

This algorithm is based on Yarowsky's supervised algorithm that uses Decision Lists. It makes a couple of assumptions regarding the language. The assumptions are One sense per Collocation, states that the sense of a word is strongly dependent on the neighboring words, and One sense per Discourse, states that every document contains a single sense of a word with high probability.

2) *Monosemous Relatives*

Monosemous relatives approach is developed as a bootstrapping algorithm to use words with single sense as possible synonyms. In this approach, all words having single sense (the sense of word itself) are found through the synset of a word.

D. UNSUPERVISED APPROACHES

Unsupervised methods are based on unlabeled corpora which are required to be trained before using them on ambiguous words. They introduces the concept that the sense of a particular word depends on its neighboring words to show the problem of knowledge acquisition bottleneck but the performance of this method is lower than the other methods. It divides the occurrence of specific word into number of classes in order to decide whether the occurrence of word has same sense or not to identify the sense clusters. The unsupervised approaches are as follows:

1) *Context Clustering*

Context clustering is a clustering technique in which first context vector is created. In this method, every occurrence of target word is represented as context vector in the corpus. These vectors are grouped into clusters to identify the sense of the target word. In this approach, large amount of manually annotated training data is not required.

2) *Word Clustering*

In word clustering method, words that are semantically similar are clustered to form a specific meaning. This technique is similar to context clustering in terms of finding sense but it clusters those words which are semantically identical. This approach uses Lin's method for clustering. It checks identical words which are similar to target word and similarity among those words are calculated from the features they are sharing. This can be obtained from the corpus.

3) *Co-occurrence Graphs*

A graph is created with vertex, which represents every word in the text, and edge, which represents syntactic relationship. Each edge of graph is assigned a weight on the basis of relationships. Word with high frequency is assigned the weight 0, and the words which are rarely co-occurring, assigned the weight 1. An iterative algorithm is applied on the graph to find the word that has the highest relative degree and at last minimum spanning tree is used to disambiguate the actual sense of the target word.

IV. COMPARISON OF DIFFERENT APPROACHES OF WSD

Table 1: comparison of different approaches for WSD

Knowledge based approaches	**Supervised approaches**	**Semi-supervised approaches**	**Unsupervised approaches**
They use knowledge sources of MRD such as WordNet, dictionary and thesaurus etc., grammar and hand coded rules.	They rely on sense tagged corpora for disambiguation and use corpus evidence.	They make some assumptions about the language and discourse.	They are based on unlabeled corpora and combines knowledge based and supervised methods.
They give higher precision.	They found to be better than other two approaches that is knowledge based and unsupervised approach.	They require small amount of labeled training data and sometimes unlabeled data to improve performance.	They do not need any sense inventory and sense annotated corpora and less expensive.
They require large training data and expensive.	They do not give satisfactory results.	They are too expensive to carry out for natural language processing.	They are difficult to implement.

V. CONCLUSION

Word sense disambiguation is a very complex task in natural language processing and ontology. In this paper, various approaches of WSD have been described and compared. It has been concluded that Knowledge based approach use knowledge sources of MRD to decide upon the senses of words in a particular context. Supervised methods rely on large scale resources for disambiguation. The construction of the training data is time consuming and expensive. Unsupervised methods use unlabeled corpus for training which are easy to build and hence are less expensive. Semi-supervised methods use a combination of the above methods and make some assumptions about language to avoid the restrictions of few languages.

REFERENCES

[1] Agirre E, Edmonds P.G, "Word Sense Disambiguation: Algorithms and Applications", Springer, Secaucus, NJ, USA, 2006.

[2] Banerjee, S., and Pedersen, T., "An adapted lesk algorithm for word sense disambiguation using wordnet". In CICLing '02: Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing, London, UK. Springer-Verlag, Pp. 136–145, 2002.

[3] Cucerzan, R.S., Schafer, C., and Yarowsky, D., "Combining classifiers for word sense disambiguation", Natural Language Engineering, Vol. 8, No. 4, Cambridge University Press, Pp. 327-341, 2002.

[4] Diab, M., and Resnik, P., "An unsupervised method for word sense tagging using parallel corpora". In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, ACL '02, Morristown, NJ, USA, Pp. 255–262, 2002.

[5] Hwee, T., Ng Lee Yoong, K., and Tee K. Chia, "Supervised word sense disambiguation with support vector machines and multiple knowledge sources". In Proceedings of Senseval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Pp. 137–140, 2004.

[6] Ide, N., Véronis, J., "Word Sense Disambiguation: The State of the Art". In:Computational linguistics, Vol. 24, No. 1, Pp. 1-40, 1998.

[7] Nameh, M. S., Fakhrahmad, M., Jahromi, M.Z., "A New Approach to Word Sense Disambiguation Based on Context Similarity". In:Proceedings of the World Congress on Engineering, Vol. I, 2011.

[8] Navigli, R., "Word Sense Disambiguation: a Survey", ACM Computing Surveys, Vol. 41,No.2, ACM Press, Pp. 1-69, 2009.

[9] PranjalProtim Borah, GitimoniTalukdar, Arup Baruah, "Approaches for Word Sense Disambiguation : A Survey". In:International Journal of Recent Technology and Engineering (IJRTE),Volume-3, Issue-1, ISSN:2277-3878, 2014.

[10] Ravi Mante, Mahesh Kshirsagar, Prashant Chatur, "A review of literature on Word Sense Disambiguation". In:International Journal of Computer Science and Information Technologies, Vol. 5 (2), 1475-1477, 2014.

[11] Roshan G. Belsare, Shrikant P. Akarte, "A Review on Strategies of Word Sense Disambiguation". In: International Journal of Science and Engineering Applications Volume 5– No.4, ISSN:2319-7560, 2016

[12] Samit Kumar, Neetu Sharma, Niranjan, S., "Word Sense Disambiguation Using Association Rules: A Survey". In:International Journal of Computer Technology and Electronics Engineering (IJCTEE) Volume 2, Issue 2, 2012.

[13] William Gale, Kenneth Church, and David Yarowsky, "A method for disambiguating word senses in a large corpus", Computers and the Humanities, 26:415-439, 1992.