Word Sense Disambiguation: Hybrid Approach with Annotation Up To Certain Level – A Review

Mr. Roshan R. Karwa[1], Mr.M.B.Chandak[2]

M.Tech Scholar, CSE Department, SRCOEM, Nagpur, India[1]

Associate Professor and Head, CSE Department, SRCOEM, Nagpur, India[2]

Abstract- Word sense disambiguation is concerned with determining correct meaning of word in a given particular context. Disambiguation of a word is required in Machine Translation for lexical choice for words that have different translations for different senses and that are potentially ambiguous within a given domain, Information Retrieval for Resolving ambiguity in questions and in Information Extraction for distinguish between specific instances of concepts:. It is one of the challenging issues in Natural Language Processing which is ability of computer program being able to processes human like language like Hindi, English, and French etc. This paper presents a review on methods for WSD.

Keywords: Word Sense Disambiguation; Corpora; Natural Language Processing; Knowledge Sources

I. INTRODUCTION

In a language, Words can have more than one meaning. For example, Mouse in English can either have an animal meaning or a device meaning. To know correct meaning of particular word in a given context is largely unconscious and automatic in human but it's quite tough for computers as it has the shortage of real world knowledge necessary between word meanings i.e. Computer program has no basis for knowing which one is appropriate. Hence, determining correct meaning for words in context is called Word Sense Disambiguation. Important step in Word Sense Disambiguation are as follows: given a set of word, a classifier is applied which makes use of one or more sources of Knowledge to find out the most appropriate senses with words in context. Sources is of two types, one is corpus evidence based which is either unlabelled or annotated with word senses, and other is dictionaries related machine readable dictionaries, dictionaries, thesauruses etc. Without knowledge sources, it is difficult for both humans and machines to identify the correct sense i.e. meaning. Several WSD techniques have been proposed in the past ranging from knowledge based to supervise to unsupervised methods. Supervised and unsupervised rely on corpus evidence. Knowledge based relies on knowledge resources like Machine Readable dictionaries, dictionaries, thesauruses. Supervised with fewer annotation is to be used as fully supervised requires a large annotated corpus [1].

This paper is organized as follows: Section 2 comprises basic elements of WSD; section 3 comprises different methods of WSD task and Conclusion is in section 4.

II. BASIC ELEMENTS OF WSD

WSD can be outlined as a classification task: word senses are the classes, and a self regulating classification method is used to assign each occurrence of a word to one or more classes based on the evidence from the context and from external knowledge sources [2].

A word sense is a correct meaning of a word. Consider the following two sentences, One is "Mouse is not working" and other is "Mouse is the vehicle of Lord Ganesha." The word MOUSE is used in two senses. One is of type of device and other is of animal. Selection of appropriate word sense is one of the elements. A Knowledge sources provide data which are essential to associate senses with words. Sources is of two types, one is corpus evidence based which is either unlabelled or annotated with word senses, and other is dictionaries related machine readable dictionaries, dictionaries, thesauruses etc. As text is in unstructured form, Preprocessing of the input text is usually performed, which includes the following steps: Converting word to lower case, remove punctuations, stemming, removing stop words and parsing. The final step is the choice of a classification approach.
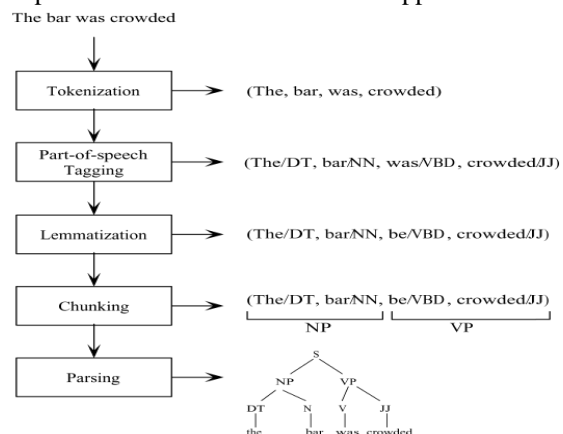

Fig.1. Preprocessing (Representation of context)

III. RELATED WORK IN WSD

There are following methods used to perform disambiguation of word in a particular context

A. Knowledge based learning Methods

The aim of knowledge-based or dictionary-based WSD is to exploit knowledge resources (such as dictionaries, thesauri, ontologies etc.; to ascertain the senses of words in context. Different integration of resources used to accumulate the information to get the sense of the word. In NLP, we referred these knowledge bases as a Lexical Knowledge Base (LKB) [10].

Word Sense Disambiguation with Conceptual Density method [3] selects a sense based on the relatedness of that word-sense to the context. How close the concept represented by the word and the concept represented by its context words is Conceptual Distance)

Selectional preferences [4] capture information about the possible relations between word categories, and represent commonsense knowledge about classes of concepts. DRIVE-VEHICLE, PLANT-TREES, 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

Overlap based approaches is depend on the finding the overlap between features of the senses definitions of two or more target words (Lesk algorithm) [5].

Problems with knowledge based approach: dictionary is restrained for sense of target word. It does not have plenty material to create classifier.

B. Supervised Learning Methods

Supervised WSD uses machine-learning techniques rely on corpus evidence for inducing a classifier from manually sense-annotated corpus, includes training and testing module where Training module requires a sense annotated training corpus from which syntactic and semantic features are picked up to build a classifier using machine learning techniques such as support vector machine. In testing module, classifier extracts the winner i.e. best sense for a word on the basis of its surrounding words [10].

Decision list method [6] is relied on 'One sense per collocation' property as nearby words providing strong and uniform hint as to the sense of a target word. Calculate the log-likelihood ratio. Higher the log-likelihood value, more the foretelling evidence. Collocations are ordered in a decision list, with most foretelling collocations ranked highest [2]

A Naive Baye's classifier [7] is a simple probabilistic approach based on the application of mathematical Baye's theorem. It depends on the calculation of the conditional probability of each sense of a word given the features in the context [2]. The sense $S\wedge$ which maximizes the following formula is chosen as the most proper sense in context.

$$S\wedge = \underset{Si \in SensesD(w)}{argmax} \quad P(Si \mid f1, f2, \ldots, fm)$$

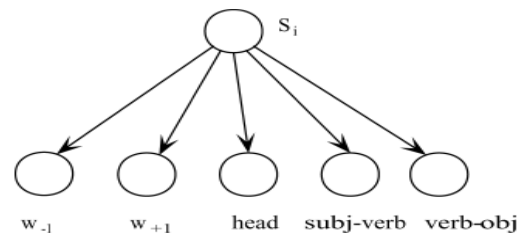Where f1, f2 .. fm are features for Naïve Baye's classifier.

Fig 2 . Naïve Baye's Feature set for WSD

Exemplar-based learning is a supervised algorithm in which the classification model is built from examples. The model preserves examples in memory as points in the feature space and, as new examples are subjected to classification, they are gradually added to the model [7].
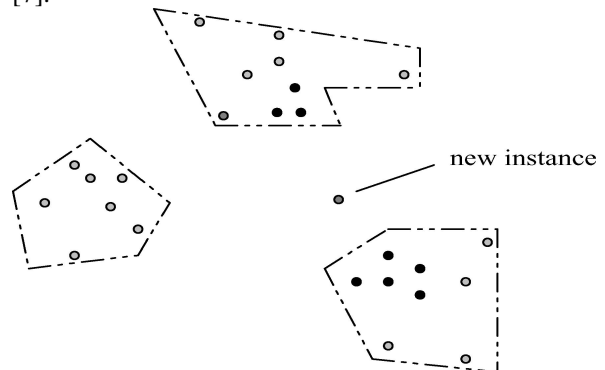

Fig 3. Exemplar based learning

Support Vector Machines (SVM) is based on binary classifier which finds hyper plane with the largest margin that separates training example into two classes [2].

Problem with many supervised WSD system is that the effort of creating the training corpus- annotated sense marked corpus has always been a matter of concern.

C. Unsupervised Learning Methods

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 neighboring 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, first, 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

IV. CONCLUSIONS

Since knowledge Based requires exhaustive knowledge resources, supervised learning method requires large annotated data due to which Knowledge acquisition problem is there and in unsupervised approach, the training data is totally unlabelled and to form cluster from the corpus is quite difficult. So effort should be made to use classifier such as Hybrid approach with annotation up to certain level.

REFERENCES

[1] Mitesh M. Khapra, Anup Kulkarni, Saurabh Sohoney, and Pushpak Bhattacharyya. 2010. All words domain adapted WSD: Finding middle ground between Supervision and unsupervision. In Jan Hajic, Sandra Carberry, and Stephen Clark, editors, ACL, pages 1532-1541.

[2] Navigli, roberto, "*word sense disambiguation: a survey*", *ACM* computing surveys, 41(2), ACM press, pp. 1-69, 2009.

[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] Ping Chen and Chris Bowes, *University of Houston-Downtown and* Wei Ding and Max Choly, *University of Massachusetts, Boston* Word Sense Disambiguation with Automatically Acquired Knowledge, 2012 IEEE INTELLIGENT SYSTEMS published by the IEEE Computer Society.