IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 24, Issue 3, Ser. I (May. –June. 2022), PP 26-31
www.iosrjournals.org

Machine Learning
Investigation of machine learning and there application to WSD

Harshada Suresh Patil.
Assit. Prof. Sheetal A Wadhai
Universal College of Engineering and Research, Pune.
Head of The Department
Computer Engineering,

Abstract:
Word Sense Disambiguation (WSD) is a task in Natural Language Processing (NLP) that determines the right meaning (sense) for a given word in a text or speech that is distinguished from alternative meanings. senses that could be attributed to the term These senses could be considered the target. A classification problem's labels That is, Machine Learning (ML) appears to be a viable option. A solution to this problem This paper investigates the potential applications of the methods and techniques of Machine Learning was used to handle the WSD problem. The first issue addressed was the adaption of various ML algorithms to deal with word senses serving as classifications Following that, a comparison of various methodologies is carried out under the the same circumstances The conventional precision and recall measurements, as well as agreement rates and kappa statistics, are used to compare the results of these approaches.The second topic investigated is the cross-corpora use of supervised Machine Learning.WSD learning systems to assess generalisation capacity across corpora and domains.The results found are quite unsatisfactory, calling into serious question the possibility of creating a sufficiently broad training corpus (labelled or unlabeled), and the manner in which it is used to create a general-purpose Word Sense Tagger, samples should be used. The application of the use of unlabeled data to train classifiers for Word Sense Disambiguationis a relatively common practise. Difficult line of research in order to create a truly robust, full, and accurate. Tagger for Word Sense As a result of this, the following topic The application is the subject of this paper. Considering two WSD bootstrapping methods: Transductive Support Vector Machines and Steven Abney's Greedy Agreement bootstrapping algorithm
Keywords: *Natural Language Processing (NLP); supervised, unsupervised; knowledge Base; information retrieval; information extraction; machine translation, context; ambiguity; polysemous terms*

Date of Submission: 08-05-2022 Date of Acceptance: 23-05-2022

I. Introduction:

In natural languages, there are words that have distinct meanings in different contexts yet are spelled the same. and automatically assigning its correct sense to polysemous terms in a given context WSD [18] is a significant but difficult technique in the field of natural language processing (NLP). Many real-world applications require it, including machine translation (MT), semantic mapping (SM), semantic annotation (SA), and ontology learning (OL). It is also thought to help with the performance of various applications, including information retrieval (IR), information extraction (IE), and speech recognition (SR). Many natural languages, such as English, Hindi, French, Spanish, and Chinese, feature words whose meanings differ for the same spelling in different contexts (polysemous words). Polysemous words in English include words like run, execute, book, and so on. Humans are endowed with the ability to learn. Theycan quickly determine the correct meaning of a word in context. However, it is a difficult assignment for a computer. As a result, we must create an automatic system that can act like humans, i.e. a system that can determine the correct meaning of a word in a given context and automatically assign the optimal sense to the target word. Context refers to the text or words that surround the uncertain term. Humans can quickly perceive the correct meaning of a word in context by using context. So we also need the computer to follow some rules that will allow the system to determine the absolute meaning from a set of multiple interpretations. the phrase Consider a text T as a series of words, such as Word 1, Word 2, Word 3,.....Word n. Then, WSD is a task in which you must assign the right sense to all or some of the words in the text T. Figure 1 depicts the WSD conceptual model [3].

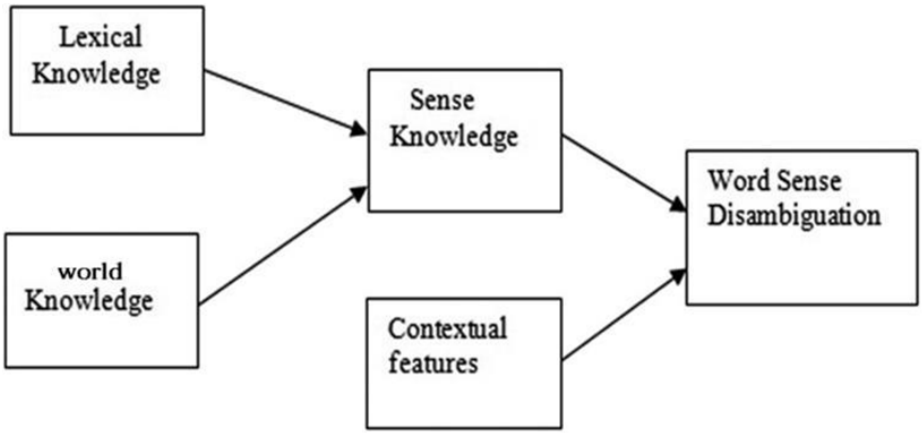

Figure 1: Word Sense Disambiguation Conceptual Model.

Machine-Learning-based techniques and Dictionary-Based approaches are the two basic approaches applied to WSD. Machine-Learning techniques are used to train systems to perform WSD tasks. A classifier is used to learn traits and assign senses to examples that have not yet been viewed. The initial input in these systems is the word to be disambiguated, known as target word, and the text in which it is placed, known as context. Dictionary-based techniques retrieve all of the senses of a term that need to be disambiguated from the dictionary. These senses are then compared to the dictionary definitions of all remaining context terms. The WSD techniques to Word sense can be further subdivided. Deep approach and Shallow approach are disambiguation.

WSD APPROACHES:

For Word Sense Disambiguation (WSD), two methodologies are used: Machine-Learning Based Approach and Knowledge Based Approach. Machine learning-based approaches teach systems to handle the task of word sense disambiguation. External lexical resources, such as Word Net, dictionaries, and thesauruses, are required in a knowledge-based approach.

Machine Learning Based Approach:

A classifier is used to learn traits and assign senses to examples that have not yet been viewed. The initial input in these systems is the word to be disambiguated, known as target word, and the text in which it is placed, known as context. The words themselves function as features in this approach. The amount of times the word appears in the region surrounding the target word determines the feature's value. The region is frequently a fixed window with the target word at the middle. There are three categories of machine learning techniques: supervised techniques, unsupervised techniques, and semi- supervised techniques.

Supervised Techniques:

It employs machine-learning techniques [5] to create a classifier from manually sense-annotated data sets. Typically, the classifier (also known as a word expert) is concerned with a single word and conducts a classification task to give the correct sense to each instance. Of that phrase Typically, the training set used to train the classifier involves a set of examples in which a given target word is manually tagged with a sense from a reference dictionary's sense inventory. Let us consider the learning process of a tiny child. The toddler has no idea how to read/write. He or she is taught at home by his or her parents and later at school by their teachers. The Children are taught and modules are developed to help them recognise the alphabets. , numbers, and so forth. Each and every one of their actions is overseen by the teacher Actually, a youngster works on the basis of the product that he or she is required to produce. Produce. A word sense disambiguation system is similarly trained using a representative collection of tagged words. Instances will be picked from the same distribution as the test set. Essentially, this WSD method provides good results. Compared to other ways The following are the methods used in

Supervise WSD: Lists of options: It is a well-organized collection of criteria for classifying test occurrences (in the case of WSD, for assigning the proper [24] sense)to a certain term). It may be thought of as a set of weighted [if-then-else] rules. A training set is employed in causing a collection of characteristics when considering any term, Its incidence is initially estimated, followed by its in terms of feature representation The vector is used to generate the decision list, from which the score is calculated. Calculated. A vector's greatest score indicates its sense. A decision tree [17] is a decision tree. Splits the training data recursively and expresses the classification rules in a tree structure. The internal nodes reflect feature tests, and each branch depicts how

the choice is made. Being constructed, while the leaf node denotes the outcome or forecast A decision tree example for WSD is depicted in Figure 2. The ambiguous word "banknoun "'s meaning is classed in the phrase, "I shall be on the Narmada River's bank in the afternoon" The tree is produced in Figure 2.and explored before deciding on a sense bank/RIVER. The leaf node's empty value indicates that there are no children. For that feature value, an option is provided. Bayes, Nave: A Naive Bayes [25] classifier is a basic classification algorithm. The Bayes theorem is used to create a probabilistic classifier. It is based on the result of the computation. Each sense's conditional probability Given the qualities f j in the context, the si of a word w S is the sense .The most acceptable sense in context is the one that optimises the following formula.

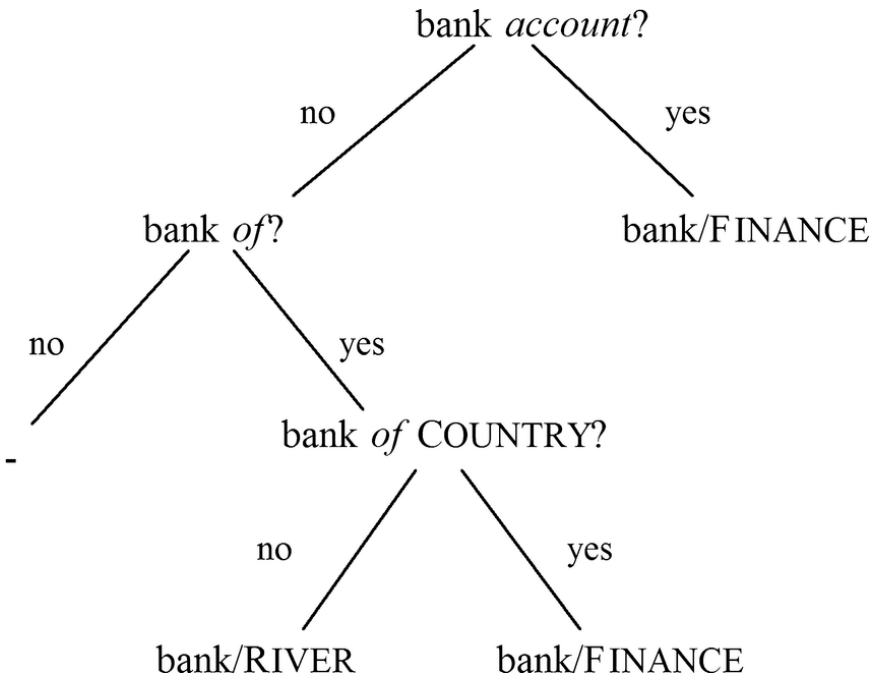

Figure 2: An example of a decision tree.

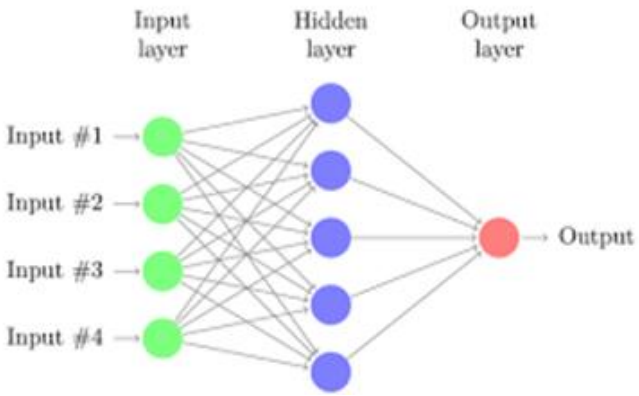

Figure 3: Neural Network Conceptual Model.

Neural Networks:

Neural networks [23] process data using a connectionist approach's computational paradigm. The goal output is included in the input, as well as the input attributes. Based on the intended responses, the training dataset is separated into non-overlapping groups. When the network meets fresh input pairs, the weights are modified such that the output unit that produces the desired output has a higher activation.

Un-supervised Techniques:

For disambiguation, an unsupervised technique, unlike a supervised approach, does not require hand labelled knowledge of sensory information in large scale resources. It is based on the concept that words with similar meanings will be surrounded by words with similar meanings. The aim is to categorise the new

occurrence to the derived clusters, which are formed by constructing clusters of word occurrences. Instead of assigning sense labels, this method discovers clusters. Context Clustering: This approach is based on clustering techniques [15], in which context vectors are constructed first, and then they are organised into clusters to determine the word's meaning. This approach employs vector space as a word space, with only words as dimensions. A word is also used in this technique. It will be designated as a vector in the corpus, and the number of times it appears will be tallied inside its context [16]. The co-occurrence matrix is then constructed, and similarity metrics are applied. After that, any clustering approach is used to accomplish discrimination. Word Clustering: Words with comparable meanings are grouped together in this strategy. Finding the same sequence of words as the target term was one of the techniques discussed in [12]. Syntactical dependence determines how related the words are. If W contains words that are similar to wm, a tree is built with only one node wm at first, and when wi is discovered to be the word, it will have a child node wm. has the closest meaning to wm Each word is represented as a feature vector in another technique termed clustering by committee algorithm [14]. When target words are found, a similarity matrix Smn is built, with each member representing a similarity between two words wm and wn. In the next phase of this technique, recursive committees are constructed for a collection of words W. The clustering algorithm then looks for terms that aren't comparable to any committee's words. These terms, which aren't part of any committee, are utilised to create new ones. In the end, each target word from W will become a member of the committee based on its resemblance to the centroid. committee. Average-link clustering was employed as the clustering approach. Co-occurrence Graphs: This approach generates a cooccurrence [13] graph with vertex V and edge E, where V represents the words in the text and E is added if the words co-occur in the same paragraph or text according to syntax. The graph for a particular target word is initially produced, followed by the graph's adjacency matrix. The Markov clustering approach is then used to determine the meaning of the word. Each graph edge is given a weight based on the frequency with which those words appear together. The formula for wmn= 1- maxP(wm| wn), P(wn| wm) gives the weight for edge m,n.

.....(2) Where P(wm|wn) is the frequency/frequency, and freqmn is the frequency. The frequency of occurrence of the word wn is freqn. Those that appear frequently are given a weight of 0, whereas words that occur infrequently are given a weight of 1. Edges with weights greater than a specified threshold are removed. The network is then subjected to an iterative procedure, with the greatest relative degree node chosen as the hub. When the frequency of a word to its hub falls below a certain level, the algorithm ends. Finally, the entire hub is marked as the meaning of the specified target term. The target word's hubs with zero weight are linked, and the graph is used to generate a minimal spanning tree. This spanning tree is used to determine the target word's true meaning.

Semi-Supervised Techniques: Information is there in semi-supervised learning approaches, just as it is in supervised learning techniques, although there may be less information offered. Only critic information, not accurate information, is provided here. For example, the system may indicate that just a portion of the intended output is valid. Semi-supervised or minimally supervised approaches are gaining popularity due to their ability to work with very little annotated reference material while outperforming completely unsupervised methods on huge datasets. There are a variety of methods and procedures for extracting key qualities from auxiliary data and clustering or annotating data with the information obtained.

Dictionary Based Approach: They may employ grammatical rules for disambiguation in a knowledge-based method based on machine-readable dictionaries in the form of corpus, WorldNet, and so on. WSD's Knowledge-based method (Dictionary-based approach) aims to use knowledge resources to deduce the meanings of words in context. Dictionaries, thesauri, ontologies, collocations, and other knowledge resources are examples. Although the preceding approaches perform worse than their supervised counterparts, they offer the advantage of a larger range.

Overlap Based Approaches: This strategy necessitates the use of a machine-readable dictionary (MDR). It entails determining the many characteristics of ambiguous word senses as well as characteristics of words in context. Lesk's formula: W is a word that creates disambiguation, C is the set of words in the context collection in the surrounding, S is the senses for W, B is the bag of words derived from glosses, synonyms, hyponyms, glosses of hyponyms, example sentences, hypernyms, glosses of hypernyms, meronyms, example sentence of meronyms, example sentence of hypernyms, glosses of hypernym then calculate the overlap using the interaction similarity criteria and output the feeling that is the most comparable.

PINE 1. evergreen tree species with needle-shaped leaves
1. deteriorate due to grief or sickness

CONE

1. A solid body that constricts to a point

2. A solid or hollow object with this form

3. The fruit of several evergreen trees

The best intersection of pine and cone, as can be observed, is

Cone #3 = 2..... Pine #1 (3)

W alke r ' s ap p r oach :
Each word in the thesaurus is assigned to one or more categories of subjects, according to this process. In distinct meanings of the term, different subjects are allocated.

Selection Preferences:
Preferences in selection [17] use the knowledge source to acquire information about the potential relationships between word categories and to signify common sense. Modeling-dress and Walk-shoes, for example, are semantically related terms. Improper word senses are excluded in this strategy, and only those senses are chosen that are in accordance with common sense principles. The primary concept behind this method is to count how many times a given word pair with syntactic connection appears in the corpus. Word senses will be identified based on this count. Other approaches, such as conditional probability, can be used to identify this type of relationship between words.

II. Conclusion:

This work compiled a list of WSD methods and categorised current WSD algorithms based on their methodologies. In this study, we present a summary of the many techniques available in word meaning disambiguation, with a particular focus on machine learning approaches and dictionary-based knowledge-based approaches. We concluded that the supervised approach outperforms the unsupervised approach, but one of its drawbacks is that it requires a large corpora without which training is impossible, which can be overcome in the unsupervised approach because it does not rely on such a large scale resource for disambiguation. The knowledge-based approach, on the other hand, relies on knowledge sources to determine the meanings of words in a given context, as long as the information is machine-readable. The information base is ready to use.

References:

[1]. C. D. Manning and H. Schutze, 1999. Statistical Natural Language Processing Foundations. Cambridge, Massachusetts: MIT Press London, United Kingdom.
[2]. F. Tacoa, D. Bollegala, and M. Ishizuka. A Supervised Word Sense Disambiguation Method Using Context Expansion. The IEEE Sixth International Conference on Semantic Computing was held this year.
[3]. J. Sreedhar, S. Viswanadha, A. Raju, V. Babu, A. Shaik, and P. Kumar An Empirical Study of Word Sense Disambiguation 2.
[4]. Agirre, E., and Edmonds, P. 2006. International Journal of Soft Computing and Engineering (IJSCE). Algorithms and Applications for Word Sense Disambiguation (Text, Speech and Language Technology). Secaucus, NJ, USA: Springer-Verlag New York, Inc.
[5]. Navigli, R. 2009. A Survey of Word Sense Disambiguation. ACM Computing Surveys, Universita di Roma La Sapienza, 41.
[6]. Lesk, M. 1986. Automatic perception How to distinguish a pine cone from an ice cream cone using machine-readable dictionaries In SIGDOC '86 Proceedings.
[7]. D. Jurafsky and J. H. Martin, 2008. Natural Language Processing, Computational Linguistics, and Speech Recognition are all covered in this course. Pearson Education is a company that specialises in education.
[8]. S. G. Kolte and S. G. Bhirud, 2008. Disambiguation of words using Word Net Domains. In ICETET'08 Proceedings.
[9]. Bala, P. 2013. Selectional Restriction for Word Sense Disambiguation. Journal of Scientific and Research Publications International.
[10]. Zheng, Z., and Z. Shu. 2009. In the MT System, a Novel Approach to Word Sense Disambiguation. Computer Science and Information Engineering World Congress
[11]. D. Lin and P. Pantel, "Discovering word senses from text," in Proceedings of the 8th ACM SIGKDD Conference. 613–619. 613–619. International Conference on Knowledge Discovery and Data Mining (Edmonton, Alberta, Canada).
[12]. Hyperlex: Lexical cartography for information retrieval, J. V'Eronis, J. V'Eronis, J. V'Eronis, J. V'Eronis, J. V'
[13]. Automatic retrieval and grouping of comparable phrases, D. Lin, 1998. 768–774 in COLING's 17th International Conference on Computational Linguistics (Montreal, P.Q., Canada).
[14]. T. Pedersen and R. Bruce. 1997. In untagged text, distinguishing word senses. 197–207 in Proceedings of the Empirical Methods in Natural Language Processing Conference (EMNLP, Providence, RI).
[15]. H. Schutze, H. Schutze, H. Schutze, H. Schutze, H. Schutze, H. Schutze, H. Schutze, H. Schutze, H. Schutze, H

[16]. R. J. Mooney, 1996. A comparison of word sense disambiguation trials demonstrates the role of bias in machine learning. In the Proceedings of the Conference on Empirical Methods in the Social Sciences, 82–91 in Natural Language Processing (EMNLP).
[17]. E. Palta, M. Tech. dissertation, department. CSE Indian Institute of Technology, Mumbai, 2006. Word Sense Disambiguation.
[18]. Structural ambiguity and lexical linkages, D. Hindle and M. Rooth, 1993. 103–120 in Computat. Ling.
[19]. P. S. Resnik, editor. Selection and information: Ph.D. dissertation on lexical associations using a class- based method. University of Pennsylvania, Philadelphia, Pennsylvania, 1993

Harshada Suresh Patil., et. al. "Machine Learning Investigation of machine learning and there application to WSD." *IOSR Journal of Computer Engineering (IOSR-JCE),* 24(3), 2022, pp. 26-31.