International Journal on Emerging Technologies **11**(4): 96-103(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Contrastive Study and Review of Word Sense Disambiguation Techniques

S.S. Patil[1], R.P. Bhavsar[2] and B.V. Pawar[3]
[1]Research Scholar, Department of Computer Engineering,
SSBT's COET Jalgaon (Maharashtra), India.
[2]Associate Professor, School of Computer Sciences,
KBC North Maharashtra University Jalgaon (Maharashtra), India.
[3]Professor, School of Computer Sciences,
KBC North Maharashtra University Jalgaon (Maharashtra), India.

(Corresponding author: S.S. Patil)
(Received 25 February 2020, Revised 10 June 2020, Accepted 12 June 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Ambiguous word, phrase or sentence has more than one meaning, ambiguity in the word sense is a fundamental characteristic of any natural language; it makes all the natural language processing (NLP) tasks vary tough, so there is need to resolve it. To resolve word sense ambiguities, human mind can use cognition and world knowledge, but today, machine translation systems are in demand and to resolve such ambiguities, machine can't use cognition and world knowledge, so it make mistakes in semantics and generates wrong interpretations. In natural language processing and understanding, handling sense ambiguity is one of the central challenges, such words leads to erroneous machine translation and retrieval of unrelated response in information retrieval, word sense disambiguation (WSD) is used to resolve such sense ambiguities. Depending on the use of corpus, WSD techniques are classified into two categories knowledge-based and machine learning, this paper summarizes the contrastive study and review of various WSD techniques.

Keywords: Computational Linguistics, Machine Learning, Machine Translation, Natural Language Processing, Word Sense Disambiguation, Word Sense Induction.

Abbreviations: CBC (Clustering by Committee), HAL (Hyperspace Analogue to Language), LSA (Latent Semantic Analysis), NLP (Natural Language Processing, NLU (Natural Language Understanding), WSD (Word Sense Disambiguation.

I. INTRODUCTION

Natural language processing is an interdisciplinary field of study, which includes linguistics and computer science and their blend emerges new branch called as computational linguistics. The objectives of natural language processing are to deal with the development of real life software applications and pertaining to natural languages using computers.

General Tasks in Natural Language Applications

Followings are the general tasks in Natural Language Applications:

Phonological Analysis: It deals with analysis of the language sounds patterns.

Morphological Analysis: It deals with words (morphemes), includes tokenization.

Syntactic Analysis: It deals with chunking and detail parsing for verifying syntactic adequacy of the sentences.

Semantic Analysis: It deals with meaning representation.

Pragmatic Analysis: It deals with understanding the actual meaning of a sentence,

II. WORS SENSE DISAMBIGUATION

Ambiguous word has same lexeme but unrelated multiple meanings, for example the word 'bank' may stands for land slide of river, financial institute and objects in the row, Word sense disambiguation is used to identify the most proper sense of such ambiguous words [1].

Example: the word '**पूर्व**' (pūrva) in Marathi language, which is the official language for Maharashtra state in India, has different meanings depends upon context, such as the sentences घराचा दरवाजा **पूर्व** दिशेला आहे (Gharācā daravājā pūrva diśēlā āhē/The door of the house is at the east) and **पूर्व** परीक्षेचा अभ्यासक्रम पूर्ण बदलेला आहे (Pūrva parīkēcā abhyāsakrama pūrṇa badalēlā āhē/The pre-examination course is completely changed), the word पूर्व (Pūrva) is an ambiguous; It creates ambiguities due to masculine noun + adjective in both the sentences.

Sense of the word पूर्व (pūrva) in the first sentence is used as दिशा (Diśā/East Direction), which is a masculine noun (पुल्लिंग नाम) and the sense of the same word पूर्व (pūrva) in the second sentence is used as पहिला, प्राचीन (Pahilā / First), which is an adjective (विशेषण).

Applications of Word Sense Disambiguation: Machine translation and semantic web search are the most highlighted applications of WSD; but proper sense disambiguation is needed in almost every NLP application [1].

Table 1: Types of Ambiguities.

Type	Description	Example
Phonetic Ambiguity	Words have same phony	Right = Write
Lexical Ambiguity	Words have multiple meanings	Bat = flying mammal/ wooden club
Syntactic Ambiguity	A sentence has multiple parse trees	Black dogs and cats
Semantic Ambiguity	A sentence has multiple meanings	I saw the man with the binoculars

— **Machine Translation**: to select the most proper sense of source words under translation.
— **Information Retrieval**: to resolve the ambiguities in the source queries for question answering.
— **Personalization**: to resolve the referential ambiguities in the user profiles.
— **Content Management**: to resolve the semantic ambiguities in content relations.
— **Lexicography:** to group and contextual indicate the sense for the lexicographers.
Types of Ambiguity: Followings are the different types of ambiguities found in the natural languages.

III. RELATED WORK

WSD has been researched over the years, depending on the use of knowledge sources (corpora) basic approaches of WSD are classified into two categories, which are knowledge-based WSD and Machine learning WSD. Machine learning approach is further classified into THREE categories, supervised WSD, semi-supervised WSD and unsupervised WSD [41]. Taxonomy of WSD approaches is shown in below tree.

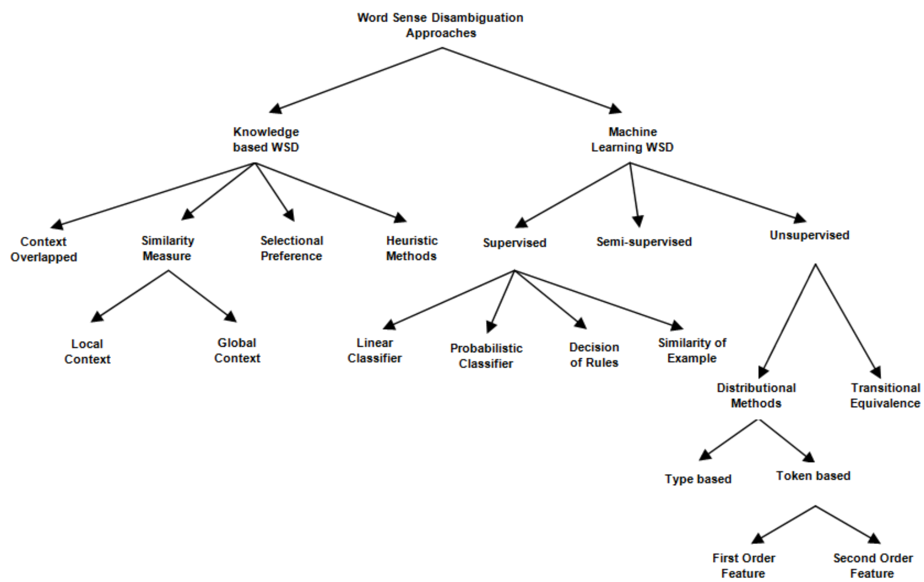

Fig. 1. Taxonomy of WSD Approaches.

A. Knowledge-based Techniques for WSD
Knowledge-based or dictionary-based WSD techniques extract the sense of word by exploiting knowledge resources such as dictionaries, thesauri and ontologies [41]. Depending upon the way of exploiting knowledge, these techniques are classified into four categories i.e. context overlap, semantic similarity, sectional preference and heuristic.
(i) Context Overlap Techniques for WSD: A context overlap technique uses contextual text overlap among dictionary definitions and the context of a target word for identifying the most likely meaning of the ambiguous word. The Lesk algorithm, Michel Lesk (1986) [30] is a pioneer work in WSD, uses overlaps between sense definitions and ambiguous word. Another variation in Lesk algorithm is by Banerjee and Pedersen (2002) [3], explores the interconnected set of relations and their associated definitions of semantically related concepts. Liu *et al.,* (2005) [33], makes the use of guess of senses or search the web for sense determination. Khapra *et al.,* (2008) [25], stated a domain specific iterative algorithm to perform WSD. Han *et al.,* (2017) [18], build overlapped Lesk based WSD using semantic and syntactic relations between context vector and sense definitions. Laatar *et al., (*2018) [27], uses the cosine distance metric semantic relation between the context of the Arabic ambiguous word and its sense definitions. Pillai *et al.,* (2018) [47], used semantic role labeling to assign the semantic role to each ambiguous word and then used Lesk algorithm. Oussalah *et al.,* (2018) [43], incorporated path length measures in WordNet and transforms of word morphology for semantic similarity, improved the chance for hitting the common sense from the context. Sharma and Joshi (2019) [52], utilized Hindi WordNet for overlapped WSD.
Review of various context overlapped knowledge-based WSD techniques is shown in Table 2.
(ii) Similarity Measures Techniques for WSD: These methods for WSD find the semantic or syntactic distance between the concepts. Depending on the features of context these techniques are divided in to two categories**.**
Local context measures: uses syntactic relations and their locality to derive the semantic measures for WSD.

Global context measures: uses lexical chain of entire text to derive the semantic measures for WSD.
Leacock (1998) [28] and Stetina (1998) [53], measured semantic distance between the syntactically parsed input sentence. Semantic relatedness is used in Patwardhan (2003) [37] and computed the cumulative Hirst score to select the proper sense. Erkan and Radev (2004) [15], proposed a sentence similarity called lexical Page Rank or LexRank, calculated an Eigen vector centrality called random walks to identify the sentence relative importance. Mihalcea (2005) [44], measured the number of common tokens between the two definitions and used Hirst (2001) [6] similarity measures for the calculation of semantic distance, and build the labeled graph.

Fuzzy graph connectivity is used in Jain and Lobiyal (2015) [21], Wattiheluw and Sarno, (2018) [58], constructed word vectors using word2vec algorithm. Review of various similarity measures knowledge-based WSD techniques is shown in Table 3.
(iii) Selection Preference Techniques for KB WSD: These techniques correlate the relations between the words in the knowledgebase and then generate the commonsense information about the groups of concepts [1]. Word to word model, word to class model and class to class model are developed by Agirre and Martinez (2001) [2]. Stevenson and Wilks (2001) [54], used shallow syntactic analysis. Kang *et al.,* 2018 [23], build a context vector and sense vector.
Review of various selectional preference knowledge-based WSD techniques is shown in Table 4.
(iv) Heuristics Techniques for KB WSD: Word sense is predicted from heuristics drawn from linguistic properties observed on long text. Gale *et al.,* (1992) [17], estimated an upper bound and lower bounds that could expect to obtain. Heuristic of a word tendency is used in David Yarowsky (1994) [59]. McCarthy *et al.,* (2004) [35], investigated a method of automatically ranking WordNet senses from raw text, used heuristics of distributional similarity. Distributional similarity is a measure indicating the degree that two words, a word and its neighbor, occur in similar contexts.
Review of various heuristics knowledge-based WSD techniques is shown in Table 5

Table 2: Review of Context Overlapped Knowledge-based WSD Techniques.

S.No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	Michel Lesk, (1986)	Non Syntactic Overlapped	OALD	English, 0.60
2.	Banerjee and Pedersen, (2002)	Relation based	English WordNet	English, Improved Lesk by 32%
3.	Liu *et al.*, (2005)	Association based	English WordNet	English, 0.77
4.	Khapra *et al.*, (2008)	Domain Specific	Multilingual WordNet	English and Marathi, 0.65
5.	Han *et al.*, (2017)	Overlapped based	SemEval-2015	English, 0.67
6.	Laatar *et al.,* (2018)	Overlapped based	Arabic Dictionary	Arabic
7.	Pillai *et al.,* (2018)	Overlapped based	English WordNet	English
8.	Oussalah *et al.,* (2018)	Overlapped based	Senseval-2	English, 0.69
9.	Sharma and Joshi (2019)	Non Syntactic Overlapped	Hindi WordNet	Hindi, 0.714

Table 3: Review of Similarity Measure Knowledge-based WSD Techniques.

S. No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	Leacock et al. (1998)	Statistical Classifier	English Sentence	English, 0.95
2.	Stetina (1998)	Semantic based	English WordNet	English, 0.83
3.	Patwardhan *et al.* (2003)	English Context	Senseval-2	English, 0.38
4.	Erkan and Radev (2004)	Random Walks	DUC 2003 News dataset	English
5.	Rada and Mihalcea (2005)	Semantic Distance based	Senseval-2	English, 0.50
6.	Jain and Lobiyal, (2015)	Fuzzy Graph Based	Hindi WordNet	Hindi
7.	Wattiheluw and Sarno, (2018)	Word2vector Technique	Twitter Dataset	English

Table 4: Review of Selectional Preference Knowledge-based WSD Techniques.

S. No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	Agirre and Martinez (2001)	Selectional Preference	English Open Text	English, 0.21
2.	Stevenson (2001)	Shallow Syntactic Tech.	English Running Text	English, 0.94
3.	Kang *et al.* 2018	Sense Context Vector Model	Korean Sense Corpus	Korean

Table 5: Review of Heuristics Knowledge-based WSD Techniques.

S. No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	Gale (1992)	Upper and Lower Measures	English Open Text	English
2.	David Yarowskey (1995)	Majority of Sense Occurrences	English News, Novels	English, 0.96
3.	McCarthy (2004)	Distributional Similarity	English Open Text	English, 0.64

B. Machine Learning Techniques for WSD

Machine learning WSD can learn the sense of word by exploiting the examples or external structured lexical sources, which is labeled, partially labeled or unlabeled [41]. Depending on the nature of source, it is categorized into three techniques, viz. supervised, semi-supervised and unsupervised.

(i) Supervised Techniques for WSD: To induce the sense for ambiguous words supervised WSD techniques uses learning to classifiers like Decision Tree, Naïve Bayes theorem, Neural Networks and Support Vector Machine [41]. To train the classifiers, this technique uses sense annotated corpus as a training data, and untagged corpus as a test data.

Followings are supervised WSD techniques:

Linear Classifier for WSD: Liner classifiers are used for WSD, which uses feature space that is represented by feature vectors of the input sample. Watanabe and Ishizaki (2006) [57], used Japanese associative ontology for WSD task, calculated associative and concept distance. Collocation features to train SVM for WSD is used by Zhong and Ng (2010) [61].

Probabilistic Classifier for WSD: A context of a target word is represented by a set of probabilistic classifier parameters. These parameters are used to generate context features for WSD. Pederson and Bruce (1997a) [45], induced a probabilistic classifier on sense tagged text and selects appropriate features; employs forward sequential search and backward sequential search. Madeeh Nayer El-Gedawy (2013) [14], adopted a fuzzy classifier for marking the senses of Arabic language, used fuzzy. Jaccard semantic word similarity. Naïve Bays classifier for the disambiguation of Gurmukhi (Panjabi) words is used by Walia *et al.,* (2017) [55]. Faisal *et al.,* (2018) [16], used TF and IDF features and Wikipedia in SVM algorithm and solve the WSD problem in Indonesian language.

Discrimination of Rules for WSD: These WSD technique selects the proper rule which is satisfied by the example features and then assigns sense based on the prediction. David Yarowsky (1996) [59], used the syntactic and semantic property of surrounding context. Dhopavkar *et al.,* (2014) [11], developed rules to train the maximum entropy model of Marathi words. Kaysar *et al.,* (2019) [24], used FP growth algorithm to create the rules for disambiguation.

Similarity of Examples for WSD: It compares a new example to a set of learned examples. Ng and Lee (1997) [42], integrated diverse set of knowledge sources. Cosine Bayesian, decision models and k-means clustering for the detection of ambiguities in seven languages is proposed in Yarowsky *et al.,* (2001) [60].

Review of various supervised WSD techniques is shown in Table 6.

(ii) Semi-supervised Techniques for WSD: Semi-supervised WSD techniques uses partially annotated corpus, it has two approaches like bootstrap and heuristic. In bootstrap approach, it starts with small seed (few labels) and uses decision list to label rest of corpus. In heuristic approach, it derives sense by observing one sense per discourse or one sense per collocation. Mihalcea (2002) [36], build disambiguation context patterns using SemCor, WordNet and GenCor dataset and employed pattern learning and instance based learning. Topic signature from large quantity of text is collected and used topic signature as a word sense in Cuadros and Rigau (2006) [8]. Diou *et al.,* (2006) [13], extracts fuzzy relations from WordNet using depth first tree traversal collected fuzzy association between context bag and sense bag to WSD. Mihalcea and Faruque (2011) [38], developed minimally supervised sense tagger, uses memory based learning and explore the sense from English WordNet. Review of various Semi-supervised WSD techniques is shown in Table 7.

Table 6: Review of Supervised WSD Techniques.

S. No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	David Yarowsky (1996)	Decision List Classifier	Spanish and French	Spanish, 0.98
2.	Ng and Lee (1996)	Example Based Classifier	English Wordnet 1.4	English, 0.89
3.	Peterson and Bruce (1997a)	Naïve Byes, K-nn	English Text	English, 0.98
4.	Yarowsky *et al.*, (2001)	Cosine and Bayesian Similarity	Seven Languages	Seven languages 0.97
5.	Watanabe and Ishizaki, (2007)	Neural Based	Associative Ontology	Japanese
6.	Zhong and Ng, (2010)	SVM	English-Chinese Parallel Corpora	English, 0.72
7.	Madeeh Nayer El-Gedawy (2013)	Fuzzy Classifier	Arabic Inventory	Arabic, 0.85
8.	Dhopavkar *et al.*, (2014)	Maximum Entropy Model	Marathi Rules	Marathi, 0.98
9.	Walia *et al.*, (2017)	Probabilistic Classifier	Gurmukhi (Panjabi) Dictionary	Gurmukhi (Panjabi)
10.	Faisal *et al.*, (2018)	Linear Classifier	Indonesian Wikipedia	Indonesia, 0.88
11.	Walia *et al.*, (2018)	Linear Classifier	Punjabi Corpora	Gurmukhi (Panjabi)
12.	Keysar *et al.*, (2019)	FP Growth	Bengali	Bengali

Table 7: Review of Semi-Supervised WSD Techniques.

S. No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	Rada Mihalcea, (2002)	Iterative	GenCor	English, 0.65
2.	Cuadros and Rigau, (2006)	Bootstrap	Central Repository	English, 0.69
3.	Diou *et al.,* (2006)	Imprecise Association	Brown Corpus	English, 0.629
4.	Mihalcea and Faruque, (2011)	Memory Based	SemCor	English, 0.64

(iii) Unsupervised Techniques for WSD: Unsupervised WSD techniques, a priorly does not require sense tagged corpus, these techniques identify the sense of ambiguous word from the neighboring words, called as context. Prepares the clusters of the word occurrences in the input text and then induce senses of new occurring word into the proper cluster [41]. Depends on the context clustering, these techniques are classified in to distributional and transitional WSD.

Distributional WSD: Distributional techniques of WSD use the similar contexts of the ambiguous words in which they appears [1].

Types based Distributional for WSD: These techniques create word co-occurrence using contextual similarity, then creates cluster of words of same meanings. Latent Semantic Analysis (LSA), Hyperspace Analogue to Language (HAL) and Clustering by Committee (CBC) algorithms are used. Pedersen and Bruce (1997b) [46], used McQuitty's similarity analysis (MSA), word minimum variance (WMV) and Expected Maximization (EM) for WSD. To determine the exact sense, collocation matrix is created by Schutze H. (1998) [50] from the dimension for each word.

Dekang Lin (1998**)** [31] proposed Cluster by Committee by using cosine similarity between two words. High dimensional context space for WSD is crested by Burgess and Lund (2000)[9], from the Hyperspace Analogue to Language (HAL) co-occurrence.

Token based Distributional Technique for WSD: Based on the similarity of the contexts, these techniques create clusters, differentiate the incidence of the each target word and then maintains the incident of the target word (token). These techniques uses first and second order features.

The first order features are contextual features and the second order features includes first order feature without the context. Lin and Pantel (2002) [32], presented Cluster by Committee, used to discover the sense from text. First and second order token based features used by Purandarae and Pederson (2004) [48], for the discrimination of word sense. Martinez *et al.,* (2006) [34], proposed relational based WSD and obtained polysemous relations between word sense and words. Graph based approach for word sense induction is proposed by Hope and Keller (2013)[19]. Bhingerdive *et al.,* (2013) [4], computed semantic relatedness of neighboring words from Wordnet graph for the disambiguation of Marathi and Hindi verbs.

Transitional Equivalence for WSD: This technique uses parallel corpus of two languages and word translation mappings. Diab and Resnik (2002) [12], developed stochastic French-English parallel translative approach for WSD, employs pseudo translation on Brown corpus. Graph based approach for WSD is used by JeanVeronis, (2004) [22], for French web pages and constructed co-occurrence graphs for polysemous words. Mihalcea *et al.,* (2004) [39], explores graph based ranking algorithm and performs open text disambiguation. Construction of graph from local measures like eigenvector centrality and global measures like compactness, graph entropy and edge density of each word by Navigli and Lapta (2007) [40], and proposed graph based WSD approach for English. Lefever *et al.,* (2011) [29] proposed a language independent framework. Fuzzy c-means Spanish and German word mapping is proposed by Ren and Ren (2015) [49]. Bilingual Expected Maximization based WSD is proposed by Khapra *et al.,* (2011) [25]. Bhingerdive and Bhattachraya (2017) [5], prepared a context of two languages.

Table 8: Review of Distributional Un-supervised WSD Techniques

S. No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	Pedersen and Bruce (1997)	Agglomerative and Probabilistic Clustering	Wall Street Journal Corpus	English
2.	Schutze H. (1998)	Similarity Context Clustering	English Corpora and a Thesaurus	English
3.	Dekang Lin (1998)	CBC Similarity	English Wall Street Journal	English
4.	Burgess and Lund (2000)	HAL Association Approach	English Verb	English
5.	Lin and Pantel (2002)	CBC Similarity	SemCor	English
6.	Purandarae and Pederson (2004)	First and Second Order Context Clustering	Senseval 2 English Dataset	English
7.	Martinez *et al.*, (2006)	All Words Unsupervised Approach	Raw Corpora and Thesaurus	English, 0.98
8.	Hope and Keller (2013)	Graph Based Clustering	English Verbs	English, 0.72
9.	Bingerdive *et al.* (2013)	EM Based Clustering	Hindi and Marathi Verb	Hindi and Marathi

Table 9: Review of Transitional Un-supervised WSD Techniques.

S.No.	Author(s)	Methodology	Corpora	Language and Accuracy
1.	Schutze (1992)	Context Clustering	New York Times News	English
2.	Ide *et al*., (2002)	Hierarchical Clustering	Orwell Parallel Corpora	Seven Languages
3.	Diab and Resnik (2002)	Stochastic Based	Brown Corpus	English. 0.60
4.	Jean Vernosis, (2003)	Semantic Distance Based	French Web Page Corpus	French, 0.97
5.	Mihalcea *et al*., (2004)	Iterative Graph Based	English WordNet	English, 0.62
6	Navigli and Lapta, (2007)	Depth First Search	English WordNet	English, 0.31
7.	Khapra *et al*., (2011)	EM Based Bilingual	Bilingual Dictionary	Hindi and Marathi
8.	Lefever *et al*., (2011)	Cross Lingual	Europarl Corpus	French, Dutch, Italian etc.
9.	Ren and Ren (2015)	FCM and KCM Clustering	Medline Biomedical Data	English, 0.81
10.	Bhingerdive and Bhattacharyya (2017)	Context and EM	Hindi News Paper	Hindi and Marathi

IV.CONTRASTIVE ANALYSIS

During this survey, we observed the following contrast between knowledge-based and machine learning techniques, the analysis is presented below:

Knowledge-based techniques disambiguate all the words in open text, while machine learning targets to only few selected words. Knowledge-based techniques have the wider coverage in the availability of large-scale knowledge resources, this is not the case of there supervised machine learning counter parts, for wide coverage, supervised techniques needs large scale training corpus. To infer the proper sense for the ambiguous word knowledge-based technique needs to exploits complete structured knowledge resources like thesauri, machine readable dictionaries and ontologies, this is not the case with remaining WSD techniques.

In the availability of sense labeled training corpus, supervised WSD techniques yield higher precision (crosses 90%), this is not in the case of their alternate WSD techniques. However, availability of large training labeled corpora is not a realistic assumption in all scenarios; also it is language expert specific efforts, which may be time consuming.

During our study, we found that there is no clear cut boundary between semi-supervised and supervised techniques. Semi-supervised techniques acquire knowledge through self iterations and then they perform the sense labeling on partial annotated data, so unlike supervised techniques, it involve sense iteration on partial annotated data. These methods start with a proper randomly chosen seed point.

Unsupervised techniques are better choice in the lack of large-scale structured knowledge and manually sense annotated corpus. They induce the sense of ambiguous word from the context of neighboring words. They overcome the problem of knowledge acquisition bottleneck, witnessed in the knowledge-based and supervised techniques. Due to their minimal requirements and linguistic knowledge independence, they can be easily extended to other languages.

So, from the above discussion, it is apparent that each WSD technique has its own pros and cons and choice of the technique is constrained by the problem under study and available resources.

V. CONCLUSION

During this study, we have investigated knowledge-based, supervised, semi-supervised and unsupervised WSD techniques in detail on different parameters. Different reported works have also been surveyed and reviewed for national as well as international languages. In the computational scenario, the WSD community has mostly shifted from hand crafted to automatic learning techniques. On the language side, we found that maximum work has been done for English WSD using knowledge-based, supervised, semi-supervised and unsupervised techniques. It is observed that, the knowledge-based and supervised techniques have yielded accuracy around 90% or beyond 90%, but they lack completeness. The work in Indian language scenario research is picking up for the languages like Hindi, Marathi, Guajarati and Punjabi. Due to the applications of WSD in machine translations and semantic web, overall WSD community is now shifting towards unsupervised techniques, where context is used as a main clue for sense disambiguation.

VI. FUTURE SCOPE

In future, topic modeling and fuzzy context clustering can be uses to disambiguate the words in the resource scared Indian languages like Hindi, Marathi, Guajarati, and Punjabi.

Conflict of Interest. No potential conflict of interest was reported by the authors.

REFERENCES

[1] Agirre, E., and Edmonds, P. (2007). *Word Sense Disambiguation: Algorithms and Applications* (Vol. 33). New York: Springer.

[2] Agirre, E., and Martinez, D. (2001). Learning class-to-class selectional preferences. *Workshop on Computational Natural Langauge Learning.* France: ACM.

[3] Banerjee, S., and Pederson, T. (2002). An Adaptive Lesk Algorithm for Word Sense Disambiguation using WordNet. *Proceedings of the Third International Conference on Computational Linguistics and Intelligent Text Processing.* Verlag, London: Springer.

[4] Bhingardive, S., Shaikh, S., and Bhattacharyya, P. (2013). Neighbors Help: Bilingual Unsupervised WSD Using Context. *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, pages , Sofia, Bulgaria, Au* (pp. 538–542). Sofia, Bulgaria: ACL.

[5] Bhingardive, S., and Bhattacharyya, P. (2017). Word Sense Disambiguation Using IndoWordNet. *The WordNet in Indian Languages*, 243-260.

[6] Budanitsky, H., and Graeme, A. (2001). Semantic Distance in WordNet: An Experimental, Application-Oriented Evaluation of Five Measures. *Proceedings of the NAACL Workshop on WordNet and Other Lexical Resources*, (pp. 29-34). Ittsburgh.
[7] Burgess, C., and Lund, K. (1998). Automatic retrieval and clustering of similar words. *ACL '98/COLING '98* (pp. 768-774). Quebec, Canada: ACL.
[8] Cuadros, M., and Rigau, G. (2006). Quality assessment of large scale knowledge resources. *EMNLP '06* (pp. 534-541). Sydney, Australia: ACM.
[9] Curt Burgess, K. L. (2000). Modelling Parsing Constraints with High-dimensional Context Space. *Language and Cognitive Processes, 12*(2/3), 177-210.
[10] David, Y. (1995). Unsuprevised Word Sense Disambiguation Rivaling Supervised Method. *33rd Annual Meeting of the Association for Computational Linguistics* (pp. 189-196). Cambridge, Massachusetts, USA: Association for Computational Linguistics.
[11] Dhopavkar, G., Kshirsagar, M., and Malik, L. (2014, April). Handling Word Sense Disambiguation in Marathi Using a Rule Based Approach. *IJERA* , 13-16.
[12] Diab, M., and Resnik, P. (2002). An Unsupervised Method for Word Sense Tagging using Parallel Corpora. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL)* (pp. 255-262). Philadelphia: ACL.
[13] Diou, C., Katsikatskos, G., and Delopoulos, A. (2006). Constructing Fuzzy Relations from WordNet for Word Sense Disambiguation. *SMAP06*. IEEE.
[14] El-Gedawy, M. N. (2013). Using Fuzzifiers to Solve Word Sense Ambiguation in Arabic Language. *International Journal of Computer Applications , 79* (2), 1-8.
[15] Erkan, G., and Radev, D. R. (2004). LexRank: Graph-based Lexical Centrality as Salience in Text Summarization. *Journal of Artificial Intelligence and Reasearch, 22*, 457-479.
[16] Faisal, E., Nurifan, F., and Sarno, R. (2018). Word Sense Disambiguation in Bahasa Indonesia Using SVM. *Semantic-2018*, (pp. 239-243).
[17] Gale, W. (1992). Estimating Upper and Lower Bounds on the Performance of Word-Sense Disambiguation Programs. *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics* (pp. 249-256). Newark, Delaware, USA: Association for Computational Linguistics.
[18] Han, L., Deng, X., and Wu, G. (2017). A Knowledge-Based Word Sense Disambiguation Algorithm Utilizing Syntactic Dependency Relation., (pp. 85-70).
[19] Hope, D., and Keller, B. (2013). MaxMax: A Graph-Based Soft Clustering Algorithm Applied to Word Sense Induction. *CICLing 2013* (pp. 368-381). Berlin, Heidelberg: Springer .
[20] Ide, N., Erjavec, T., and Tufis, D. (2002). Sense Discrimination with Parallel Corpora. *SIGLEX/SENSEVAL Workshop on WSD Recent Success and Future Directions* , (pp. 54-60). Philadelphia.
[21] Jain, A., and Lobiyal, D. K. (2015). Fuzzy Hindi WordNet and Word Sense Disambiguation Using Fuzzy Graph Connectivity Measures. *ACM Trans. Asian and Low-Resource Lang. Inf. Process, 15* (2), 8:1-8:31.
[22] JeanVéronis. (2004). HyperLex: lexical cartography for information retrieval. *Computer Speech and Language* , 223-252.
[23] Kang, M. Y., Min, T. H., and Lee, J. S. (2018). Sense Space for Word Sense Disambiguation. *IEEE International Conference on Big Data and Smart Computing* (pp. 669-672). IEEE.
[24] Kaysar, M. S., Khaled, M., Hasan, M., and Khan, M. I. (2019). Word Sense Disambiguation of Bengali Words using FP-Growth Algorithm. *ECCE2019*. IEEE.
[25] Khapra, M. M., Joshi, S., and Bhattacharyya, P. (2011). It Takes Two To Tango: A Bilingual Unsupervised Approach For Estimating Sense Distributions Using Expectation Maximization. *5th International Joint Conference on Natural Language* (pp. 695-704). Chiang Mai, Thiland: AFNLP.
[26] Khapra, M., Bhattacharyya, P., Chauhan, S., S, N., and Sharma, A. (2012). Domain Specific Iterative Word Sense Disambiguation in a Multilingual Setting. *International Conference on NLP (ICON08).* Pune, India.
[27] Laatar, R., Aloulon, C., and Belguith, L. H. (2018). Word Sense Disambiguation using Skip Gram Model to Create a Historical Dictionary for Arabic. *AICCSA-2018.* IEEE.
[28] Leacock, C., Chodorow, M., and Miller, G. A. (1998). Using Corpus Statistics and WordNet Relations for Sense Identification. *Association for Computational Linguistics 24(1), 24* (1), 147-165.
[29] Lefever, E., Hoste, V., and De Cock, M. (2011). ParaSense or How to Use Parallel Corpora forWord Sense Disambiguation. *Association for Computational Linguistics* , 317-322.
[30] Lesk, M. (1986). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. *International Conference on System Documentation* (pp. 24-27). Toronto: ACM.
[31] Lin, D. (1998). Automatic Retrieval and Clustering of Similar Words., (pp. 768-774).
[32] Lin, D., and Pantel, P. (2002). Concept Discovery from Text. *COLING 2002*.
[33] Liu, S., Yu, C., and Meng, W. (2005). Word Sense Disambiguation in Queries. *14th ACM international conference on Information and knowledge management(CIKM).* Bremen, Germany: ACM.
[34] Martinez, D., Agirre, E., and Wang, X. (2006). Word Relatives in Context for Word Sense Disambiguation. *Australasian Language Technology Workshop 2006* (pp. 42-50). Sydney: ALTA.
[35] McCarthy, D., Koeling, R., Weeds, J., and Carroll, J. (2004). Finding Predominant Word Senses in Untagged Text. *42nd Meeting of the Association for Computational Linguistics (ACL'04), Main Volume* (pp. 279-286). Barcelona, Spain: Association for Computational Linguistics.
[36] Mihalcea, R. (2002). Bootstrapping Large Sense Tagged Corpora. *Third International Conference on Language Resources and Evaluation.* Canary Islands - Spain: European Language Resources Association (ELRA).
[37] Mihalcea, R. (2005). Unsupervised Large-Vocabulary Word Sense Disambiguation with Graph-based Algorithms for Sequence Data Labeling. *In Proceedingsof Human Language Technology Conference and Conference on EmpiricalMethods in*

Natural Language Processing, (pp. 411-418). Vancouver.
[38] Mihalcea, R., and Faruque, E. (2011). SenseLearner: Minimally Supervised Word Sense Disambiguation for All Words in Open Text. *Proceedings of Sen-seval-3: Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text, Barce.* Barcelona, Spain.
[39] Mihalcea, R., Tarau, P., and Figa, E. (2004). PageRank on semantic networks, with application to word sense disambiguation. *COLING '04.* Geneva, Switzerland.
[40] Navigli, R., and Lapata, M. (2007). Graph Connectivity Measures for Unsupervised Word Sense Disambiguation. *Proceedings of the 20th International Joint Conference on Artifical Iintelligence* (pp. 1683-1688). Hyderabad, India: ACM.
[41] Navigli, R. (2009). Word Sense Disambiguation: A Survey, Journal ACM Computing Survey. *ACM Computing Survey, 41* (2), 1-69.
[42] Ng, H. T., and Lee, B. L. (1997). Integrating Multiple Knowledge Sources to Disambiguate Word Wense: An Exemplar-Based Approach. *34th Annual Meeting of the Association for* (pp. 40-47). Santa Cruz, CA: ACL.
[43] Oussalah, M., Väisänen, M., and Gilman, E. (2018). On Wordsense Disambiguation Through Morphological Transformation and Semantic Distance and Domain link knowledge*. *IEEE International Conference on Information Reuse and Integration for Data Science*, (pp. 117-121).
[44] Parwardhan, S., Banerjee, S., and Perderson, T. (2003). Using measures of semantic relatedness for word sense disambiguation. *4th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2003)*, (pp. 214-257). Mexico City.
[45] Pedersen, T., and Bruce, R. (1997). A New Suprvised Learning Algorithm for Word Sense Disambiguation. *fourteenth National Conference on AAAI* (pp. 197-207). AAAI.
[46] Pedersen, T., and Bruce, R. (1997). Distinguishing Word Sense in Untagged Text. *Second Conference on Empirical Methods in Natural Language Processing*, (pp. 197-207). Providence, RI.
[47] Pillai, L. R., G, V., and Gupta, D. (2018). A Combined Approach Using Semantic Role Labelling and Word Sense Disambiguation for Question Generation and Answer Extraction. *ICAECC-2018.* IEEE.
[48] Purandare, A., and Pedersen, T. (2004). Word Sense Discrimination by Clustering Contexts in Vector and Similarity Spaces. *CoNLL 2004* (pp. 41-48). Boston: ACL.
[49] Ren, K., and Ren, Y. F. (2015). Kernel Fuzzy C-Means Clustering for Word Sense Disambiguation in BioMedical Texts. *Journal of Digital Information Management , 13* (6), 411-421.
[50] Schutze, H. (1998). Automatic Word Sense Discrimination . *Computational Linguistics* , 97-123.
[51] Schutze, H. (1992). Dimensions of Meaning. *ACM/IEEE Conference on Supercomputing* (pp. 787-796). Minnesota, USA: IEEE.
[52] Sharma, P., and Joshi, N. (2019). Design and Development of a Knowledge-Based Approach for Word Sense Disambiguation by using WordNet for Hindi. *International Journal of Innovative Technology and Exploring Engineering, 8* (3), 73-78.
[53] Stetina, J., Kurohashi, S., and Nagao, M. (1998). General Word Sense Disambiguation Method Based on a Full Sentential Context. *Proceedings of the workshop on Usage of WordNet in Natural Language Processing*, (pp. 1-8). Montreal, Canada.
[54] Stevenson, M., and Wilks, Y. (2001). The Interaction of Knowledge Sources in Word Sense Disambiguation. *Computational Linguistics, 27* (3), pp. 321-349.
[55] Walia, H., Rana, A., and Kansal, V. (2017). A Naïve Bayes Approach for working on Gurmukhi Word Sense Disambiguation. *ICRITO-2017* (pp. 432-435). Noida, India: IEEE.
[56] Walia, H., Rana, A., and Kansal, V. (2018). A Supervised Approach on Gurmukhi Word Sense Disambiguation using k-NN Method. *8th International Conference on Cloud Computing, Data Science and Engineering* (pp. 743-746). IEEE.
[57] Watanabe, N., and Ishizaki, S. (2006). Neural Network Model for Word Sense Disambiguation using Up/Down State and Morphoelectrotonic Transform: Learning Words Distance with Small-world Structure. *SCIS andISIS2006* (pp. 451-456). Tokyo, Japan: World Congress of International Fuzzy Systems Association .
[58] Wattiheluw, F. H., and Sarno, R. (2018). Developing Word Sense Disambiguation Corpuses using Word2vec and Wu Palmer for Disambiguation. *iSemantic-2018* (pp. 244-248). IEEE.
[59] Yarowsky, D. (1994). Decision Lists For Lexical Ambiguity Resolution: Application to Accent Restoration in Spanish and French. *32nd Annual Meeting of the Association for Computational Linguistics* (pp. 85-90).
[60] Yarowsky, D., Cucerzan, S., and Florian, R. (2001). The Johns Hopkins SENSEVAL2 System Descriptions. *SENSEVAL2* (pp. 163-166). Toulouse, France: ACM.
[61] Zhong, Z., and Ng, H. T. (2010). It Makes Sense: A Wide-Coverage Word Sense Disambiguation System for Free Text. *ACL 2010 System Demonstration* (pp. 78-83). Uppsala, Sweden: ACL.

How to cite this article: Patil, S. S., Bhavsar, R. P. and Pawar, B. V. (2020). Contrastive Study and Review of Word Sense Disambiguation Techniques. *International Journal on Emerging Technologies*, *11*(4): 96–103.