A Review: Word Sense Disambiguation

1 Charushila Dadabhau Bhadane, Department of Computer Engineering SITRC, Nashik
2 Naresh Thoutam, Assistant Professor in Computer Engineering Department SITRC, Nashik
3 Mansi Narendra Mahajan, Department of Computer Engineering SITRC, Nashik
4 Vrushali Ramdas Suryawanshi, Department of Computer Engineering SITRC, Nashik

Abstract

Word Sense Disambiguation (WSD) is the task of removing ambiguity in different senses of words. It is a core research field in computational linguistics dealing with the automatic assignment of senses to words occurring in a given context. Humans are inherently good at WSD and distinguish senses used in words through spoken language. Computers on the other hand have difficulties identifying correct senses of words. Various advancements have been made in the task of disambiguation using mainly four approaches: Knowledge based, Supervised, Semi-Supervised, and Unsupervised. Better understanding of the human language will help computer's performance in various applications such as search engines, assistant software, and voice command interpretation. The objective of this work is to increase the accuracy of existing system to improve communication between computers and humans. Our hypothetical model uses supervised approach to create database using WordNet dictionary upon which we use a two-level algorithm. First level uses our improvised system, whereas second level uses Lesk algorithm together providing accurate sense detection

Keywords: ambiguity, supervised learning, unsupervised learning, Semi-supervised learning

---

1. Introduction

Word Sense Disambiguation is comes under the Artificial Intelligence that In Natural Language Processing (NLP), Word-Sense Disambiguation (WSD) is an crucial problem concerned with identifying which sense of a word is used in a sentence to get exact and accurate meaning of sentence. There are so many words used in the English language have various different senses and meanings. WSD is concerned with the problem of selecting the correct meaning with respect to sense. The solution to this problem impacts improving relevance of search engines with real time information.
The human brain is quite adroit at word-sense disambiguation. Simple context is all that is needed for humans to understand the correct sense or meaning of a word. The human language developed in a way that reflects the innate ability provided by the brain's biological neural networks. In computer science and Information technology, it has been a long-term challenge to train computers for natural language processing and predictions by machine learning or Artificial Intelligence. Presently so many application or technologies are there that they working on this problem of disambiguation but still there is scope for improvement and accuracy.
In today's modern world, people are heavily invested in a computer's ability to solve various problems in their daily lives. From finding directions via GPS to calculating their tax returns, most people are reliant on computer devices in one way or another. For better user experience and improved interfacing between man and machine, there needs to be clear communication between them. One obstruction in the way is the problem of ambiguity in word senses. In an effort to reduce this problem and enhance the intelligence of computers, we propose our system for Word Sense Disambiguation.
For example, consider a word bank in English which has different meanings: any of various as a commercial bank and other is Blood bank. Word sense disambiguation replaces the ambiguous word by the proper one depending on the surrounding context of the sentence.

2. Literature Survey

A rich variety of techniques have been studied, from knowledge-based methods that use the data encoded in lexical resources such as MRDsto supervised machine learning methods in which a classifier is trained for each word on a corpus of manually solved examples, to completely unsupervised methods that cluster occurrences of words, thereby inducing word senses.

B] There are four conventional approaches to WSD:

1. **Knowledge Based**

Knowledge based algorithms use various lexical resources such as Machine Readable Dictionaries (MRDs), WordNet to identify the correct sense of words.

These Algorithms are easy to implement and were the first to be developed while trying to solve the problem of WSD. A knowledge based system only needs access to commercial dictionary resources to start process of disambiguation.

Drawback of these algorithms is that their performance is limited on the speed of searching and retrieval of these resources. As the size of the resources increase, so does the latency and hence performance decreases.

1.1 **Lesk Algorithm**

The Lesk algorithm is the most influential dictionary-based method. It is based on the hypothesis that words used together in a sentence are related to each other and that the relation can be observed in the definitions of the words and their meanings. Two or more words are disambiguated by finding the pair of senses with the greatest word overlap in their dictionary definitions.

For example, when differentiating the words in "pine cone", the definitions of the appropriate senses both include the words evergreen and tree.

2. **Machine Learning Based**

1.1. **Supervised methods**

Supervised methods are based on the hypothesis that the context can provide enough indication on its own to disambiguate words (hence, common sense and reasoning are deemed unnecessary).

A learning set is prepared for the system to predict the actual meaning of an ambiguous word using a few sentences, having a specific meaning for that particular word. A system finds the actual sense of an ambiguous word for a particular context based on that defined learning set.

Supervised approach always gives superior performance than any other methods.However, these supervised methods are subject to a new knowledge acquisition holdup since they rely on considerable amounts of manually sense-tagged resources for training, which are arduous and expensive to create.

2.1.1 **Naïve Bayes Method**

Naïve Bayes Method is a supervised method. This method makes use of probabilistic approach which is one of the statistical methods, used to estimate probabilistic parameters. This probabilistic method usually expresses joint probability distribution or conditional probabilities in a given context and categories. Naïve Bayes algorithm uses classifiers which are mainly based on Bayes theorems to calculate the conditional probability for each sense (say k) of a word for which the features are defined (x1, x2, …, xm). Let P(k) and P(xi/k) are the probabilistic parameters of the model and they can be projected from the training data, using relative frequency counts as,

$$\arg\max_{k} P(k \mid x_1,\ldots,x_m) = \arg\max_{k} \frac{P(x_1,\ldots,x_m \mid k)P(k)}{P(x_1,\ldots,x_m)}$$

$$= \arg\max_{k} \; P(k)\prod_{i=1}^{m} P(x_i \mid k) .$$

Figure 2.1.1: Naive Bayes Formula [7]

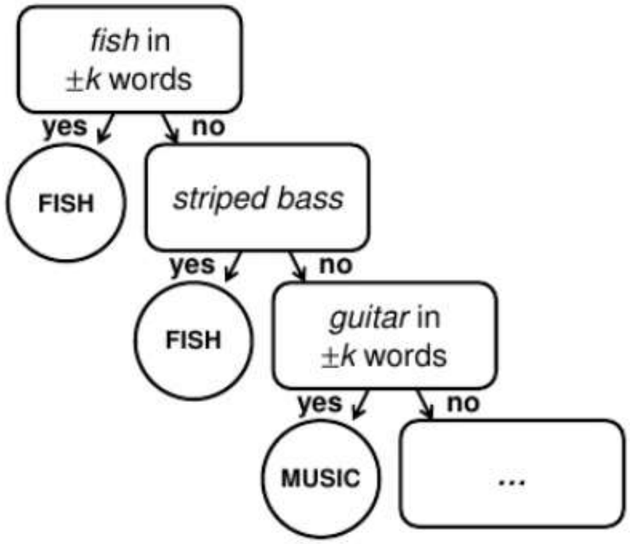

Figure 2.1.2: Decision Tree Example

2.2 Semi-supervised methods

Many word sense disambiguation algorithms use semi-supervised learning which allows both labelled and unlabelled data because of the lack of training data. The bootstrapping method starts from a small amount of seed data for each word: either a small number of sure fire decision rules (e.g., 'play' in the context of 'bass' almost always states the musical instrument) or manually tagged training corpus. Using any of the supervised method, seeds are used to train an initial classifier. This classifier is then used on the untagged portion of the corpus to extract a larger training set, in which only the most assured classifications are included. This procedure repeats, each new classifier being trained on a successively larger training data, until the complete data is consumed, or until a given maximum number of iterations are reached.

2.2.1. Yarowsky Bootstrapping Method

One of the most effective uses of the bootstrapping approach in Natural Language Processing (NLP) is made by the Yarowsky in 1995. The Yarowsky method is incremental and one of simple iterative algorithm which does not requires large training sets and depends only on relatively small number of instances of each sense. As semi-supervised method uses labelled instances, these labelled instances are then used as raw information to train the classifier initially using other supervised methods. The trained initial classifiers are then used to extract a larger training set from the remaining untagged corpus. The trained sets which are obtained above the particular threshold are kept for future to train the other untrained sets of data for next iteration.

2.3. Unsupervised methods

Unsupervised learning methods are the greatest challenge for WSD researchers. The underlying assumption is that similar meaning words occur in similar contexts, and thus senses can be induced from text by clustering word occurrences using some measure of similarity of context, a task referred to as word sense discrimination or induction. To disambiguate a word they use some measure of similarity in context to get the correct sense.

Performance has been observed to be lower than for the other methods described above, but it is hoped that unsupervised learning will overcome the knowledge acquisition bottleneck because they are not dependent on manual effort.

2.3.1. Co-occurrence Graphs/ Hyperlex Algorithm

Where the previous techniques use vectors to represent the words, the algorithms in this domain make use of graphs. Every word in the input text becomes a vertex and syntactic relations become edges between respective vertices. The context units (e.g. paragraph, sentence) in which the target words occur, are used to create the graphs. The edge weights are inversely proportional to the frequency of co-occurrence of these target words.

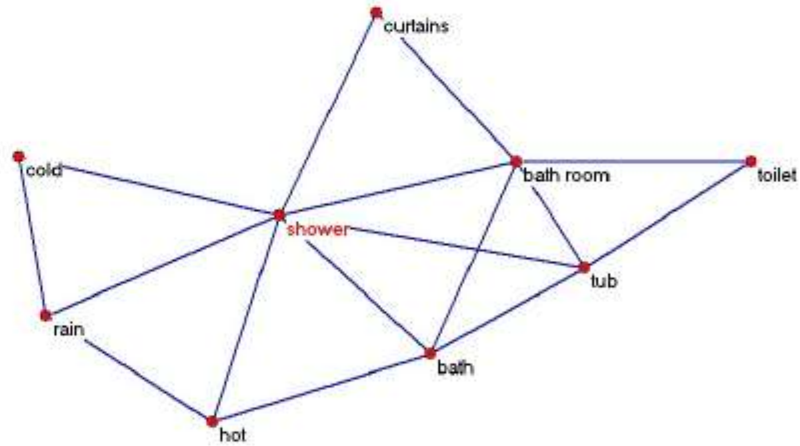

Figure 2.3.1: Co-occurrence Graph Example

Summary of Survey:

Table 2.1: Literature Survey Summary

Year	**Author**	**Title**	**Method**	**Advantages**	**Limitations**
2019	Chandrakant D. Kokane et.al.	Supervised Word Sense Disambiguation with Recurrent Neural Network Model	Supervised RNN	Superior performance for polysemy words	Training is complex
2020	Chandrakant D. Kokane et.al.	An Adaptive algorithm for lexical ambiguity in WSD	DNN	High accuracy for lexical ambiguity than RNN	Creating manual sense instances for training is complicated
2018	Myung Yun Kang, Tae Hong Min, Jae Sung Lee	Sense Space for Word Sense Disambiguation	Word Space Model	Sense Space Model more effective than Word space model	Not practical because sense context is not available in normal queries.
2017	Bartosz Broda, MaciejPiasecki	Semi-Supervised Word Sense Disambiguation Based on Weakly Controlled Sense Induction	LexCSD algorithm ,Clustering algorithm, NB algorithm	Reduced human involvement	Limited to unsupervised methods which has a lower accuracy and produce results which are not satisfying for many applications
2017	Pratibha Rani, VikramPudi, Dipti M. Sharma	Semi-supervised Data-Driven Word Sense Disambiguation for Resource-poor Languages	Contextual similarity property based on hypothesis of Yarowsky (1993)	Generic Method issuitable for resource-poor languages and it can be used for various languages without requiring a largesensetaggedcorpus.	Provides comparatively low accuracy for English language as compared to other systems.
2016	Ignacio Iacobacci, Mohammad	Embedding for Word Sense Disambiguation: An	Knowledge-based word	Word Embedding can be used to	Bestperformance is obtained when

	TaherPilehvar, Roberto Navigli	Evaluation Study	embedding	improve state-of-the-art Supervised WSD.	standard WSD features are augmented with the additional knowledge from Word2vecvectors.
2015	Udaya Raj Dhungana, SubarnaShakya, KabitaBaral and Bharat Sharma	Word Sense Disambiguation using WSD Specific WordNet of Polysemy Words	Adapted Lesk algorithm	Accuracy is increased by 3.484% in compared to the accuracy of previous systems of Nepali Language.	Requires large amount of training data.
2013	Alok Pal,AnupamMunshiand DigantaSaha	An Approach To Speed-up the Word Sense Disambiguation Procedure Through Sense Filtering	Filtering Method using unsupervised method	Correct sense of an ambiguous word is found using the Part-of-Speech Tagging before the disambiguation procedure.	Execution time might be increased
2013	LokeshNandanwar, KalyaniMamulkar	Supervised, Semi supervised, Unsupervised WSD Approaches	Survey Paper which summarizes approaches regarding WSD	Provides useful information regarding basics of WSD	Does not go into details about any of the approaches it mentions
2009	Niladri Chatterjee and RohitMisra	Word-Sense Disambiguation using Maximum Entropy Model	Principle of Maximum Entropy	High accuracy of around 85% through the language independent model.	Same concept applied to Supervised methods might reward higher accuracy.
1986	Michael Lesk	Automatic Sense Disambiguation using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone	Supervised Method, Lesk Algorithm	Laid the foundation of Supervised Learning, providing 50-70% accuracy	Accuracy is limited to technology of lexical resources.

3. Conclusion and Future Work

Word Sense Disambiguation (WSD) is an open problem concerned with selecting the correct sense of a polysemous word. To bridge the gap between humans and computers and to provide better interfacing, the accuracy of the systems working on this task, need to be improved. In this report, we have addressed the need to increase accuracy of existing WSD systems, and to put forward our hypothetical model based on supervised approach. The model uses a two level algorithm that uses our improvised system at the first level and Lesk algorithm at second level to maximize accuracy.

As future work, proposed system can be developed. Carefully planned execution and programming of the system will provide required high accuracy. Then after development, test cases will be generated and accuracy of system can be tested successfully

REFERENCES

[1] Kokane C. D., Babar S.D. “Supervised Word Sense Disambiguation with Recurrent Neural Network Model,” International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-2, December-2019.

[2] Kokane C. D., Babar S.D., Mahalle P.N.“An Adaptive Algorithm for Lexical Ambiguity in Word Sense Disambiguation,” Lecture notes in Network and systems May-2020.

[3] Myung Yun Kang, Tae Hong Min, Jae Sung Lee. "Sense Space for Word Sense Disambiguation". 2018 IEEE International Conference on Big Data and Smart Computing

[4] Bartosz Broda and Maciej Piasecki. "Semi-Supervised Word Sense Disambiguation Based on Weakly Controlled Sense Induction". 2009 Proceedings of the International Multiconference on Computer Science and Information Technology

[5] Pratibha Rani, Vikram Pudi, Dipti M. Sharma "Semi-supervised Data-Driven Word Sense Disambiguation for Resourcepoor Languages" 2017 14th International Conference on Natural Language Processing (ICON)

[6] Ignacio Iacobacci, Mohammad Taher Pilehvar, Roberto Navigli "Embedding for Word Sense Disambiguation: An Evaluation Study" 2016 Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, pages 897-907

[7] Udaya Raj Dhungana, Subarna Shakya, Kabita Baral and Bharat Sharma "Word Sense Disambiguation using WSD Specic WordNet of Polysemy Words" 2015 Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing

[8] Alok Pal, Anupam Munshi and Diganta Saha "An Approach To Speed-up the Word Sense Disambiguation Procedure Through Sense Filtering" 2013 International Journal of Instrumentation and Control Systems (IJICS) Vol.3, No.4, October 2013 26

[9] Lokesh Nandanwar and Kalyani Mamulkar "Supervised, Semi-Supervised and Unsupervised WSD Approaches: An Overview" 2013 International Journal of Science and Research (IJSR)

[10] Niladri Chatterjee and Rohit Misra// Word-Sense Disambiguation using Maximum Entropy Model// International Conference on Methods and Models in Computer Science, 2009

[11] Michael Lesk// Automatic Sense Disambiguation using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone// Bell communications Research Morristown,NJ 07960

[12]George A. Miller WordNet: A Lexical Database for English. Communications of the ACM Vol. 38, No. 11: 39-41 Bell communications Research Morristown,NJ 07960 https://wordnet.princeton.edu/