ISSN 2320 2602

Volume 1, No.1, November – December 2012
International Journal of Advances in Computer Science and Technology
Available Online at http://warse.org/pdfs/ijacst01112012.pdf

Sense Disambiguation Techniques: A Survey

Rekha Jain[1], Sulochana Nathawat[2], Dr. G. N. Purohit[3]
[1]Department of Computer Science, Banasthali Vidyapith, Jaipur, Rajasthan
[1]rekha_leo2003@yahoo.com
[2]Department of Computer Science, Banasthali Vidyapith, Jaipur, Rajasthan
[2]nathawat.sulochana@gmail.com
[3]Department of Computer Science, Banasthali Vidyapith, Jaipur, Rajasthan
[3]gn_purohitjaipur@yahoo.co.in

ABSTRACT

In today's era most of the people are depended on the web to search some contents. At the time of searching they never bother about ambiguities that exist between words. An ambiguous word is a word that has multiple meaning in different contexts. The sense of the word is determined by the context in which the ambiguous word appears. When the user performs the search related to ambiguous word, web displays all the results related to senses of the word. Some of them are relevant and some are irrelevant according to user perspective. Word Sense Disambiguation (WSD) is the process of identifying the senses of word in textual context, when word has multiple meanings. The purpose of the research is to elaborate the methodology, approaches of WSD that can handle all issues with better performance and accuracy. In this paper the authors are discussing both the approaches and their roles in various applications like IR, MT, IE, KM etc.

Keywords : Sense Ambiguity, Word Sense Disambiguation, Supervised, Unsupervised approaches.

1. INTRODUCTION

Ambiguity is the possibility of interpreting sense of the word used in the sentence or query in distinct ways. Ambiguous word or sentence has multiple meaning. For a particular language, grammar provides pairs of ambiguous forms with more than one meaning. For example crane word can have pairs of ambiguous forms with multiple meaning as "crane - bird", "crane-machine". There are majorly three types of ambiguity, lexical, syntactic and Semantic [1]. The lexical ambiguity refers to having more than one meaning of a word or phrase in the language to which the word belongs. For instance, the word "crane" has several distinct lexical definitions, including "bird" and "machine". The intended meaning of an ambiguous word can be identified by the contexst in which an ambiguous word is used. Lexical ambiguity can be resolved by algorithmic methods that automatically associate the appropriate meaning with a word in context [2]. Syntactic ambiguity occurs when a phrase or sentence can be parsed in more than one way. Such phrases has more than one underlying structure can be assigned different interpretations. For example, 'The girl hit the boy with a book' could mean that girl hit the boy by using book, or it could mean that girl hit the boy who is having the book.[1] Semantic ambiguity arises when a word or concept has an inherently different meaning based on widespread or informal usage. For example, with idiomatic expressions whose definitions are rarely or never well-defined [2]. "A little bird told me" is an idiomatic expression which is referred when people do not want to reveal the source of information, often gossip. The clause "bird told me" presents a statement with such wide possible interpretation as to be essentially meaningless as a bird never say something [4].

1.1 Sense Ambiguity

Word sense is most common accepted meaning of the word. A Word Sense Ambiguity is some uncertainty about the precise Word Sense [5]. Human language is ambiguous so a particular word with a particular syntactic category is associated with more than one meaning [3]. It is an important characteristic of natural language. For example, the word "cold" has various senses like a disease, a temperature sensation, or an environmental condition. The sense of the word is determined by the context in which the ambiguous word appears. "I am taking aspirin for my cold" is intended for the disease sense, "Let's go inside, I'm cold" is intended for the temperature sensation, while "It's cold today, only 1.5 degrees", refers the environmental condition sense [6].

1.2 Word Sense Disambiguation

On Web we are facing problem of word sense ambiguity while searching for ambiguous words. Web display all the results related to sense of the word. For the solution of this we use Word Sense Disambiguation. WSD is the process of identifying the sense of word in textual context, when word has multiple meaning [7]. WSD associate a word in a text or sentence having different meaning [9]. WSD is an important and open problem of natural language processing (NLP). It improved the performance of many applications such as information retrieval (IR), information extraction (IE), and speech recognition (SR). Figure 1 shows the conceptual model [12] for WSD.

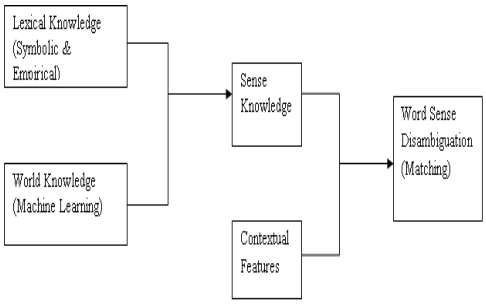

Figure 1: Conceptual Model for Word Sense Disambiguation

In 1940 WSD was first developed due to fast research in machine translation as a distinct computational task. In 1949 Warren Weaver first introduced the problem in computational context. Then in following years many researchers developed various methods to solve the problem of WSD: Artificial Intelligence based method, knowledge-based method, supervised machine learning techniques, knowledge-based systems via graph-based methods [7].

2. APPLICATIONS OF WSD

2.1. Machine Translation: It is the original and most direct application of WSD. WSD is required in MT not to predict the sense of the word but to choose the meaning of word that has different translations in different senses [11].

2.2. Information Retrieval: WSD is used to improve performance in Information Retrieval. It is clear that an automated disambiguation system should provide benefits to IR if it achieves higher precision.

2.3. Lexicography: In an application a word can have different senses so lexicon has been developed accordingly [11].

2.4. Information extraction and Knowledge Mining: WSD is used in many applications for analysis of text [11].

2.5. Speech Processing and Part of Speech tagging: WSD for speech recognition i.e. for homophones words which are spelled differently but pronounced exactly same. For example 'base' and 'bass' [9].

2.6. Text Processing: WSD has also an important role in text to speech translation i.e. words are pronounced in differently depending upon the meaning and the context in which it is used [9].

2.7. Bioinformatics: Bioinformatics research requires the relationships between genes and gene products however genes and their proteins often have the same name.

2.8. Semantic Web: The Semantic Web requires automatic annotation of documents according to reference ontology.

3. WSD METHODOLOGY

According to Ide and V'eronis(1998),WSD task involves two steps:

3.1. Sense Repository: Identify all the different meaning of all the words relevant to the text under consideration. Example from list of senses in dictionaries, from synonyms in thesaurus, from translations in a translation dictionary.

3.2. Sense Assignment: It involves the assignment of appropriate sense to each occurrence of word in textual context.

Third step also involved in which computer learns about the association meaning with the word in textual context using either machine learning or manual creation of rules [9].

4. WSD APPROACHES

There are two main approaches of WSD [99]:

4.1. Deep Approaches: This is based on world knowledge. But such knowledge is not available in computers readable format except in some limited domain so this approach is not very popular. However if such knowledge available than this approach will be much more accurate than shallow approaches.

Example: Man goes fishing for some bass.

Here knowledge is used to identify the meaning of 'bass' in sentence because one can go for fishing for a type of fish but not for low frequency sound. So here bass will refer to fish.

4.2. Shallow Approaches: This approach does not use the world knowledge. One can understand the text through the surrounding words.

Example: If 'crane' has words sky or fly nearby then it will point to bird. If 'crane' has words parts or manufacture nearby it refer to machine.

These rules can be automatically derived by the computer, using a training corpus of words tagged with their word senses. Our knowledge is limited so this approach gives better results but theoretically it is not very powerful as compared to deep approaches. It can, though, be confused by sentences like "The dog barked at the tree."

5. LEARNING METHODOLOGIES

WSD methods are classified into two types: Machine learning approaches and another is Dictionary based approaches [8]. Section 5.1 discusses the basic Machine learning approaches and section 5.2 discusses the basic Dictionary based approaches.

5.1 Machine Learning Approaches

In machine learning approaches, systems are trained to perform the task of WSD. In these approaches classifier is learned that is used to assign fixed number of senses to unseen examples. In most of these approaches the initial input are "target word" which is the word to be disambiguated and the "context" which is the portion of the text to which target word is embedded. This initial input is then processed after which it consist of fixed set of features to abstract information relevant to learning task. This task consist of two steps: selecting the relevant linguistic features and encoding them in a form usable in a learning algorithm. Linguistic features are divided into two classes: collocational and co-occurrence features. Here the collocation refers quantifiable position-specific relationship between two lexical items. Collocational feature encode information of specific positions located to the left or right of the target word. Typical features include the word, the root form of the word and the word's part of speech.

Let take an example [8] where we will disambiguate the word bass:

Example 1.1: An electric guitar and bass player stand off to one side, not really part of the scene. Here bass is the target word following feature vector consist of two words to the right and left of the target word:

[guitar, NN1, and, CJC, player, NN1, stand, VVB]

In co-occurrence feature it considers the data about the neighboring words while ignoring the exact position. In this approach the word themselves serve as features. The value of the feature is the number of times the word occurs in a fixed size window with the target word at the center. For example 1.1, a co-occurrence vector consist of following 12 most frequent word from a collocation of bass sentences: fishing, big, sound, player, fly, rod, pound, double, runs, playing, guitar, band represented by the following vector with window size of 10:

[0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0]

Categorization of "Machine learning" approaches of WSD is as: supervised, bootstrapping and unsupervised.

5.1.1 Supervised Approaches

Supervised WSD uses machine learning techniques in which inputs are manually sense-annoted data and output is a classifier system. In this approach, a sense disambiguation system is learned from a representative set of labeled instances drawn from the same distribution as the test to be used. Classifier system capable of assigning labels to new feature encoded inputs. Generally supervised approaches give better results than unsupervised approaches. Following are some of supervised techniques:

(i) Naïve Bayes Classifier: It is a probabilistic classifier based on the application of Bayes theorem. It is based on the premise that choosing the best sense for the input vector amounts to choosing the most probable sense.

$$\hat{s} = \operatorname{argmax}_{s \in S} P(S/V) \qquad (1)$$

Here S denotes the set of senses appropriate for the target, s denotes each of possible sense in S, V denotes the vector representation of the input context.
By applying Bayes rule we get,

$$\hat{s} = \operatorname{argmax}_{s \in S} \frac{P(S/V)P(S)}{P(V)} \qquad (2)$$

In example 1.1 the individual statistics might include the probability of the word playing occurring immediately right of a use of each of the bass sense or the probability of the word guitar one place to the left of a use of one of the bass senses.

(ii) Decision List Classifier: It is an ordered set of rules to which we can apply a sequence of test. It is a list of weighted "if-then-else" rules. This classifier is equivalent to case statement used in programming languages. If the test returns true then the sense associated with it will be returned else next test in sequence is applied. This will continue at the end of list where default test will return the majority sense. Decision list have two senses per word. Learning rules are ordered by:

$$\log\left(\frac{P(\text{sense}_A \mid \text{collocation})}{P(\text{sense}_B \mid \text{collocation})}\right) \quad (3)$$

(iii) Decision Trees: The set of rules are represented in the form of tree structure which recursively partition the training data set. Each internal node represents a test on feature value and each branch represents an outcome of the test and through the leaf node prediction is made. In practice it is not so popular and rarely applied to WSD. Some popular algorithms are C4.5 algorithm [Quainlan 1993], an extension of ID3 algorithm [Quainlan 1986], Mooney [1996] concluded that C4.5 algorithm to implement decision tree gives outperformed result than other supervised approaches. Figure 2 shows an example of decision tree [10]. For instance, to classify the noun bank in sentence "we sat on a bank of sand", the tree is traversed by following the no-yes-no path and sense bank/RIVER is made by this. Empty value (-) at leaf node indicates no choice for specific feature vale.

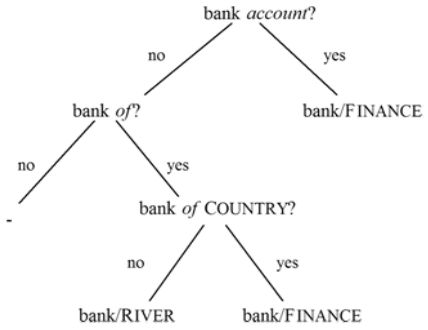

Figure 2: An example of a decision tree.

(iv) Neural Network: A neural network as an interconnected assembly of artificial neurons that uses computational model for processing data based on connectionist approach. Training data set are partitioned into non-overlapping set corresponding to desired responses. Inputs are given with adjusted weight so that desired response has larger activation than other output unit. Weights can be positive or negative. Neural networks are trained until the output of the unit that corresponds to desired response is greater than the output of the any other unit in training set. Whenever a node gets activated it causes all the nodes to be activated with which it is connected by excitory link and deactivation of nodes that are connected by inhibitory link. Veronis and Ide developed the neural network from the dictionary definition of the Collins English Dictionary. Following figure 3 [10] shows a multilayer perceptron with four feature values which outputs the corresponding value of three senses of target word in context.

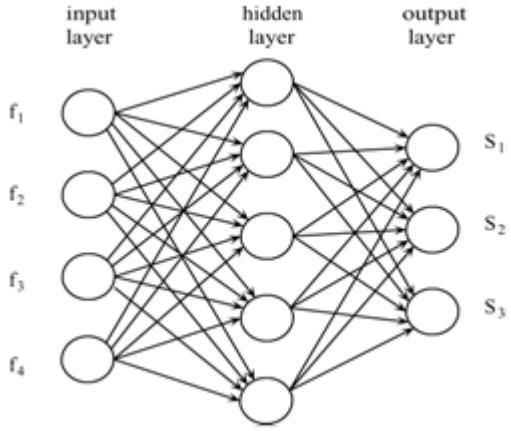

Figure 3: An illustration of a feed forward neural network for WSD with four features and three responses, each associated to a word sense.

Several studies has proved that neural network perform well compared to other supervised methods.

5.1.2 Bootstrapping Approaches

The drawback of supervised approaches is the requirement of a large sense tagged training set. Bootstrapping approach does not require the large training data set. It works on few numbers of instances of each sense of target word. Initial classifier is trained using a labeled instance which is known as seed. The task of initial classifier is to extract large training set from the remaining untagged corpus. On repeating this process we will get a series of improved classifier. This method creates a large training set from small set of seeds. The larger training set is again used to create a new more accurate classifier. In each iteration training set increases and untagged corpus size reduced. This process is repeated until low error rate on the training set is reached or until no instances from untagged corpus is above threshold. The initial seed can be generated in this approach in different ways. Hearst [1991] generates a seed set through hand label a small set of examples from initial corpus. Another technique is to search the sentence which contains the

words or phrases that are strongly associated with target sense. This is given by Yarowsky [1995] and known as one sense per collocation constraint which gives better results.

5.1.3 Unsupervised Approaches

This approach uses the sense tagged data of any type during the training. The concept behind this technique is that the words which have same sense will also have similar neighboring words. Input to this approach is unlabeled instances which are represented as feature vector. Then these are grouped into clusters based on similarity metric. From input text word sense can be assigned from cluster to which they are closest based on similarity metric. It don't use machine readable resources like dictionaries, thesauri, ontology etc. So the main disadvantage of this technique is that they can't rely on a shared reference inventory of senses. Unsupervised WSD performs word sense discrimination i.e. it divides the occurrence of word into a number of classes by determining for any two occurrences whether they belong to the same sense or not. Evaluation of these methods is more difficult. Main task of unsupervised approaches are identifying sense clusters. Problems associated with unsupervised technique are:

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.
Schutze's experiment [Schutze, 1992, 1998] has shown the application of unsupervised clustering to WSD.

5.2 Dictionary – Based Approaches

The major drawback of all the approaches is the scale. To create a classifier it needs a great amount of work. The first implementation of this approach is done by Lesk [1986]. All the sense of the word to be disambiguated retrieved from the dictionary. Each sense is then compared to the dictionary definition of remaining word in context. The sense which meets the context word is chosen as sense. For example: for selection of sense of pine in context of pine cone given the following definitions of pine and cone.

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

Lesk's method would select cone3 as the correct sense because two of its entry evergreen, tree overlaps with words in pine definition.

Problems associated with this technique are:

- Dictionary is limited for sense of target word.
- It does not have sufficient material to create classifier.
- The solution to problem is one can expand lists of words used in classifier.

6. CONCLUSION

In this paper we have surveyed the area of Word Sense Disambiguation. We have discussed various WSD methods that are useful in disambiguation the senses of words. In future we will implement algorithm that is based on supervised approach for refinement of result in the area of Information Retrieval.

REFERENCES

[1] "Ambiguity - SFSU". Available at online.sfsu.edu/~kbach/ambguity.html.

[2] "Ambiguity - Wikipedia, the free encyclopedia". Available at http://en.wikipedia.org/wiki/Ambiguity.

[3] "Human Communication 1 Lecture 13". Available at http://www.inf.ed.ac.uk/teaching/courses/hc1/slides1011/slides16.pdf.

[4] "Idiomatic Expressions". Available at http://idiomaticexpressions.org/frontpage

[5] "Word Sense Ambiguity - GM-RKB - Gabor Melli". Available at http://www.gabormelli.com/RKB/Word_Sense_Ambiguity.

[6] "Word Sense Disambiguation (WSD) Test Collection". Available at http://wsd.nlm.nih.gov/

[7] "Word-sense disambiguation - Wikipedia, the free encyclopedia". Available at http://en.wikipedia.org/wiki/Word-sense_disambiguation

[8] Daniel Jurafsky, James H. Martin. ***Speech and Language Processing,*** PEARSON, 2nd ed., 2009, ch. 17, pp. 657-690.

[9] Esha Palta. **Word Sense Disambiguation**, M. Tech. dissertation, dept. CSE Indian Institute of Technology, Mumbai, 2006.

[10] R. Navigli. **Word Sense Disambiguation: a Survey**, ACM Computing Surveys, Vol. 41, No. 2, pp. 1-69, Feb. 2009.

[11] Word sense disambiguation – Scholarpedia. Available at http://www.scholarpedia.org/article/Word_sense_disambiguation.

[12] Xiaohua Zhou and Hyoil Han. **Survey of Word Sense Disambiguation Approaches,** in *Proc. 18th International FLAIRS Conf.***,** Florida, May 15-17, 2005, pp. 1-6.

ABOUT THE AUTHORS

Rekha Jain completed her Master Degree in Computer Science from Kurukshetra University in 2004. Now she is working as Assistant Professor in Department of "Apaji Institute of Mathematics & Applied Computer Technology" at Banasthali University, Rajasthan and pursuing Ph.D. under the supervision of Prof. G. N. Purohit. Her current research interest includes Web Mining, Semantic Web and Data Mining. She has various National and International publications and conferences.

Sulochana Nathawat is pursuing her M.Tech degree in Computer Science and Engineering from Banasthali Vidyapith, Rajasthan. She received Master Degree in Computer Application from Apex Institute of Management & Science, Jaipur, Rajasthan in 2010. Her research interest includes Web Mining, Data Mining, Semantic Web, Information Retrieval and Natural Language Processing.

Prof. G. N. Purohit is a Professor in Department of Mathematics & Statistics at Banasthali University (Rajasthan). Before joining Banasthali University, he was Professor and Head of the Department of Mathematics, University of Rajasthan, Jaipur. He had been Chief-editor of a research journal and regular reviewer of many journals. His present interest is in O.R., Discrete Mathematics and Communication networks. He has published around 40 research papers in various journals.