3rd National Conference on Green Technology and Science for Sustainable Development
© 2020 IJSRST | Volume 5 | Issue 6 | Print ISSN: 2395-6011 | Online ISSN: 2395-602X
International Journal of Scientific Research in Science and Technology

Word Sense Disambiguation – Supervised Approaches: Present Scenario

Aparitosh Gahankari[1], Dr. Avanish S.Kapse[2], Dr. V. M. Thakre[3]

[1]Research Scholar, Department of Computer Science and Engineering, Nagpur Institute of Technology , Nagpur (MS), India

[2]Head of Department, Department of Computer Science and Engineering, Anuradha College of Engineering, Chikhli (MS),India

[3]Professor & Head, P. G .Department of Computer Science & Engineering, SGBAU, Amravati, MS, India

ABSTRACT

This paper covers the discussion of how a meaningful sense of the given word can be selected in the given context. In the domain of Natural Language Processing (NLP), Word Sense Disambiguation (WSD) is still an open problem. The use of WSD can be there in many fields including but not limited to Machine Translation, Text Pre Processing, Information Retrival, etc. To deal with Problem of correctly identification of the sense different approaches are used specifically in the Machine Learning. This paper is a kind of Survey wherein we will present the current scenario of the different Machine Learning (ML) Algorithms & the techniques used with the particular data set the said algorithm is applied on. This paper should be helpful to those who are novice in the NLP, specifically from WSD domain point of view. This survey concludes that there are some ML algorithms which works efficiently on some data sets while the others works best on data set from different languages.

Keywords : Word Sense Disambiguation (WSD), WordNet, Natural Language Processing (NLP)

I. INTRODUCTION

The Broad area of Computational Linguistics which includes many fields, has a sub domain as WSD, wherein the systems are designed to determine the closet contextual meaning of the given word. WSD is the process of filtering out the unrelated sense of the word in the given term in current context and assigning the correct sense as there are many words which can have many meanings based on the context they are used. As there are three types of words disambiguation removal techniques namely Knowledge based techniques, Machine Learning based techniques and combination of above both, better known as Hybrid techniques[1], I will limit this discussion to Supervised Machine Learning based techniques and the particular algorithms in those techniques.

For example in the plain English the word "Bank" can have different meaning depending upon the context in which the word is utilized viz. Meaning related to Finance ,River-side or reservoir.

Following table shows some major milestones achieved historically.

Table 1 :- Major milestones achieved historically

Sr.No	Year	Name of Researcher	Milestone Achieved
1.	1949	Zipf	Law of Meaning published
2.	1950	Kaplan	Concluded that Either side of the ambiguous word plays important role while understanding the correct meaning of the word in the given context.
3.	1957	Masterman	Technique to choose the heading which matches most correct sense.
4.	1975	Willks	"Preference Semantics" developed by Willks
5.	1979	Rieger & Small	The idea of "Word Expert" was developed
6.	1980	-	Large scale Lexical Resources developed
7.	1990	Willks et.al.	Different Procedures for Knowledge extraction were developed. Online Dictionary WordNet was available.
8.	1991	-	Longman Dictionary of Contemporary English (LDOCE) was used with Lesk's Algo.
9.	1992	Yarowsky	Disambiguation Rules Framed
10.	1997	Resnik & Yarowsky	SENSEVAL Discussed
11.	2000	Kilgarriff and Palmer	Corpora Discussed

APPLICATIONS OF WORD SENSE DISAMBIGUATION

Following are the major application areas where the concept of WSD is mainly used in any kind of the Linguistic Research [2]

- Machine translation (MT):

The word carries different meaning depending upon the context in which it is used, as can be seen through the example given in the Section I of this paper..

- Information retrieval (IR):

If the Question has a Ambiguous word then if no proper meaning is found it's very difficult to find the correct answer.

- Information extraction (IE):

Plays a major role in different research works are concerned as far as Information Extraction is concerned.

II. Related Work

The topic of WSD has been addressed by bulk of literary work but still it is a open issue in the world of NLP. Due to overwhelmed number of uses of the WSD concept it is expected that many more researcher will try to contribute in this field of WSD. The research in the field of WSD is also important from socio-economic impact as WSD can be directly be utilized to provide more and more life enriching experiences. There are very few papers are available wherein the exhaustive summary is written which is represented from the domain of WSD. This section shows the related work in this research field. biLM (bidirectional Language Model) representations can be used to deal with WSD[3]

In[4] the software is designed called SupWSD, which is based on (i) input parsing, (ii) text pre-processing,

(iii) Features extraction and (iv) classification is developed The experimental evaluation showed that, in addition to its flexibility, SUPWSD can replicate or outperform the state-of-the-art results reported by the best supervised models on standard benchmarks, while at the same time being optimized in terms of execution time.

In [5] It has been observed that minor differences like content category, choice of tag set, size of training sets can have an impact on accuracy of the POS taggers. In[6] the authors shows that sequence learning approaches learn a single model in one pass from the training data, and then disambiguate jointly all target words within an input text, as per observations by the authors of [6] these models are sufficiently flexible to allow them, for the first time in WSD, to be readily adapted to languages different from the one used at training time, and still achieve competitive results.

Paper [7] uses Extended WSD Incorporating Sense Embeddings (EWISE) as targets instead of discrete sense labels. This helps the model gain zero-shot learning capabilities, demonstrated through ablation and detailed analysis.

The authors of [8] implemented a state of the art WSD neural network and we showed that these methods compress the size of the underlying models by a factor of 1.2 to 2, and greatly improve their coverage on the evaluation tasks. As a result, they reach coverage of 99.99% of the evaluation tasks.

In [9] a unified evaluation framework for all-words WSD is presented. This framework is based on evaluation datasets taken from Senseval and

SemEval competitions, as well as manually and automatically sense-annotated corpora.

In [10] authors seek to address the problem of integrating the glosses knowledge of the ambiguous word into a neural network for WSD

Following table shows some of the papers based on WSD.

Table 2 :- Summary of some Papers

Sr. No	Paper title	Authors	Algorithms used	Datasets
1	Knowledge-basedWord Sense Disambiguation using Topic Models	Devendra Singh Chaplot, Ruslan Salakhutdinov	probabilistic graphical model	SensEval-2 (Palmer et al. 2001), SensEval-3 (Snyder and Palmer 2004), SemEval-2007 (Pradhan et al. 2007), SemEval-2013 (Navigli, Jurgens, and Vannella 2013) and SemEval-2015
2	Word Sense Disambiguation for Urdu Text by Machine Learning	Syed Zulqarnain Arif1, Muhammad Mateen Yaqoob1, Atif Rehman2, and Fuzel Jamil1	SVM	k-folded cross validation
3	Design and Development of a Knowledge-Based Approach for Word Sense Disambiguation by using WordNet for Hindi	Pooja Sharma, Nisheeth Joshi	LESK Algorithm	Hindi WordNet.
4	Incorporating Glosses into NeuralWord Sense Disambiguation	Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang and Zhifang Sui	GAS: a gloss-augmented WSD neural network	Senseval-2 (SE2), Senseval-3 task 1 (SE3), SemEval-07 task 17 (SE7), SemEval-13 task 12 (SE13), and SemEval-15 task 13 (SE15).
5	SUPWSD: A Flexible Toolkit for SupervisedWord Sense Disambiguation	Simone Papandrea, Alessandro Raganato and Claudio Delli Bovi	supervised Word Sense Disambiguation (WSD).	Framework of Raganato et al. (2017)8, which includes five test sets from the Senseval/Semeval series and two training corpus of different size, i.e. Sem-Cor (Miller et al., 1993) and OMSTI (Taghipour and Ng, 2015a). As sense inventory, we used WordNet 3.0 (Miller et al., 1990) for all open-class parts of speech.
6	Entity Linking meets Word Sense Disambiguation: a Unified Approach	Andrea Moro, Alessandro Raganato, Roberto Navigli	Babelfy, a unified graph-based approach to EL and WSD based on a loose identification of candidate meanings coupled with a densest subgraph heuristic which selects high-coherence semantic interpretations.	SemEval-2013 task 12 dataset for multilingual WSD, SemEval-2007 task 7 dataset for coarsegrained English all-words WSD, SemEval-2007 task 17 dataset for finegrained English all-words WSD, Senseval-3 dataset for English all-words WSD, KORE50 (Hoffart et al., 2012), which consists of 50 short English sentences, AIDA-CoNLL6 (Hoffart et al., 2011), which consists of 1392 English articles
7	Neural Sequence Learning Models for Word Sense Disambiguation	Alessandro Raganato, Claudio Delli Bovi and Roberto Navigli	Sequence Learning for Word Sense Disambiguation, bidirectional LSTM architecture	five standardized test sets from the Senseval/SemEval series: Senseval-2 (Edmonds and Cotton, 2001, SE2), Senseval-3 (Snyder and Palmer, 2004, SE3), SemEval-2007 (Pradhan et al., 2007, SE07), SemEval-2013 (Navigli et al., 2013, SE13) and SemEval-2015 (Moro and Navigli, 2015, SE15
8	RandomWalks for Knowledge-Based Word Sense Disambiguation	Eneko Agirre, Oier L´opez de Lacalle	knowledge-based Word Sense Disambiguation based on random walks over relations in a LKBThe PageRank random walk algorithm (Brin and Page 1998)	variety of English data sets and a data set on Spanish

III. Approaches for WSD

There are broadly three main categories for different approaches used in WSD,They can be categories as

a. Supervised
b. Unsupervised
c. Knowledge-based.[11]

This paper will discuss the supervised approaches only.

A. Supervised Approaches

When enough is known regarding the data the Supervised Learning method can be used as we had some knowledge regarding the known outcomes. In other words when we had the Labeled sample data (tagged with identifying information) & the correct output then the Supervised Learning method can be used effectively. The sense-annoted data is used for the training purposes, so the data which is used for creation of vector of features will be problem hence can be proved to be major bottleneck. Comprise of word which is focused & the classifier that new occurrences of the word in focused [12].

Broadly the approaches can be subdivided into two subdivisions.

Step 1: The word in focus, which creates ambiguity, will further be broken down into the vector of features.

Step 2: Now apply any Supervised Learning algorithm on the Step 1 created vector of features.

The major limitation of this approach is that if any new word which needed to be targeted has no related learning data then creation.Following are some of the different approaches which can be used to classify as supervised approaches to deal with problem with WSD.

1. **AdaBoost**:-This method uses set of weak classifier which is all linear set available in nature. The incorrectly classified. The classification steps are repeated to effectively increase the weights assigned so that as we move on the classification we could focus on the different classification [11].Application of Adaboost can be found in paper presented by [11].The main feature of this method is that it corrects the weights assigned in each iteration, so that the classifier can focus on the incorrect classified examples. From limitations point of view this approach is that it took huge amount of time if the input terms are large.
2. **Support Vector Machines (SVM):-**The main concept behind Support Vector Machines is learning using a set of training data by using the hyperplane. By using this hyperplane SVM is able to distinguish between the more positive & the results which are towards negative side from the chosen set of examples .The biggest optimization achieved through the use of SVM is separation in the training examples , the said concept is applied using the dataset SENSEVAL-3.The review of the final results from the SVM revels that SVM has reasonable accuracy. The regularization parameter is plying a lead role in case of performance of SVM, which can further be developed to build many applications of AI.
3. **Decision Trees:** This is a well utilised technique for the problem solving related to classification. This is Boolean value tree using the concept of Binary trees. The Gain feature decides the length of the tree from root to leaf . The approaches sues the SENSEVAL -1 & SENSEVAL -2 dataset on English Language.[13]concluded that this technique has very low accuracy. Decision Trees can be applied in case of data maintenance. Minute change in the input parameter can play a big role in changing the data results.
4. **Neural Networks:-** It is a Supervised method of Artificial Neurons. The genetic approach can be used to provide data processing Local & Global context is used to calculate the score in case of NN. The dataset used is a Wordsim353 on English language shows Multi-prototype NN model performs better than other methods.
5. **Memory based Learning:-** The current model acts as a depository of newer & newer models as they are getting added one by one on the same model. The K-NN can be used for the WSD as the distance to measure proximity if K is greater than is used. The long run time can be a bottleneck in the performance.
6. **Naive Bayes:-** Naive Bays , the simplified probabilistic learning algorithm works well to classify the words whose sense is ambiguous. SENSEVAL -1 & SENSEVAL -2 datasets are used. The said implementation is on Arebic Language. The number of iterations which needs to be implemented can be a major Bottleneck of this algorithm.

IV. Conclusion

This paper has summarized the Supervised machine Learning approaches to solve the WSD issue which is still a open research area in NLP domain. The different algorithms & their summary has been quoted here.

Techniques	Methods	Language
Supervised	Decision Lists	General
Supervised	Examples	English
Knowledge Based	Distance Between Semantic Simility	English
Supervised	Naïve Bayes	English
Supervised	Svm	English
Unsupervised	Co-Occurence	English
Knowledge Based	Lesk Algo	Bengali
Unsupervised	Clustering	English
Supervised	Dt Algo	English
Knowledge Based	Knowledge Types	English

V. REFERENCES

[1]. Knowledge-based Word Sense Disambiguation using Topic Models by Devendra Singh Chaplot, Ruslan Salakhutdinov, arXiv:1801.01900v1 [cs.CL] 5 Jan 2018

[2]. Design and Development of a Knowledge-Based Approach for Word Sense Disambiguation by using WordNet for Hindi by Pooja Sharma, Nisheeth Joshi, International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-8 Issue-3, January 2019

[3]. Deep contextualized word representations, Matthew E. Peters , Mark Neumann , Mohit

Iyyer , Matt Gardner, Proceedings of NAACL-HLT 2018, pages 2227–2237, New Orleans, Louisiana, June 1 - 6, 2018. c 2018 Association for Computational Linguistics

[4]. SUPWSD: A Flexible Toolkit for SupervisedWord Sense Disambiguation by Simone Papandrea, Alessandro Raganato and Claudio Delli Bovi, Proceedings of the 2017 EMNLP System Demonstrations, pages 103–108 Copenhagen, Denmark, September 7–11, 2017. c 2017 Association for Computational Linguistics

[5]. Research on Machine Learning Techniques for POS Tagging in NLP Aparna Bulusu, Sucharita V, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019

[6]. Neural Sequence Learning Models for Word Sense Disambiguation Alessandro Raganato, Claudio Delli Bovi and Roberto Navigli , Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pages 1156–1167 Copenhagen, Denmark, September 7–11, 2017. c 2017 Association for Computational Linguistics

[7]. Zero-shotWord Sense Disambiguation using Sense Definition Embeddings Sawan Kumar1 Sharmistha Jat1 Karan Saxena2;_ Partha Talukdar Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5670–5681 Florence, Italy, July 28 - August 2, 2019. c 2019 Association for Computational Linguistics

[8]. Sense Vocabulary Compression through the Semantic Knowledge of WordNet for NeuralWord Sense Disambiguation, Loïc Vial Benjamin Lecouteux Didier Schwab, dearXiv: 1905.05677v3 [cs.CL] 27 Aug 2019

[9]. Word Sense Disambiguation: A Unified Evaluation Framework and Empirical Comparison Alessandro Raganato, Jose Camacho-Collados and Roberto Navigli, Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pages 99–110, Valencia, Spain, April 3-7, 2017. c 2017 Association for Computational Linguistics

[10]. Incorporating Glosses into NeuralWord Sense Disambiguation By Fuli Luo, Tianyu Liu, Qiaolin Xia, Baobao Chang and Zhifang Sui, Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Long Papers), pages 2473–2482, Melbourne, Australia, July 15 - 20, 2018. c 2018 Association for Computational Linguistics

[11]. Word Sense Disambiguation:Survey Study,Ahmed Aliwayand Hawraa Taher,Journal of Computer Science pages 1004-1011,22/07/2019

[12]. Bridging the Gap between Training and Inference for Neural Machine Translation by Wen Zhang, Yang Feng, Fandong Meng, Di You, Qun Liu, arXiv:1906.02448v2 [cs.CL] 17 Jun 2019

[13]. A Simple Theoretical Model of Importance for Summarization Maxime Peyrard_ Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 1059–1073 Florence, Italy, July 28 - August 2, 2019. c 2019 Association for Computational Linguistics