CAREER POINT UNIVERSITY HAMIRPUR

CPUH-Research Journal: **2015,** ***1(2),*** 54-58
ISSN (Online): 2455-6076
http://www.cpuh.in/academics/academic_journals.php

Word Sense Disambiguation and Its Approaches

Vimal Dixit[1*], Kamlesh Dutta[2] and Pardeep Singh[2]

[1 &2] *Department Of Computer Science and Engineering, National Institute of Technology, Hamirpur-177005, India*

[*] *Correspondance E-mail: vimaldixit2@gmail.com*

ABSTRACT: Word Sense Disambiguation (WSD) is an important but challenging technique in the area of natural language processing (NLP). Hundreds of WSD algorithms and systems are available, but less work has been done in regard to choosing the optimal WSD algorithms. This paper summarizes the various Approaches used for WSD and classifies existing WSD algorithms according to their techniques. We have discussed about the machine learning and Dictionary–based approaches for WSD. Various supervised learning, unsupervised learning and semi-supervised techniques have been discussed. WSD is mainly used in Information Retrieval (IR), Information Extraction (IE), Machine Translation (MT), Question Answering (QT), Content Analysis, Word Processing, Lexicography and Semantic Web.

Keywords: Word Sense Disambiguation (WSD); Natural Language Processing (NLP); supervised, unsupervised; knowledge Base; information retrieval; information extraction; machine translation, context; ambiguity; polysemous words Machine readable dictionary and WordNet.

INTRODUCTION: There are words in Natural languages which have different meaning for different context but they are spelled same. Those words are called polysemous words. Word sense disambiguation [2] (WSD) is the solution to the problem. Word Sense Disambiguation is a task of finding the correct sense of the words and automatically assigning its correct sense to the words which are polysemous in a particular context. WSD [18] is an important but challenging technique in the area of natural language processing (NLP). It is necessary for many real world applications such as machine translation (MT), semantic mapping (SM), semantic annotation (SA), and ontology learning (OL). It is also believed to be helpful in improving the performance of many applications such as information retrieval (IR), information extraction (IE), and speech recognition (SR).Many Natural languages like English, Hindi, French, Spanish, Chinese, etc. are the languages which have some words whose meaning are different for same spelling in the different context (polysemous words). In English, Words likes Run, Execute, book, etc. can be considered example of polysemous words. Human beings are blessed with the learning power. They can easily find out what is the correct meaning of a word in a context. But for computer it is a difficult task. So, we need to develop an automatic system which can perform like humans do i.e. the system which can find out the correct meaning of the word in particular context and automatically assign the optimal sense to the target word. Context is the text or words which are surrounding to the ambiguous word. Using the context, human can easily sense the correct meaning of the word in that context. So we also need the computer to follow some rules using which the system can evaluate the absolute meaning out of multiple meanings of the word. If we consider a text T a sequence of words i.e. Word $_1$, Word $_2$, Word $_3$……..Word $_n$. Then, WSD is a task to assign the correct sense for all or some words in the text T. The conceptual model [3] for WSD is shown in figure1.

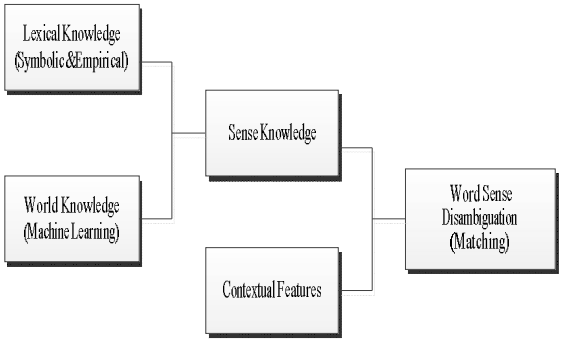

Figure 1: Word Sense Disambiguation Conceptual Model.

Two main approaches which are used to WSD are Machine-Learning based approaches and Dictionary Based approaches. Machine-Learning based approaches, the systems are trained to perform the task of WSD. A classifier is used to learn features and assigns senses to unseen examples. In these approaches, the initial input is the word to be disambiguated called target word, and the text in which it is embedded, called as context. Dictionary Based approaches, all the senses of a word that need to be disambiguated are retrieved from the dictionary. These senses are then

compared to the dictionary definitions of all the remaining words in context. We can further classified the WSD approaches to Word sense Disambiguation are Deep approach and Shallow approach.

WSD APPROACHES: There are two approaches that are followed for Word Sense Disambiguation (WSD): Machine-Learning Based approach and Knowledge Based approach. In Machine learning-based approach, systems are trained to perform the task of word sense disambiguation. In Knowledge based approach, it requires external lexical resources like Word Net, dictionary, thesaurus etc.

Machine Learning Based Approach: A classifier is used to learn features and assigns senses to unseen examples. In these approaches, the initial input is the word to be disambiguated called target word, and the text in which it is embedded, called as context. In this approach features are themselves served by the words. The value of feature is the number of times the word occurs in the region surrounding the target word. The region is often a fixed window with target word as center. Three types of techniques of machine learning based approaches are: supervised techniques, unsupervised techniques, and semi-supervised techniques.

Supervised Techniques: It uses machine-learning techniques [5] for inducing a classifier from manually sense-annotated data sets. Usually, the classifier (often called word expert) is concerned with a single word and performs a classification task in order to assign the appropriate sense to each instance of that word. The training set used to learn the classifier typically contains a set of examples in which a given target word is manually tagged with a sense from the sense inventory of a reference dictionary. Let us take the example of the learning process of a small child. The child doesn't know how to read/write. He/she is being taught by the parents at home and then by their teachers in school. The children are trained and modules to recognize the alphabets, numerals, etc. Their each and every action is supervised by the teacher. Actually, a child works on the basis of the output that he/she has to produce. Similarly, a word sense disambiguation system is learned from a representative set of labeled instances drawn from same distribution as test set to be used. Basically this WSD algorithm gives well result than other approaches. Methods in Supervise WSD are as follow:

Decision Lists: It is an ordered set of rules for categorizing test instances (in the case of WSD, for assigning the appropriate [24] sense to a target word). It can be seen as a list of weighted [if-then-else] rules. A training set is used for inducing a set of features. When any word is considered, first its occurrence is calculated and its representation in terms of feature vector is used to create the decision list, from where the score is calculated. The maximum score for a vector represents the sense.

Decision Tree: A decision tree [17] divides the training data in a recursive manner and represents the rules for classification in a tree structure. The internal nodes represent test on the features and each branch shows how the decision is being made and the leaf node refers to the outcome or prediction. An example of a decision tree for WSD is described in the Figure 2.

The noun sense of the ambiguous word "bank" is classified in the sentence, "I will be at the bank of Narmada River in the afternoon" In the Figure 2, the tree is created and traversed and the selection of sense bank/RIVER is made. Empty value of leaf node says that no selection is available for that feature value.

Naïve Bayes: A Naive Bayes [25] classifier is a simple probabilistic classifier based on the application of Bayes' theorem. It relies on the calculation of the conditional probability of each sense S_i of a word w given the features f_j in the context. The sense S which maximizes the following formula is chosen as the most appropriate sense in context.

$$\hat{S} = \underset{S_i \in Senses_D(w)}{\operatorname{argmax}} P(S_i \mid f_1, \ldots, f_m) = \underset{S_i \in Senses_D(w)}{\operatorname{argmax}} \frac{P(f_1, \ldots, f_m \mid S_i) P(S_i)}{P(f_1, \ldots, f_m)}$$

$$= \underset{S_i \in Senses_D(w)}{\operatorname{argmax}} P(S_i) \prod_{j=1}^{m} P(f_j \mid S_i) \quad \ldots\ldots (1)$$

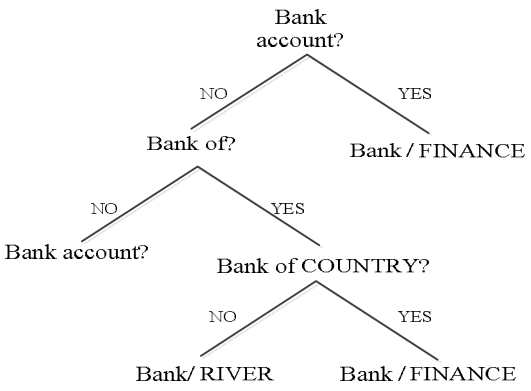

Figure 2: An example of a decision tree.

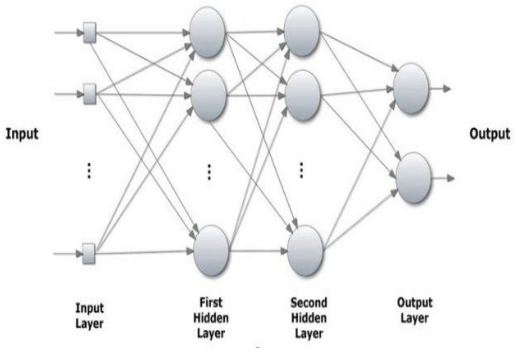

Figure 3: Neural Network Conceptual Model.

Neural Networks: Neural networks [23] processes information based on computational model of connectionist approach. The input includes the input features and the target output. The training dataset is divided into sets which are non- overlapping based on desired responses. When the network encounters new input pairs the weights are adjusted so that the output unit giving the target output has the larger activation.

Un-supervised Techniques: Unsupervised approach unlike supervised approach does not need the hand labeled knowledge of sense information in large scale resources for the disambiguation. It is based on the fact that words having similar senses will have similar surrounding words. Word senses are derived by forming clusters of occurrences of words and the task is to classify the new occurrence to the derived clusters. This approach instead of assigning sense labels detects the clusters.

Context Clustering: In this method is based on clustering techniques [15] in which first context vectors are created and then they will be grouped into clusters to identify the meaning of the word. This method uses vector space as word space and its dimensions are words only. Also in this method, a word which is in a corpus will be denoted as vector and how many times it occurs will be counted within its context [16]. After that, co-occurrence matrix is created and similarity measures are applied. Then discrimination is performed using any clustering technique.

Word Clustering: In this technique words having similar meanings are assigned to the same cluster. One of the approaches [12] mentioned in was to find the sequence of words same as the target word. The similarity between the words is given by syntactical dependency. If W consist of words which are similar to wm then a tree is formed initially with only one node wm and a node w_i will have a child node w_m when w_i is found to be the word with most similar meaning to wm. Another approach mentioned in called clustering by committee algorithm [14] represents each word as a feature vector. When target words are encountered a matrix called similarity matrix S_{mn} is constructed whose each element is a similarity between two words wm and w_n. In the subsequent step of this algorithm committees are formed for a set of words W in recursive manner. The clustering algorithm then tries to find those words not similar to the words of any committee. These words which are not part of any committee are again used to form more committees. In the final step each target word belonging to W will be a member of committee depending on its similarity to the centroid of the committee. The clustering technique used is average-link clustering.

Co-occurrence Graphs: This method creates co-occurrence [13] graph with vertex V and edge E, where V represents the words in text and E is added if the words co-occur in the relation according to syntax in the same paragraph or text. For a given target word, first, the graph is created and the adjacency matrix for the graph is created. After that, the Markov clustering method is applied to find the meaning of the word. Each edge of graph is assigned a weight which is the co-occurring frequency of those words. Weight for edge {m,n} is given by the formula:

$$w_{mn} = 1 - \max\{P(w_m|w_n), P(w_n|w_m)\} \quad \text{.....(2)}$$

Where $P(w_m|w_n)$ is the $freq_{mn}/freq_n$ where $freq_{mn}$ is the co-occurrence frequency of words w_m and w_n, $freq_n$ is the occurrence frequency of w_n. Word with high frequency is assigned the weight 0, and the words which are rarely co-occurring, assigned the weight 1. Edges, whose weights exceed certain threshold, are omitted. Then an iterative algorithm is applied to graph and the node having highest relative degree, is selected as hub. Algorithm comes to an end, when frequency of a word to its hub reaches to below threshold. At last, whole hub is denoted as sense of the given target word. The hubs of the target word which have zero weight are linked and the minimum spanning tree is created from the graph. This spanning tree is used to disambiguate the actual sense of the target word.

Semi-Supervised Techniques: In semi-supervised learning techniques, the information is present like in supervised but might be less information is given. Here only critic information is available, not the exact information. For example, the system may tell that only particular about of target output is correct and so. The semi-supervised or minimally supervised methods are gaining popularity because of their ability to get by with only a small amount of annotated reference data while often outperforming totally unsupervised methods on large data sets. There are a host of

diverse methods and approaches, which learn important characteristics from auxiliary data and cluster or annotate data using the acquired information.

Dictionary Based Approach: Knowledge based approach based on knowledge resources of machine readable dictionaries in form of corpus, WorldNet etc. they may use grammar rules for disambiguation. The aim of Knowledge based approach (Dictionary based approach) WSD is to exploit knowledge resources to infer the senses of words in context. The knowledge resources are dictionaries, thesauri, ontology's, collocations etc. The above methods have lower performance than their supervised alternative methods but they have an advantage of a wider range.

Overlap Based Approaches: This approach calls for the requirement of machine readable dictionary (MDR). It includes determination of the different features of the senses of words which are ambiguous along with features of the words in the context.

Lesk's algorithm: The Lesk's algorithm [7] used by overlap based approach can be stated as if W is a word creating disambiguation, C be the set of words in the context collection in the surrounding, S be the senses for W, B be the bag of words derived from glosses, synonyms, hyponyms, glosses of hyponyms, example sentences, hypernyms, glosses of hypernyms, meronyms, example sentence of meronyms, example sentence of hypernyms, glosses of meronyms then use the interaction similarity rule to measure the overlap and output the sense which is the most probable having the maximum overlap.

PINE
1. kinds of evergreen tree with needle-shaped leaves
2. waste away through sorrow or illness

CONE
1. solid body which narrows to a point
2. something of this shape whether solid or hollow
3. fruit of certain evergreen trees

As can be seen, the best intersection of pine and cone is

Pine #1 ⋂ Cone #3 = 2 …..(3)

Walker's approach: This algorithm can be stated as each word is assigned to one or more categories of subjects in the thesaurus. Different subjects are assigned to different senses of the word.

Selection Preferences: Selection preferences [17] find information of the likely relations of word types, and denote common sense using the knowledge source. For example, Modeling-dress, Walk-shoes are the words with semantic relationship. In this approach improper word senses are omitted and only those senses are selected which have harmony with common sense rules. The basic idea behind this approach is to count how many times this kind of word pair occurs in the corpus with syntactic relation. From this count, senses of words will be identified. There are other methods, which can find this kind of relation among words using conditional probability.

CONCLUSION: This paper summarized the various approaches used for WSD and classified existing WSD algorithms according to their techniques. In this paper we have put forwarded a survey of comparison of different approaches available in word sense disambiguation with primarily focusing on the Machine Learning Approaches and Dictionary based approaches knowledge based. We concluded that supervised approach is found to perform better but one of its disadvantage is the requirement of a large corpora without which training is impossible which can be overcame in unsupervised approach as it does not rely on any such large scale resource for the disambiguation. Knowledge based approach on the other hand makes use of knowledge sources to decide upon the senses of words in a particular context provided machine readable knowledge base is available to apply.

REFERENCES:

1. Manning, C. D. and Schutze, H. 1999. Foundations of Statistical Natural Language Processing. *MIT Press, Cambridge, Massachusetts London, England*.
2. Tacoa, F., Bollegala, D. and Ishizuka, M. 2012. A Context Expansion Method for Supervised Word Sense Disambiguation. *In IEEE Sixth International Conference on Semantic Computing.*
3. Sreedhar, J. Viswanadha, S., Raju, A., Babu, V., Shaik, A. and Kumar, P. 2012. Word Sense Disambiguation: An Empirical Survey. *International Journal of Soft Computing and Engineering (IJSCE).* 2.
4. Agirre, E. and Edmonds, P. 2006. Word Sense Disambiguation: Algorithms and Applications *(Text, Speech and Language Technology). Springer-Verlag New York, Inc. Secaucus, NJ,* USA.
5. Navigli, R. 2009. Word Sense Disambiguation: A Survey. *Universita di Roma La Sapienza, ACM Computing Surveys*. 41.
6. Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from a ice cream cone. *In Proceedings of SIGDOC* '86.

7. Jurafsky, D. and Martin, J. H. 2008. An Introduction to Natural Language processing, *Computational Linguistics, and Speech Recognition. Pearson Education*.
8. Kolte, S. G. and Bhirud, S. G. 2008. Word Sense Disambiguation using Word Net Domains. *In Proceedings of ICETET'0*8.
9. Bala, P. 2013. Word Sense Disambiguation Using Selectional Restriction. *In International Journal of Scientific and Research Publications*.
10. Zheng, Z. and Shu, Z. 2009. A New Approach to Word Sense Disambiguation in MT System. *In World Congress on Computer Science and Information Engineering*.
11. Lin, D. and Pantel, P. 2002. Discovering word senses from text", *In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Edmonton, Alta., Canada*). 613–619.
12. V´Eronis, J. 2004. Hyperlex: Lexical cartography for information retrieval. *Comput. Speech Lang*. 18: 223–252.
13. Lin., D. 1998. Automatic retrieval and clustering of similar words. *In Proceedings of the 17th International Conference on Computational* linguistics (COLING, Montreal, P.Q., Canada). 768–774.
14. Pedersen, T. and Bruce, R. 1997. Distinguishing word senses in untagged text. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP, Providence*, RI), 197–207.
15. Schutze, H. 1998. Automatic word sense discrimination. *Computat. Ling.* 24: 97–124.
16. Mooney, R. J. 1996. Comparative experiments on disambiguating word senses: An illustration of the role of bias in machine learning. *In Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP), 82–91.
17. Palta, E. 2006. *Word Sense Disambiguation, M. Tech. dissertation, dept. CSE Indian Institute of Technology, Mumbai*.
18. Hindle, D. and Rooth, M. 1993. Structural ambiguity and lexical relations. *Computat. Ling.* 19: 103–120.
19. Resnik, P. S. 1993. Ed. Selection and information: A class-based approach to lexical relationships, *Ph.D. dissertation. University of Pennsylvania, Pennsylvania, Philadelphia*, PA, 1993.
20. Collins, M. 2004. Parameter estimation for statistical parsing models: Theory and practice of distribution free methods. *In New Developments in Parsing Technology, H. Bunt, J. Carroll, and G. Satta, Eds. Kluwer, Dordrecht, The Netherlands*, 19–55.
21. Mcculloch, W. and Pitts, W. 1943. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys*. 5: 115–133.
22. Rivest, R. L. 1987. Learning decision lists. *Mach. Learn.* 2: 229–246.
23. Marquez, L., Escudero, G., Mart`Inez, D. and Rigau, G. 2006. Supervised corpus-based methods for WSD. *In Word Sense Disambiguation: Algorithms and Applications, E. Agirre and P. Edmonds, Eds. Springer, New York, NY*, 167–216.
24. Schutze, H. 1998. Automatic word sense discrimination. *Computat. Ling*. 24: 97–124.