International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

An Empirical Analysis of Word Sense Disambiguation through Machine Learning Approaches

Vaishnavi Wasankar[1], Yashika S. Nimje[2], Kajal Pardhi[3], Divyani C. Shende[4], Aparitosh Gahankari[5]
[1234]BE Scholar, Department of Computer Science engineering, Nagpur Institute of Technology, Nagpur, Maharashtra, India
[5]Assistant Professor, Department of Computer Science engineering, Nagpur Institute of Technology, Nagpur, Maharashtra, India

Article Info
Volume 8, Issue 2
Page Number : 104-114

Publication Issue :
March-April-2022

Article History
Accepted: 10 March 2022
Published: 25 March 2022

ABSTRACT

The procedure to identify the appropriate meaning for the particular word in an ambiguous statement is considered as Word Sense Disambiguation. It is a complicated problem since it necessitates the utilization of information from a variety of sources. Since the start of machine learning, a significant amount of time and effort has been devoted to overcoming this challenge, and the work is currently ongoing. In WSD, a variety of methodologies were employed and executed on a variety of corpora representing practically all languages. WSD algorithms are grouped into three groups in this paper: Supervised algorithms, unsupervised algorithms and knowledge-based algorithms. Every subcategory will be examined thoroughly, with details elaborated for nearly all of the algorithms within each area. As a result, work samples for every technique were selected based on the language being used, the corpora being used, and other considerations. Each method's advantages and disadvantages were meticulously documented. Some of these strategies have limits in certain scenarios, and our work will assist scientists in the fields of machine learning in selecting the most appropriate algorithms to tackle their specific problem in WSD. When comparing the works that have been used and indeed the procedures that were employed, it is possible to notice the distinctiveness of the piece of work that was created. As a result of this research, it was observed that (i) size of the dataset has an considerable impact on algorithm's performance, (ii) some methodologies provide high performance accuracy for one language where as it gives low performance for some other, (iii) a few of these methodologies can be run quickly but with a limitation on accuracy, and (iv) the large number among those methodologies have been implemented successfully for a wide range of different languages.

Keywords: Classification of WSD Technique, Word Sense Disambiguation, Applications of WSD

I. INTRODUCTION

There are many terms throughout all primary international languages which relate to diverse meanings in different settings, but there are many of them. Such multi-sensory phrases are referred to as "ambiguous words," and the method of evaluating the proper significance of a puzzling word from its context is referred to as "Word Sense Disambiguation." The task of picking the appropriate word sense from a specified collection of word senses in accordance with the context can be termed as Word Sense Disambiguation (WSD).

WSD has a lengthy history, having been restored in the late 1960s as component of Machine Intelligence Investigation in order to better understand the complete variety of natural language expressions. In the 1980s, a significant step forward was made in the clarification of the definitions of words. Since that time, academics have been experimenting with different algorithms in order to complete this goal.

WSD is a critical task and an open problem in the field of Natural Language Processing (NLP), and it deserves attention. There are several real-world applications that require WSD, such as “Semantic Mapping (SM), Machine Translation (MT), Ontology Learning (OL), and Semantic Annotation (SA)." WSD is also required for many academic research projects. It can also be beneficial in increasing the performance of a variety of applications, including speech recognition (SR), information extraction (IE), and information retrieval (IR) [1] among many others. For example, the English term "bank" has several multiple meanings, including financial institution, riverside, and reservoir, among others. Any phrase using this phrase cannot be translated into another language without first understanding the correct connotation of the word in the sentence. Given how difficult it is to determine the precise meaning of this word in the context of a sentence, several algorithms and strategies have been developed to assist in accomplishing this essential issue [2].

There are a variety of applications that demand WSD in order to aid with in comprehension of text's constituent parts. Some of these will be discussed in further detail later in this section:

- Machine Translation (MT): It makes use of WSD to resolve ambiguities in word meaning in a phrase in order to provide an accurate translation. For example, the English words (He scored a goal) and (It truly was his life's desire) cannot be translated correctly unless the correct connotation has already been deduced from either the term (goal), which seems to have distinct implications in these lines and so cannot be translated accurately [2].
- Extraction of Information: WSD is used in Internet Explorer and text mining to ensure that text is accurately analysed. As a rule, a semantics analysis is extremely beneficial in IE because perceptions and analogues play such a significant role in the language.
- Content Analysis (CA): WSD is an important point in qualitative research since it may assist in categorizing data in accordance with user needs and solving a variety of difficulties [3].
- Retrieval of Information: Information Retrieval is one of the most popular application of WSD in the real world. In this, the cluster of data having semantically similar with user specific query is obtained through WSD. WSD also contributes to the improvement of the precision in the infrared imaging systems [4].

II. RELATED WORKS

The field of Word Sense Disambiguation has seen a lot of progress; some of the approaches are accurate enough for diverse languages, but there has been little work done to examine the various approaches. The works that are associated with this section will be shown. According to [1] a summary of the various sources of knowledge used in WSD was provided, as well as a classification of the contemporary WSD algorithms are based on their approaches. They also reviewed the reasoning, tasks, efficiency, information resources necessary, computation time, assumptions, and highly value for every class of WSD algorithms, as well as the advantages and disadvantages of the each class of WSD methodologies. In [2], the author published a review of several research approaches and the technical scenario in existence during this whole domain; these studies were focused at numerous Languages of India, and finally the author presented a study in Bengali. In [4], the authors explored the task of WSD, as well as the various techniques and algorithms that have been developed. They additionally discussed several natural language processing (NLP) applications that will be useful when incorporating a deformation system. It also describes the evaluation metrics that are utilised to calculate WSD for performance. A survey on WSD was published in [5], and the results were used to assist researchers and users in selecting the algorithms and approaches that will best address their problems and work in their individual applications. The purpose of this work to offer users with a general understanding of the selection of WSD algorithms for usage in certain scenarios or to solve specific challenges.

WSD Approaches

There are three main categories to classify Word Sense Disambiguation Approaches namely, unsupervised, supervised and knowledge-based.

Knowledge-Based WSD

When compared to machine learning, knowledge-based approaches are less complicated. The use of external lexical resources, such as dictionaries, in conjunction with Machine Learning methods where extra training corpora are required instead of knowledge base (dictionaries, thesaurus, wordnet, etc.) [5]. Knowledge-based WSD involves a large number of methods, the most extensively used of which being the LESK algorithm for solving WSD issues. Advantage: When dealing with complicated phenomena in languages, knowledge-based WSD methodologies can indeed be employed to solve the problems. It also provides WSD practitioners with practical materials derived from a variety of concepts. The disadvantages of this method are that it necessitates the efforts of linguistic experts, that it does not accurately reflect the occurrences in pragmatic text (corpora), and that it is unable to meet the expected overall performance for certain methods.

LESK Algorithm

Using definition overlap, the LESK algorithm determines the right senses for all of the terms in the context all at the same time. It was discussed and implemented by a large number of researchers for a variety of languages. Lesk proposed an approach in which the interpretations of the each phrase were checked with both the expressions that appear in the sentence in question. The original Lesk method "analyses overlapping between sense descriptions for any and all words in the text and identifies concurrently the proper senses for all terms in the text," according to the author's description [4][5]. The author in [6] extended this technique to the English language, utilising WordNet, Wikipedia, and BablelNet as data sources. In [7] the authors used the Bengali WordNet to apply this technique to the Bengali language, which did not have a synset structure in the Bengali Word vectors at the time. The authors in [8] extended this technique to the Arabic language, employing the Elwatan data corpus as a source of information. The Simplified Lesk technique is faster than its predecessor Lesk algorithm due to its reduced computationally expensive complexity. The improved Lesk method also distinguishes between word meanings significantly more accurately. The downsides are that it requires extensive knowledge and that the basic technique is impractical.

Semantic Similarity

And one of the Knowledge-Based approaches that has been utilised to solve WSD. Moreover, there is substantially less space between text segments that include two or more words. In part, this concentration on word-to-word analogies might be attributed to the release of databases that describe connections among words or concepts, such as WordNet [9], which defines interconnections among words or concepts. Semantic similarity assessments can be used to explain ambiguity and examine patterns for coherence. All metrics fall into one of four categories: information included, features, average distance, and combination measures. [10]. According to the authors in [9], they applied this strategy to the English language using WordNet, and they ensured that IR applications would gain as much from the structural similarity measures when both the document and the query are brief. In [11] author describes how they used this methodology in association with Corpus-based Measures, (i) position wise similarity measurement, and (ii) semantic representations evaluation, all of which have been performed on the English language, as well as other techniques. This technique was adapted to the Local languages by the author, details how he did it by utilising the Bengali WordNet. The author of [12] used private data on the English language to demonstrate the effectiveness of this strategy. Benefit: The ability to produce harmony for the entire speech is referred to as semantic similarity. This is because the minimum difference between these two words suggests that they have been linguistically linked in terms of meaning. Arithmetic will be used extensively when enough than two words are taken into consideration arithmetic. The disadvantage is that the strength of semantic interaction of any related constructs with much the same separation distance seems to be the same uniformity separation problem as the previous uniform distance problem.

Selectional Preferences

A knowledge-based method that relies on the source of knowledge to function is one that obtains information about the potential interconnections between distinctive types of words while also making a reference to rational thought throughout the process. Many researchers have employed selectional preferences to address the WSD problem, which is another technique of solving the problem. Selectional preferences are described in terms of classifying objects but instead of simple words [2], rather than simple words itself. Selectional Preferences techniques express some constraints on the semantic type of the target word when picking it, and imposing actual significance of something like the word sequence, which would be collected through the morphological connection in the paragraph when selecting it. [13]

Heuristic Method

The Heuristic Approach is a type of knowledge-based method in which the linguistic features of a word serve as the central component of the method, which is used to determine the correct interpretation of an ambiguous word. Since human languages are so complex, and a single word can have several interpretations, WSD has really been adopted by a variety of sectors of interest, including search engines and automatic machine translation. Most recently, academics have concentrated on the application of meta-heuristic techniques to identify the optimal solutions that make the most logical sense. Meta-heuristic techniques, on the other hand, are still in their infancy in terms of applicability, necessitating the efficient study and development of the issue space. So the current study seeks to propose an improved hybrid meta-heuristic system that incorporates the advantages of optimization algorithm (PSO) and evolutionary programming to determine the global optimal interpretation of a given text. Different semantic measurements have been used as optimization algorithms for the hybrid PSO in this model, which is based on the semantic measurements. Specifically, these measures are comprised of JCN and expanded Lesk techniques, which are effectively coupled in this work. The effectiveness of this strategy is evaluated using three benchmark datasets (SemCor 3.0, SensEval-2, and SensEval-3).

Walker's Algorithm

Walker's algorithm is based on thesaurus-based search techniques. Beginning with the identification of the equivalents with which this interpretation corresponds, this algorithm goes on to determine the outcome among each sensitivity by matching words from either the word embedding's to the outcomes. If the meaning of the word's synonyms is equivalent towards the interpretation of this sense, this should add 1 towards the sense in which it appears. The advantage of this strategy is that it provides great accuracy because that relies on synonyms. Cons: It is tough to come up with synonyms that can assist in resolving the ambiguities inside the word problem at hand.

Supervised WSD

Supervised methods are a type of machine learning methodology that uses data that has been manually produced and labelled in the senses. The classifier will be trained on a training sets that contains examples of target words that are related to the target word. According to [3], the primary aim is to develop a classifier that appropriately and accurately classifies the new examples in accordance with the context in which they are used. According to the authors, "a supervised WSD technique consists of two parts: Specifically, (i) converting every learning instances of an unambiguously phrase into the feature space, and (ii) running a supervised learning algorithm after having encoded all training cases in the feature vector. [21]" This section will cover a number with some well supervised classification of WSD that are commonly used. Advantage: Almost all of the algorithms that are employed are not dependent on the language being used. The disadvantage of the supervised WSD algorithm is that it is limited in terms of learning data, and as a result, problems with uncertain work or uncertain sense for with this word will arise.

Decision List

It is a typical problem in computational linguistics and text processing technologies to perform word sense disambiguation (WSD) on words. That whenever a word has numerous meanings, WSD determines which one is correct in a statement or phrase by analyzing the context. WSD is often divided into two distinct phases. The first phase involves determining and extracting all of the possible senses for each word. Second, comparable terms are labelled with the proper sense to assist with disambiguation in the third phase. Many scholars who have worked on WSD have suggested that decision lists offer a straightforward method of solving ambiguity concerns [1].

Decision Lists were widely employed in a broad variety of applications and activities, with one of these tasks being the WSD. The decision list is a collection of rules (if-then-else) that are presented in an ordered list fashion. It is difficult to do word sense differentiation with single terms or short searches because of the lack of contextual information, and it is not necessary for the continued inquiries because the other phrases limit down the search regardless. In their study, the author stated that WSD is a two-step procedure. In the first stage, every word related to the phrase is disambiguated in the context of the sentence by determining probable senses with the use of auxiliary semantic resources such as dictionaries, thesaurus, and WordNet, among others. Second, overall language understanding can be ascribed based on the frequency with which words appear in a context. The second strategy, which makes use of the decision list algorithm, is used to define sense features. This work is primarily concerned with Telugu information retrieval, specifically with just how Word Sense Disambiguation enhances retrieval performance in this language.

Decision Tree

It is one of the most often used techniques in classification, and it is utilised for WSD by picking the appropriate concepts using a Yes-No tree, which is among the most well-known techniques in classification. It is possible to think of a decision tree as a binary tree in which each external node is classed by a variable and each layer is categorized with 0 or 1. The length of the longest path from the source to the leaf increases the intensity of the decision tree. Using the highest-gain feature, the training samples are separated, and the process is repeated until good DT is obtained for each group. Researchers used SENSEVAL-2 and SENSEVAL-1 information on English language to develop DT and other techniques, and discovered that SVM performed the best with image segmentation, whereas the NB operates greatest with some feature selection, and that DT provides better results with some feature selection. The researchers used senseval-3 data on the English language to develop a Decision Tree. They discovered that just a small number of terms yield reliable answers, and as a result, the total accuracy of this strategy is quite poor (45.14 percent). The advantage of using DT to filter data is that it is an effective and robust solution if the size of both the tree is small. Furthermore, it is simple to comprehend and interpret. After a brief explanation, decision tree models are easily understood by most people. The disadvantage of DT is that it is a time-consuming and complicated process when it comes to data upkeep. Instability refers to the fact that even a minor change in the data can result in a considerable shift in the efficient decision tree structure.

Naïve Bayes

Its efficiency and capacity to incorporate data from a vast number of attributes have allowed it to be successfully used to a wide range of applications and tasks in a wide range of sectors, including medicine. It can be used if the workbook is dependent on a number of different features. In our approach sense, Naive Bayes selects the category with the highest likelihood of being selected. It can function by collecting information from the words that are immediately surrounding the target word. Naive Bayes Method can be used to classify ambiguous phrases because it is considered as the most straightforward representation of probabilistic learning algorithms. The Naive Bayes algorithm was used in the studies, which were based on data from "Sense-Tagged Corpora Extracted from of the Clinical Storage Device and the MEDLINE Abstracts, on the English Language." It selects the row with both the greatest back-end probability as a starting point. With the help of the SENSEVAL-2 and SENSEVAL-1 data on the English language, Lee and Ng (2002) developed and deployed NB. El-Gamml et al. (2011) constructed the Nave Bayes Classifier on Arabic language utilising relatively tiny Personal information (lexical samples of five words) and a Nave Bayes Classifier on the same data. Benefit: Naive Bayes is a very simple, straightforward, and fast algorithm that requires no training data. Can make probabilistic forecasts about the future. A disadvantage is that there is a scarcity of data, which causes an issue. You must evaluate the possible value of a feature using an iterative process in order to gain any commercial potential for that feature.

Neural Networks

An methodology within supervised classification, neural networks replicate the interaction of artificial neurons and are used to train neural networks. A genetic technique is utilised to process data in artificial neurons, which are created in a lab. The input aspect of the learning programme seems to be the input of something like the learning programme. The objective is to split the training environment into groups that do not overlap. In order to describe words by contract, neural networks are employed, and all these phrases will stimulate the thoughts that are related with them. The intermediary layers are responsible for transferring inputs from the source to the destination layer and vice versa. [14] Although the information can be easily distributed around the network and modified to produce outputs, it is expensive to replicate a distinct output from the input. When creating a string, NN WSD employs two assessment elements that each contribute to the final outcome (word sequence, document). Using two neural networks, the score components are calculated: a first network that captures local context, and a second network that catches a global context. The researchers used WordSim-353 data on the English language to develop their NN algorithm. They came to the conclusion that the proposed multi neural language model outperformed earlier neural language models on the specific dataset. Advantage: Information can be stored over the entire network. Working with partial knowledge and the ability to absorb information in parallel are important skills. The disadvantage of NN is that it is hardware dependent and requires a parallel processing unit.

Support Vector Machines

Using the SVM, an optimization process is carried out in order to select a higher dimensional space with the biggest margin that divides training dataset into two classes. According to which side of both the hyperplane a test instance is located, the example is categorised. Prior to completing the optimization and classification, it is possible to map the input features into a high-dimensional space. It is possible to lower the computation complexity of testing and training in high-dimensional space by utilising a kernel function to do so. Support Vector Machines were used in a supervised learning technique, with nothing but the extensive training data being used as input. There would be other additional resources utilised in this project. Individual words in the immediate surroundings, regional prepositional phrases, and grammatical relationships were all utilised as knowledge sources. In the interpretation and meaning sequence of either the multilingual linguistic sampling task, the English sense of both the lexical item was also used as an extra knowledge source in addition to the translations and sense subtask. When it came to the English lexical sample task, we achieved fine-grained and coarse-grained scores (for both recall and precision) of 0.724 and 0.788, respectively, for the fine-grained and coarse-grained scores.

AdaBoost

The AdaBoost Method is a technique for developing strong classifiers from a large number of weak classifications in a linear fashion. Using this strategy, we may identify the cases that were mistakenly classified by the prior classifier and use them as input for our subsequent work in the future. The classifications are acquired from of the balanced training group, and at the start of the training group, all values are equal in importance. At each stage, it does a number of repeats, with each repetition increasing the weight of something like the corrected work, allowing some other two classifiers to concentrate on the cases that were incorrectly classified. Its central notion is to provide more importance to the misclassified training examples, which enables the subsequent classification to concentrate on examples that are truly difficult to classify [14]. This is accomplished by giving more weight to misclassified training examples. Lee and Ng (2002) used SENSEVAL-2 and SENSEVAL-1 data to apply AdaBoost to the English language. They found that it worked well. Advantage: AdaBoost can obtain equivalent classification results with significantly less adjusting of parameters or settings than traditional classification methods. The disadvantages of this algorithm include that it has a high time complexity, that it can be susceptible to measurement noise and outliers, and that it is difficult to implement on a real-time platform.

Unsupervised WSD

Unsupervised algorithms do not necessitate a training corpus, nor do they necessitate a large amount of processing time or power to run. It may be easier to obtain unidentified information from the computer than it is to obtain labelled data. In addition, when compared to supervised categorization, there may be less complexity. "It performs worse than even the supervised learning approach because it is based on less information and the data required is unknown," says the author.

Context Clustering

When it comes to Natural Language Processing, one of the most difficult problems is word sense disambiguation (WSD). The difficulty in completing this assignment stems from the fact that the context elements and the accompanying numerical analysis of data collected are specific to the individual word in the sentence. In the Bayesian framework, we provide an unique context clustering strategy that is based on modelling the similarity among bilateral circumstances at the category level, rather than at the context level. Then, using heterogeneous features that include both cooccurring keywords and parsing structures, a generative maximum entropy classifier is constructed to characterize the generating probabilistic model of paired context differences based on pairwise context similarities. The final context clusters are created via statistical annealing, which involves matching the bilateral context similarities to the final context clusters. Essentially, the finding is that the regularity of correlations between sense differentiation and context differentiation can be recorded at the category level, rather than at the level of individual words. In order to disambiguate the complete vocabulary, this approach simply determines the quantity of preexisting annotation training corpus, which is already available. In the context of the Bayesian framework, a context clustering strategy is devised. In the following step, a maximum entropy network is constructed to describe the generating posterior distribution of environment commonalities depending on heterogeneous features, such as key phrases and parsing structures, is constructed. The statistical annealing method is used to generate the ultimate context clusters by matching the bilateral contextual similarity distribution to the final context clusters.

Word Clustering

Word clustering is the process in which words are grouped together based on their semantic similarity as determined by attributes of the words. It checks for identical words that sound similar to the word goal, and the degree of similarity between some of these keywords is calculated based on the characteristics they have in common. The identical words appear to be of the same sort as the other words in the group. Because of this, the clustering technique is used to distinguish between the different types of senses. Taking a combination of words as an example, the measure of similarity is used to determine the first degree of similarity. The words are then organised in descending order of similarity, resulting in the creation of a similarity tree. The grammatical features of the target word and the Context words are used to identify the degree to which they are similar to one another.

Co-Occurrence Graph

Co-occurrence Graph is a method for unsupervised learning that is built on graph-based algorithms. It can be applied to the determination of the interpretation of a specific phrase or word. The network headings are words that occur in the very same paragraph as the target word (which will be dismantled), and they have been connected with only an edge if they exist in the same paragraph as the target word. This diagram "generates a graph with an edge of E and a vertex V representing the words in the text, and an edge E is added if those words co-occur in the same sentence or text, based on their relationship with one another according to the syntax." It is first necessary to construct the graph for a given target word, after which it is necessary to compute the graph's adjacency matrix. To determine the definition of the phrase, the Markov assembly approach is then employed." The advantage of co-occurrence is that it may be used to build strong characteristics by assembling weak data. Limitations include the fact that this does not recognize references that have been annotated with a single meaning.

The comparative analysis of the above review is summarized in table 1. The precision accuracy of the above algorithms is compared in figure 1.

Table 1 : Comparison of different algorithms and techniques on WSD

Algorithm	Language	Precision average
Lesk Algorithm	English, Arabic, Bengali	65%
Semantic	English	75%
Selectional Similarity	Bengali	78%
Heuristic Method	Bengali Language	74%
Walker's algorithm	different languages	75%
Decision List	English	58%
Decision Tree	English	60%
Naïve Bayes	Arabic, English	83%
Neural network	English	74%
Support Vector Machines	English	89%
AdaBoost	English	65%
Co-occurrence Graph	English	80%

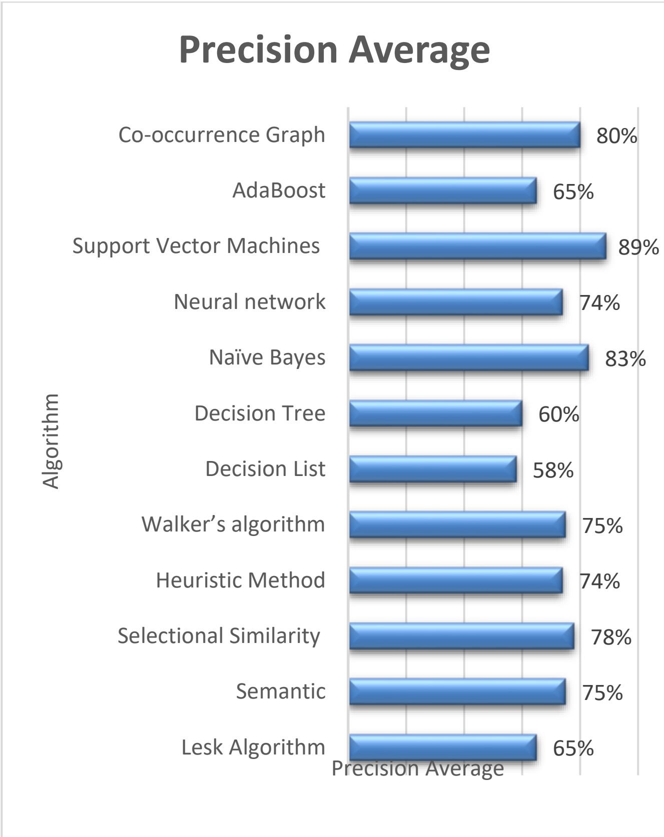

Figure 1. Comparative Analysis of the Precision of various WSD Approaches

III. CONCLUSION

Trying to figure out what a single word means is a challenging undertaking because it includes dealing with the whole complexity of the language and relying on unstructured text sources. A survey of word sense disambiguation approaches was presented in this paper, which will aid investigators in the natural language processing (nlp in selecting the methodologies that will greatest help address their specific problem in WSD. This document discussed the methodology that have been as of now have been used in sentiment classification, along with papers that introduce a study on WSD approaches. According to the findings of this study, it is impossible to make an appropriate comparison because every technique was implemented to a different data set of a different size, making it impossible to draw any conclusions. Several languages, such as with the Arabic language, demonstrate expressiveness phenomena, which might have an impact on the success of an algorithm when it is implemented. When the number of papers studied in WSD for the purpose of conducting this survey was calculated, the following conclusions were reached: While some methods continue providing high accuracy with one language while providing low accuracy for just another, the dimensionality of the data set used has an effect on the effectiveness of the automated system used, some of these methodologies can be run quickly while sacrificing accuracy, as well as the large number of the these methodologies have been implemented successfully for a multitude of distinct languages. We can design a good WSD algorithm by applying the appropriate factors into mind, which is the last but not the least: In some cases, some of the stop words can have an impact on accuracy, because of identical word which means appears to have the same neighbours. Therefore, eliminating some of the stop words can diminish precision. Additionally, the position of a word in an ambiguous term can have an impact on its meaning, which is why POS is particularly useful in word semantics. POS is also very useful in WSD. Finally, comparison tables such as those displayed in Tables 1 and 2 were created in order to survey the algorithms and publications that were chosen for this work. Table-1 compares the accuracy of all of the algorithms given, as well as the language utilised, the pros and downsides of each method. There is a comparative between each of the above works, and it contains information about each work's author as well as information about the approach category used, the methodology or technique used, the number of observations used, and the terminology employed.

IV. REFERENCES

[1]. Zhou, X. and H. Han, 2005. Survey of word sense disambiguation approaches. Proceedings of the FLAIRS Conference, May 1, pp: 307-313.

[2]. Pal, A.R. and D. Saha, 2015. Word sense disambiguation: A survey.

[3]. Giyanani, R., 2013. A survey on word sense disambiguation. IOSR J. Comput. Eng., 14: 30-33.

[4]. Sarmah, J. and S.K. Sarma, 2016. Survey on word sense disambiguation: An initiative towards an Indo- Aryan language. Int. J. Eng. Math., 6: 37-52.

[5]. Haroon, R.P., 2011. Word sense disambiguation-A survey. Proceedings of the International Colloquiums on Computer Electronics Electrical Mechanical and Civil, (EMC' 11), ACEEE, pp: 58-60. DOI: 02.CEMC.2011.01.582

[6]. Basile, P., A. Caputo and G. Semeraro, 2014. An enhanced lesk word sense disambiguation algorithm through a distributional semantic model. Proceedings of the Coling the 25th International Conference on Computational Linguistics: Technical Papers, Aug. 23-29,Dublin, Ireland, pp: 1591-1600.

[7]. Pal, A.R., D. Saha and A. Pal, 2017. A Knowledge based methodology for word sense disambiguation for low resource language. Adv. Computat. Sci. Technol., 10: 267-283.

[8]. Bakhouche, A., T. Yamina, D. Schwab and A. Tchechmedjiev, 2015. Ant colony algorithm for

Arabic word sense disambiguation through English lexical information. Int. J. Metadata, Semant. Ontol., 10: 202-211.

[9]. Jiang, J.J. and D.W. Conrath, 1997. Semantic similarity based on corpus statistics and lexical taxonomy. Proceedings of the 10th Research on Computational Linguistics International Conference, Aug. 5-7, Taipei, Taiwan, 19-33.

[10]. Meng, L., R. Huang and J. Gu, 2013. A review of semantic similarity measures in wordnet. Int. J. Hybrid Informat. Technol., 6: 1-12.

[11]. Mihalcea, R., C. Corley and C. Strapparava, 2006. Corpus-based and knowledge-based measures of text semantic similarity. Proceedings of the 21st national conference on Artificial intelligence, July 16-20, Boston, Massachusetts, pp: 775-780.

[12]. Karthikeyan, R. and V. Udhayakumar, 2015. A web search engine-based approach to measure semantic similarity between words. Transact. Know. Data Eng. 23: 977-990. DOI: 10.1109/TKDE.2010.172

[13]. Agirre, E. and D. Martinez, 2001. Knowledge Sources for Word Sense Disambiguation. In: Text, Speech and Dialogue, Matoušek, V., P. Mautner, R. Mouček and K. Taušer (Eds.), ISBN-10: 978-3-540-42557-1, pp: 229-246.

[14]. Sreenivasan, D., M. Vidya, D. Sreenivasan and M. Vidya, 2016. A walk through the approaches of word sense disambiguation. Int. J. Innov. Res. Sci. Technol., 2: 218-223.

[15]. Agirre, E. and D. Martinez, 2004. Unsupervised WSD based on automatically retrieved examples: The importance of bias. Proceedings of the Conference on Empirical Methods in Natural Language Processing, Jul. 1-3, ACL, Barcelona, Spain, pp: 25-32.

[16]. AL_Bayaty, B.F.Z. and S. Joshi, 2014. Empirical implementation decision tree classifier to WSD problem. Int. J. Adv. Technol. Eng. Sci., 2: 579-601.

[17]. Lesk, M., 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In Proceedings of the 5th annual international conference on Systems documentation (pp. 24-26). ACM.

[18]. Chatterjee, A., 2012. A survey of sense annotation, eye- tracking and wordnet linking approaches till June, 2012. Doctoral dissertation, Indian Institute of Technology, Bombay.

[19]. El-Gamml, M.M., M.W. Fakhr, M.A. Rashwan and A. Al-Said, 2011. A comparative study for arabic word sense disambiguation using document preprocessing and machine learning techniques. Proceedings of the Arabic Language Technology International Conference, Bibliotheca Alexandrina, (CBA' 11) Alexandria, Egypt.

[20]. Escudero, G., L. Màrquez and G. Rigau, 2000. Naive Bayes and exemplar-based approaches to word sense disambiguation revisited. Comput. Lang. Artificial Intel., 3: 421-425.

[21]. Liu, H., Y.A. Lussier and C. Friedman, 2001. Disambiguating ambiguous biomedical terms in biomedical narrative text: An unsupervised method. J. Biomed. Informat., 34: 249-261.

Cite this article as :

Vaishnavi Wasankar, Yashika S. Nimje, Kajal Pardhi, Divyani C. Shende, Aparitosh Gahankari, "An Empirical Analysis of Word Sense Disambiguation through Machine Learning Approaches", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 2, pp. 104-114, March-April 2022. Journal URL : https://ijsrcseit.com/CSEIT228214