Approaches for Word Sense Disambiguation: Current State of The Art

K.Neeraja
Research Scholar, Dept of CSE, JNTUCEH, Hyderabad

Dr. B. Padmaja Rani
Professor, Dept of CSE, JNTUCEH, Hyderabad

Abstract – Human languages includes many ambiguity words, each word will have different meaning or sense. Identifying correct sense of the word is purely depended on the context in which the word appears. There are many approaches to find the correct sense of the word, Word sense disambiguation is the process of differentiating among senses of words. WSD plays a vital role to reduce the ambiguity about the words in the telugu language and the dictionary is Word Net. There are many approaches for word sense disambiguation for telugu nouns. In this paper we discuss about the current state of the art of WSD and we concluded the problem of word sense disambiguation by a combination of different machine learning algorithms.

Keywords – WSD, Ambiguity, Context, Machine Learning, Word Net.

I. INTRODUCTION

Word Sense Disambiguation is an open problem in NLP Natural Language Processing. Ambiguity is the common problem in all natural languages. The main objective of WSD is identifying the correct sense of word. The correct sense of an ambiguous word can be selected based on the context where it occurs. The problem is to assign the appropriate sense to the ambiguous word in a given context. The knowledge acquisition is a bottle neck in WSD.

Since the 1950s, many approaches have been proposed for assigning senses to words in context. Currently, there are two main methodological approaches in this area: knowledge-based and corpus-based methods. Knowledge-based methods use external knowledge resources, like thesaurus, dictionaries, and corpora, which define explicit sense distinctions for assigning the correct sense of a word in context. Corpus-based methods use machine-learning techniques to induce models of word usages from large collections of text examples. Corpus-based methods use semantically annotated corpora (most commonly word net) to train machine learning algorithms. The Both knowledge-based and corpus-based methods present different benefits and drawbacks. Common problems faced in natural language processing are data sparseness and inconsistency in vocabulary.

Recently, more research is going on Graph based approach for Knowledge based disambiguation. Graph based techniques consider all the different senses for ambiguity and analyses the relation between them with respect to the whole graph. In this graph each node is assigned with a sense, and the edges in the graph represents the relations between the pairs of nodes is nothing but pair of senses. By applying Ranking algorithm over the graph we can provide correct sense of the ambiguity word.

II. CURRENT STATE OF THE ART

In this section Current State of the Art shows the up to date research approaches and some of the issues are discussed.

Walker [3,40] proposed an algorithm where each word is assigned to one or more subject categories in the thesaurus. There are several subjects assigned with a word then it is assumed that they correspond to different senses of the word. Black achieved accuracy of 50% by applying walker’s approach to five different words.

Wilk [4,40] suggested a context vector approach that expand the glosses with related words which allows for matching to be based on one or more words where dictionary glosses are too short to result reliable disambiguation, by using the Longman’s dictionary of contemporary English (LDOCE). Walker’s approach has controlled definition vocabulary of approx 2200 words which increase the likelihood of finding overlap among word sense.

Lesk [5] developed an algorithm to be used to disambiguate all the words in a sentence at once, or should it proceed sequentially, from one word to the next. If it did proceed sequentially, should the previously assigned senses influence the outcome of the algorithm for following words.

Quillian[6] explained how to use the content of a machine readable dictionary(MRD) to make inferences about word meaning and proposed the semantic network representation of dictionary contents. Where each meaning of the word is represented by a node, for defining the concept in the dictionary; all the words are connected to this node. Content words in the definitions are in turn connected to the words that are used to define them to create a large web of words.

Cowie[7,40] explained that simulated annealing method is used to search the senses in sentence of all words, he said that the Lesk algorithm is capable of disambiguation all the words in the sentence simultaneously. Computation complexity of such an undertaking is enormous and makes it difficult in practice.

Azzini, C. da Costa Pereira, M. Dragoni, and A. G. B. Tettamanzi[8] proposed a supervised approach. The viability of the approach has been demonstrated through experiments carried out on a representative set of...
polysemous words. This approach to word sense disambiguation based on neural networks combined with evolutionary algorithms. Large tagged datasets for every sense of a polysemous word are considered, and used to evolve an optimized neural network that correctly disambiguates the sense of the given word considering the context in which it occurs.

Gerard Escudero, Lluís M’arquez and German Rigau [9] described an experimental comparison between two standard supervised learning methods, namely Naïve Bayes and Exemplar–based classification, on the Word Sense Disambiguation (WSD) problem. The aim of the work is twofold. Firstly, it attempts to contribute to clarify some confusing information about the comparison between both methods appearing in the related literature. In doing so, several directions have been explored, including: testing several modifications of the basic learning algorithms and varying the feature space.

Kozima and Furugori [10] constructed a network that consist of nodes representing the controlled vocabulary and links to show the co-occurrence of these words in glosses. They used LDOC glosses and define a measure based on spreading activation that results in a numeric similarity score between two concepts.

Pedersen, Banerjee and Patwardhan [11] suggested that Semantic relatedness to perform word sense disambiguation is measured by an algorithm. It finds its root in the original Lesk algorithm which disambiguates a polysemous word. It picks that sense of the target word whose definition has the most words in common with the definitions of other words in a given window of content. Lesk’s intuition was that related word senses will be defined using similar words. The overlap in their definitions will indicate their Relatedness, an algorithm that performs disambiguation using any measure, that returns a relatedness or similarity score for pairs of word senses.

Sussna [12,42,43,44] proposed a disambiguation algorithm assigns a sense to each noun in a window of context by minimizing a semantic distance function among their possible senses. While this is quite similar to our approach of disambiguation. His disambiguation algorithm is based on a measure of relatedness among nouns that he introduces. This measure requires that weights be set on edges in the Word-Net noun hierarchy, based on the type of relation the edge rep-resents. His measure accounts for is-a relations, as well as has-part, is-a-part-of, and antonyms.

Agirre and Rigau [13] introduced a similarity measure based on conceptual density and apply it to the disambiguation of nouns. It is based on the is-a hierarchy in WordNet, and only applies to nouns. This measure is similar to the disambiguation technique proposed by Wilks, in that it measures the similarity between a target noun sense and the nouns in the surrounding context.

Rivest [14] in 1987 is proposed a decision list algorithm. It describes an ordered set of rules for categorizing test instances (in the case of WSD, for assigning the appropriate sense to a target word). It can be seen as a list of weighted “if-then-else” rules.

Kelly and Stone [15] in 1975 proposed decision tree algorithm. It explores a predictive model used to represent classification rules with a tree structure that recursively partitions the training data set. Each internal node of a decision tree represents a test on a feature value, and each branch represents an outcome of the test. A prediction is made when a terminal node (i.e., a leaf) is reached.

Rion Snow Sushant Prakash, Daniel Jurafsky, Andrew Y. Ng [16] formulated sense merging as a supervised learning problem, exploiting human-labeled sense clustering as training data. They train a discriminative classifier over a wide variety of features derived from WordNet structure, corpus-based evidence, and evidence from other lexical resources.

Naïve Bayes [25] proposed a probabilistic classifier algorithm based on the application of Bayes’ theorem. McCul-loch and Pitts [41] in 1943 proposed a neural network which is an interconnected group of artificial neurons that uses a computational model for processing data based on a connectionist approach. Pairs of input feature, desired response are input to the learning program. The aim is to use the input features to partition the training contexts into non overlapping sets corresponding to the desired responses.

Cottrell [17] in 1989 employed neural networks to represent words as nodes: the words activate the concepts to which they are semantically related and vice versa. The activation of a node causes the activation of nodes to which it is connected by excitatory links and the deactivation of those to which it is connected by inhibitory links (i.e., competing senses of the same word).

Veronis and Ide [18] in 1990 built a neural network from the dictionary definitions of the Collins English Dictionary. They connect words to their senses and each sense to words occurring in their textual definition. Tsatsaronis et al. [19] in 2007 successfully extended their approach to include all related senses linked by semantic relations in the reference resource that is WordNet.

Towell and Voorhees [20] in 1998 found that neural networks perform better without the use of hidden layers of nodes and used perceptrons for linking local and topical input features directly to output units (which represent senses).

Boser et al. [21] in 1992 is based on the idea of learning a linear hyper plane from the training set that separates positive examples from negative examples. The hyper plane is located in that point of the hyperspace which maximizes the distance to the closest positive and negative examples (called support vectors). In other words, support vector machines (SVMs) tend at the same time to minimize the empirical classification error and maximize the geometric margin between positive and negative examples.

Klein and Florian et al. [28] in 2002 studied the combination of supervised WSD methods, achieving state-of-the-art results on the Senseval-2 lexical sample task. Brody and Navigli et al.[29,45,46] in 2006 reported a study on ensembles of unsupervised WSD methods. When employed on a stand-ard test set, such as that of the Senseval-3 all-words WSD task.

Yee Seng Chan and Hwee Tou Ng, David Chiang[30] presented conflicting evidence on whether word sense disambiguation (WSD) systems can help to improve the performance of statistical machine translation (MT) systems. In this paper, we successfully integrate a state-of-the-art WSD system into a state-of-the-art hierarchical phrase-based MT system. They show for the first time that integrating a WSD system improves the performance of a state-of-the-art statistical MT system on an actual translation task. Furthermore, the improvement is statistically significant.

Brin and Page[31,32,47] in 1998 explored an alternative graph-based algorithm for inducing word senses is Pa-eRank. PageRank is a well-known algorithm for computing the ranking of web pages and is the main ingredient of the Google search engine. It has been employed in several research areas for determining the importance of entities whose relations can be represented in terms of a graph.

Gale et al. [33,48] in 1992b suggested an unsupervised methods have the potential to overcome the knowledge acquisition bottleneck which is, the lack of large-scale resources manually annotated with word senses. These approaches to WSD are 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.

Schutze [34] in 1992 described a set of unsupervised approaches which are based on the notion of context clustering. Each occurrence of a target word in a corpus is represented as a context vector. The vectors are then clustered into groups, each identifying a sense of the target word. A historical approach of this kind is based on the idea of word space.

Veronis [35] in 2004 was proposed an adhoc approach called HyperLex. Here a co occurrence 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 co occur in the same paragraph. Each edge is assigned a weight according to the relative co occurrence frequency of the two words connected by the edge.

P.Tamilselvi, S.K.Srivatsa[36] implemented disambiguation system using Neural Networks with enormous number of features, accuracy measured from 33.93% to 97.40% for words with more than two senses and 75% of accuracy for words with two senses. They developed the system with three different set of features with three different distance measuring functions combined with three different classifiers for word sense disambiguation.

M. Nameh, S.M. Fakhrhammad, M. Zolghadri Jahromi [37] presented a supervised learning method for WSD, which is based on Cosine Similarity. As the first step, they extract two sets of features; the set of words that have occurred frequently in the text and the set of words surrounding the ambiguous word. Then they presented the results of evaluating the proposed schemes and illustrate the effect of weighting strategies proposed.

A.R.Rezapour, S. M. Fakhrhammad and M. H. Sadreddini[38] presented a K-Nearest Neighbor algorithm, which is a supervised learning method for WSD. They extracted two sets of features; the set of words that have occurred frequently in the text and the set of words surrounding the ambiguous word. They proposed a feature weighting strategy to improve the classification accuracy,. The results are encouraging comparing to state of the art.

Arindam Chatterjee, Salil Joshi, Pushpak Bhattacharyya, Diptesh Kanojia and Akhlesh Meena [19] shows that in almost all disambiguation algorithms, the sense distribution parameter \( P(S/W) \), where \( P \) is the probability of the sense of a word \( W \) being \( S \), plays the deciding role. The widely reported accuracy figure of around 60% for all-words-domain-independent WSD is contributed to mainly

III. Conclusion

In this paper we presented the current state of the art about word sense disambiguation and we understood and analyzed many approaches to avoid ambiguity of words. Knowledge based disambiguation have gained much importance in NLP. In this graph based approach is advanced technique and found the right path towards WSD. So we are interested about our future work in this direction. We have presented several approaches and analyzed their performance and drawbacks. The above literature survey concludes that supervised learning algorithms performances on WSD are more appropriate than unsupervised algorithms. In some research they combined two algorithms to optimize the performance. In some research papers they have performed clustering also to improve the performance. But still there is much more to do to improve the performance of machine learning algorithms for word sense disambiguation. Using machine learning algorithm for some specific data of words reaches 86.74 % accuracy Knowledge-based approaches achieve good performance, even though below standard WSD...
baselines. Finally, we have concluded several directions for further research.

ACKNOWLEDGMENT

We are deeply indebted to all the esteemed authors in a reference list, who greatly helped me to improve this research article in a better and right path direction. Hopefully, this article will contribute to further progress on WSD.

REFERENCES

[16] Rion Snow Sushant Prakash, Daniel Jurafsky, Andrew Y. Ng , ”Learning to Merge Word Senses”, Computer Science Department Stanford University, 2007.
[34] Schutze, H. Dimensions of meaning. In Supercomputing '92: Pro-ceedings of the 1992 ACM/IEEE Conference on


[40] Boshra F. ZoponAl_Bayaty and Dr.Shashank Joshi”Word Sense Dis-ambiguation (WSD) and Information Retrieval (IR): Literature Re-view” ijarcsse,Volume 4, Issue 2,ISSN: 2277 128X, February 2014.


AUTHOR'S PROFILE

K. Neeraja
completed her M.Tech. (CSE) from JNTUH.and Pursuing Ph.D (CSE) from JNTUH. She is a member of CSI. She is having 12 years of Teaching Experience. To her credit she is having 06 international journals / conference papers. Her area of Research is Information Retrieval Systems and NLP.

Dr. B. Padmaja Rani
received B.Tech(ECE) from Osmani University, M.Tech(CSE) from JNTU Hyderabad and Ph.D from Jawaharlal Nehru Technological University. She is a member of ISTE, CSI; she is currently working as Professor in Dept. of CSE, College of Engineering, JNTU Hyderabad. She is having 18 + years of Professional Teaching Experience and 5+ years of research experience. To her credit she is having 25 international Journals/Conference Papers. Her area of Research interest includes Information Retrieval in Indian languages, NLP.