Word Sense Disambiguation Based on the Context, Structure and Meaning

Amir Hossein Rasekh*, Mohammad Hassan Dianaty†, Hossein Taghi-Zadeh‡,
Seyed Mostafa Fakhrahmad§, Mohammad Hadi Sadreddini¶
School of Computer and Electrical Engineering
Shiraz University, Shiraz, Iran
Email: {ahrasekh*, mfakhrahmad§, sadredin¶}@shirazu.ac.ir, { h-taghizadeh‡, dianaty†}@cse.shirazu.ac.ir

***Abstract*—Development of computer linguistics is a popular field in Natural Language Processing (NLP) in order to overcome some challengeable problems of translation in despite of constraints and bottlenecks. Machine translation is considered as a branch of machine intelligence and its background returns to about half century ago. Ambiguity of language is the most problematic issue in machine translation systems, which may lead to an unclear or wrong translation. One of the problems involved in NLP is the semantic and structural ambiguity of the words. The goal of this paper is focused on the word sense disambiguation (WSD) and the proposed method is based on the concept and structure of words. Experimental results prove that our approach improves the quality of machine translation.**

Keywords—Natural Language Processing; Word Sense Disambiguation; Data Mining; WordNet.

I. Introduction

Machine translation is a branch of computational linguistics where a machine is employed to understand a text completely and then translate it into another language. Using corpus linguistics techniques, translation of more complex texts is facilitated. Furthermore, these techniques provide a strong way in order to make better recognition and translation of different phrases and terminologies of all most languages.

Machine translation includes some features which justify the necessity of it from economic the perspective. For instance, despite of working about 1,200 translators at NATO National Headquarters in Brussels and Europe, machine translation is also used in order to decrease the speed and cost of translation. Also translation of text using a machine is faster than a human translator. Although, the quality and accuracy of machine translation is lower than the human translator, the machine is more applicable in point of different measure and values.

Machine translation is a very active topic in the field of machine intelligence. The first experience of machine translation was on the translation of a text from English to Russian. Despite of many studies already performed on machine translation, there are still several defects. It is important that translation of any text (simple/complex) using a machine is not done easily. Ambiguity in semantic and structure of the language are the most problematic issues in machine translation systems. The most important task of machine translation is to recognize the role of a word in a sentence and then translate it to its correct meaning in the target language.

Disambiguation requires to consider semantics, syntax and word processing issues. Unfortunately, there is a big gap between the theoretical linguistics and its application. Theoretical linguistics has failed to present a general model that covers all aspects of the language.

A major type of ambiguity is resolved using a technique known as ”Part Of Speech Tagging” which is determine and annotate the role of all words in the sentence [1]. In other words, the method specifies the right role of every word in a sentence which would facilitate determining the correct meaning of the word. Note that the word sense disambiguation is more difficult than the structural ambiguity resolution, and a stronger method is required to solve this problem. Actually, word sense disambiguation is carried out by identifying the sentence which contains the ambiguous word followed by recognition of the words surrounding the ambiguous word.

All WSD methods already proposed can be classified in three main categories. The first category includes supervised methods which use a sense-tagged dataset to train a classifier; the second one uses lexical resources such as dictionary or thesaurus culture [2]; and finally, the third category includes unsupervised methods which work on unlabeled set of words and texts [3].

A. Disambiguation Using Supervisor

In this approach, a set of texts containing the ambiguous word are used as the training data. The correct sense of each word is also annotated in each text, during the training phase.

This approach can be considered as a classification task on the words. In other words, the algorithm is trained in order to classify a new unseen case of the ambiguous word by determining its correct sense. In supervised learning approach, different well-known classifiers have been applied including Bayesian Network [4], Information Theory [6], K-Nearest Neighbor, Decision Tree and Neural Network.

B. Disambiguation based on Dictionary

This method is used when no information about the class of the words in the corpus is available. In such cases, the general information about the word is explored from the dictionary. In this approach, different types of information are used: the first type is called Lesk which uses exactly the meaning of the word which is defined in the dictionary; in the second type, the algorithm uses the classification information of the word which is defined in the dictionary [5]; and in the last type, the information about the translation which is achieved from a dictionary of two different languages is used.

C. Disambiguation Using Unsupervised Learning

In two previous disambiguation methods, the procedure needs the prior knowledge about the semantic of words, while in some cases there is no information about the meaning of the words included in the text [2, 6]. The first step of this approach is clustering of the words based on semantic of the cluster. Each cluster has a semantic, and each word is assigned to the nearest cluster. Here, the word takes the meaning of the cluster in the sentence.

This paper is organized as follows. In Section 2, the related work in this domain is reviewed. Proposed method is described in Section 3. The experimental results are presented in Section 4. Finally, Sections 5 concludes the paper.

II. Related Works

The set of all statistical and example-based WSD approaches already proposed can be divided into two main categories namely, knowledge-based and corpora-based approaches (Navigli 2009) [7]. Knowledge-based approaches use external lexical resources, such as dictionaries, ontologies, thesauri, etc., for disambiguation. However, the corpora-based approaches do not make use of any of these resources.

From another point of view, the set of WSD methods can be divided into three major categories namely, supervised, unsupervised and hybrid schemes. The first category includes methods which are based on supervised learning. These methods use classification systems to determine the correct translation of ambiguous words. The second category includes methods that use unsupervised learning. Text clustering is the main learning process used by the methods in this category. There is also another category of disambiguation methods which propose a combination of supervised and unsupervised learning.

A. Supervised Approaches

There are a lot of proposed methods for word sense disambiguation which follow supervised learning techniques, e.g., Naive Bayesian (Gale 1992) [8], Decision List (Yarowsky 1994) [9], Nearest Neighbor (Ng and Lee 1996) [10], Transformation Based Learning (Mangu and Brill 1997) [11], Winnow (Golding and Roth 1999) [12], Boosting (Escudero et al. 2000) [13], SVM (Murata et al. 2001) [14] and Naive Bayesian Ensemble (Pedersen 2000) [15]. Among the mentioned methods, the methods that use SVM and Naive Bayesian Ensemble have been reported to have the best performance for ambiguity resolution tasks, respectively (Navigili 2009, Pederson 2000). In order to determine the correct translation of each ambiguous word, all of the above methods build a classifier, using features that represent the context of the ambiguous word.

Brown et al. (1991) proposed a corpora-based disambiguation method which can be applied in machine translation systems [16]. They use data from syntactically related words in the local context of the ambiguous word. In order to obtain statistical data, a word-aligned bilingual corpus was required. Each occurrence of an ambiguous word should be labeled with a sense by asking a question about the context in which the word appears.

The other method proposed by Yarowsky et al. (1992) assumes that each word is located in a major category [17]. In order to disambiguate word senses they have used the Rogets Thesaurus data set. By searching the hundred surrounding words as indicators of each category, the most probable category of a word can be determined. During the training phase, a stemming process is performed over all words in order to achieve more useful statistics. Subsequently, by examining the hundred surrounding words for indicators of each category, the indicator words are obtained and weighted. The measure used as the weight of each indicator word is the log of words salience as shown in (1)

$$weight(w \quad in \quad cat) = log\frac{P(w|cat)}{P(w)} \qquad (1)$$

Here, w is an indicator word and cat stands for a category. $P(w|cat)$is the probability that w appears in the context of a word from the category cat and P(w) is the probability of the ws occurrence in the corpus as a whole. For useful words, the computed weight, i.e., the log of salience will be greater than one.

The system proposed by Yarowsky et al. (1992) is not limited to particular word categories and works in a wide domain. The first challenge of the system is that it cannot disambiguate topic-independent distinction words that occur in many topics. Moreover, the system does not consider the distance of words in the contexts it handles.

Another method for word sense disambiguation was proposed by Dagan and Itai (1994) [18]. The method chooses the most probable sense of a word using frequencies of the related word combinations in a target language corpus. In this method, first of all, the system identifies syntactic relations between words using a source language parser and maps those relations to several possibilities in the target corpus using a bilingual lexicon.

The other method of word sense disambiguation proposed by Justeson and Katz (1995), uses syntactically or semantically relevant clues [19]. This method disambiguates adjectives using only nouns that are combined by the adjectives. The system was evaluated on five of the most frequent ambiguous adjectives in English: right, hard, light, old, and short on large sets of randomly selected sentences from the corpus that contained the adjectives. However, for adjectives which can be differently accompanied by the same noun, this method cannot be helpful in disambiguation.

The system presented by Ng and Lee (1996) is based on the Nearest Neighbor method [20]. The prototypes are the instances of the ambiguous word in the training corpus, each containing the following features: singular/plural; POS tags of the current word; three words on either side; support for verbs, which have a different verbal morphological feature; a verbobject syntactic feature for nouns; and nine local collection features. These features are calculated for each instance of w in the sense-tagged training data. The results are stored as exemplars of their senses. By calculating the same feature vector for the current word and comparing by all the examples of that word, the given word is disambiguated choosing the closest matching instance.

The method presented by Brown et al. (1991) requires a bilingual word-aligned corpus, which is costly to build. This is one of the challenges of this method, which reduces the applicability of the method to other pairs of languages.

The other method proposed by Mosavi and Khalafi (2005) [21] is somewhat the same as the method presented by Dagan and Itai (1994) which uses a target language model. They use Persian as the target language and consider the co-occurrences of the multiple-meaning words in a monolingual corpus of the Persian language. By calculating the frequencies of these words in the corpus, the most probable sense for the multiple-meaning words is chosen. However, instead of considering syntactic tuples in the target language corpus, they consider just co-occurrences of certain words in that corpus without having a syntactic analysis for the corpus. In this method, no analysis is performed either for the source or the target language corpus from the syntactic viewpoint. The only task of the proposed algorithm, for gaining the required statistical information, is performed by determining the nearest noun, pronoun, adjective, or verb to the ambiguous word, whether it is a noun, a verb, an adjective, or an adverb. When applying this method for the comparison of English and Persian, only a small portion of ambiguous words in English can be correctly translated into Persian.

Reddy et al. (2010a), investigated WSD by modeling it in a distributed constraint optimization (DCOP) framework. The method requires information from various knowledge sources (including part-of-speech, morphology, domain information, etc), in order to be modeled in a multi-agent setting [22].

B. Unsupervised and semi-supervised approaches

In addition to supervised approaches, unsupervised approaches and combinations of them have also been proposed for word sense disambiguation.

The proposed approaches to unsupervised WSD, can be divided into some major categories, which are based on context clustering (Schutze 1998), word clustering (Lin 1998) and co-occurrence graphs (Widdows and DOROW 2002) [23], respectively.

Among these main approaches, graph-based methods have been recently explored with a certain success. These approaches are based on the notion of a co-occurrence graph. A co-occurrence graph is a graph like G = (V, E), whose vertices V correspond to words in a text and edges E connect pairs of words which have co-occurred in the same context.

For example, Veronis et al. (2004) proposed an ad hoc approach called HyperLex [24]. In the First step, a co-occurrence graph is built such that nodes are words that have occurred in the paragraphs of a corpus in which a target word occurs, and an edge between a pair of words is added to the graph if they co-occur in the same paragraph. A weight is then assigned to each edge, according to the co-occurrence frequency of the pair of words connected by the edge.

Sinha and Mihalcea (2007) proposed an unsupervised graph-based word sense disambiguation algorithm, which combines several word semantic similarity (i.e., the degree to which two words are semantically related) measures and algorithms for graph centrality (i.e., the algorithms which assign different importance degrees to the nodes) [25]. That work was the first method that addressed the problem of WSD by comparatively evaluating measures of word semantic similarity in a graph theoretical framework.

The method proposed by Reddy et al. (2010b) uses the Personalized PageRank which is a graph centrality algorithm (Agirre and Soroa 2009) over a graph representing WordNet to disambiguate ambiguous words by taking their context into consideration [26].

There are some other unsupervised methods in the literature, which are based on word clustering or context clustering. For example, Schutze (1998) [27] proposed an ambiguity resolution technique which divides the occurrences of a word into a number of clusters by determining for any two occurrences whether they belong to the same sense or not, which is then used for the full ambiguity resolution task. The approaches proposed by Litkowski (2000) [28] and Lin (2000) [29] are other examples of unsupervised learning methods. Nigam et al. (2000) [30] had proposed an unsupervised learning method using the Expectation-Maximization (EM) algorithm for text classification problems, which then was improved by Shinnou and Sasaki (2003) in order to apply it to the ambiguity resolution problem [31]. Agirre et al. (2000) combined both supervised and unsupervised lexical knowledge methods for word sense disambiguation [32]. In two other methods, Yarowsky (1995) [33] and Towell and Voothees (1998) [34], rule-learning and neural networks have been applied, respectively.

Galley and McKeown (2003) proposed a method consisting of two stages [35]. First, a graph is built representing all possible senses of the words under investigation. The words are represented as graph nodes and their semantic relationships are the weights assigned to their connecting edges. The text is processed sequentially, comparing each word against all words previously read. If a relation exists between the senses of the current word and any possible sense of a previous word, a connection is established between the appropriate words and senses. The weight assigned to each connection is a function of two factors, i.e., the type of relationship and the distance between the words in the text.

For each sense of an ambiguous word, the weights of all connections involving that sense are summed, giving that sense a unified score. The sense with the highest unified score is finally chosen as the correct sense.

Mihalcea and Moldovan (2000), proposed an iterative method for WSD [36]. This method differs from other proposed methods in that it follows an iterative process for WSD (the same as the method proposed in this paper). This method requires two sources of information, namely WordNet and a semantic-tagged corpus. The method which will be proposed in the current paper is similar to the method proposed by Mihalcea and Moldovan (2000) in that it also disambiguates the words, iteratively. However, it does not need any additional information source. Instead, it works with a set of discovered association rules and knowledge deduced from the rules. Moreover, as will be shown via experiments, the response times of these two methods are quite different.

In 2011 Rezapour and et al [37] in his paper were proposed supervise learning method for WSD based on K-Nearest Neighbor algorithm. In first step, were extract two sets of features. The first set of words that a lot has occurred in the text and the next set of words are appears surrounding the ambiguous word. In order to improve the classification accuracy used also feature weighting strategy.

III. THE PROPOSED METHODS

The key task in WSD is to explore the context of the ambiguous word in order to find some unique characteristics of the ambiguous word. In this way, there exist some proposed approaches which are described as follows. In our proposed methods, we have made use of the following methods.

A. Using N-grams

An N-gram is a continuous sequence of N items, i.e., syllables and words, from a given text and can be considered as the composition of letters. The size 1 for the N is "unigram"; size 2 is a "bigram"; size 3 is a "trigram"; larger sizes are mentioned to by the value of N, e.g., "4-gram", "5-gram", and so on. Usually, N-gram is collected from a text or speech corpus.

According to our experimental results, the best results were obtained for $N = \pm 5$ where the gram separates five items from the left and the right sides of the ambiguous words in each paragraph. Note that for selection of N surrounding words, prepositions and numbers are ignored and then all of the selected words will be classified using the Naive Bayes algorithm.

B. Using Word Frequencies Instead of Words

In this method, using 5-grams, each word in the text is replaced with its frequency value. Then the Naive Bayes algorithm is applied to all numbers in order to classify them. Disadvantage of this method is that it is faced with a large scale sparse matrix.

C. Using Word Stems

In this method instead of using the word, the algorithm uses the word stem instead of the word. For instance, the term studying is replaced by its stem study.

The proposed methods for feature extraction are described as follows.

D. Assigning a Weight to Each Word

In the proposed method, each surrounding word found by 5-gram takes a weight. The weights are obtained via two approaches: first, according to the nearest and the furthest surrounding word and secondly according to the frequency of each word. It means that the nearest word takes the highest weight while the farthest one achieves the lowest weight. The frequency influences on the weighting approach where the furthest word, from ambiguous word, with high frequency takes a high weight. After the weighting phase, words are classified using the Naive Bayes classifier.

E. Using Keywords

Automatic extraction of keywords is evolved to identify a small set of words or key phrases from document such that the set can describe the concept of the document. This task should be performed automatically or with minimal human intervention [14].

Keywords are the words that have the highest frequency compared to the Brown corpus. Brown corpus is introduced by Kucera and Francis at Brown University in USA. This corpus includes one million words which is equal to 500 texts that each text includes 2000 words in a wide variety of topics in English.

For extraction of keywords, each arbitrary text is compared with the Brown corpus and the words with the highest frequency will be selected as the keywords of that text.

We take an example to reveal the problem. For instance, this algorithm outputs the 167.00 for the word step from a text with 92 words. For calculation of this value, the word step in Brown corpus and the arbitrary text has 130 (of one million words) and 2 (of 92 words) occurrences, respectively. Taken as a proportion of 1,000,000 words, these 2 occurrences represent 2/92 1,000,000 = 21739 virtual occurrences. These 21739 occurrences are 167.22 times more numerous than the 130 occurrences in Brown.

In other words, the occurrences of each word in an arbitrary text are evaluated according to the frequency of this word in the Brown corpus. For this purpose, the algorithm takes a paragraph of the text and extracts keywords of the paragraph and then the Naive Bayes algorithm creates a model for classification of the words.

F. Using depended word to the ambiguous word for WSD

WordNet is a lexical database in English language, that can be considered as a combination of dictionary and thesaurus. WordNet has been constructed in the Cognitive Science Laboratory at Peterson University in order to make semantic connections between the words.

In this method, we search some levels of the WordNet to discover ambiguous words. In fact, in WordNet there is different level of words which depended to ambiguous word. Then, the keywords of a paragraph and surrounding word of the ambiguous word are extracted. Finally, a model of all extracted words is created using the Naive Bayes for classification.

Algorithm steps are mentioned as follows.

1) Passing the ambiguous word into the WordNet.
2) Extraction of the words which are related to the target word, from different levels of the WordNet and then assigning a weight to each of them.
3) A paragraph is selected for the disambiguation. After that, the words extracted in the previous step are explored in this paragraph and then the founded words are added to the features of the paragraph.
4) The keywords of the paragraph are extracted using the algorithm described above and then, all of them are appended to features of that paragraph.
5) The 5-gram is employed to find 5 words before and after the ambiguous word. All of sounding words added into the features of that paragraph.
6) The step 1-5 is repeated for all paragraphs of the text. After that, a vector of features is obtained for each paragraph and finally, the correct meaning of the ambiguous word is used as the class of the paragraph.
7) Classification of all data using the Naive Bayes approach.
8) Finally, a new unseen paragraph is passed to the model created in the step 7, and the achieved results are compared with the real class of the data.

IV. Experimental Results

In order to evaluate our proposed schema, we use the TWA[1] (two-way ambiguities) dataset. TWA is a standard dataset for identification of the ambiguous words. This dataset is collected at the University of North Texas by Mihalcea and Yang. The dataset contains some ambiguous words and the main focus of the TWA is on six different words including ”bass”, ”crane”, ”motion”, ”palm”, ”plant” and ”tank”.

The data set is divided into two subsets, i.e., training and test (unseen) parts. Using 10-fold cross validation approach, the size of the training set in each iteration is 80% of the whole data and the remaining data is applied for test.

Tabel 1. Accuracy values of the Proposed Methods

Words	Method 1	Method 2	Method3	Method 4	Method 5	Method 6
Bass	51.85%	57.14%	52.73%	61.9%	75%	**85.49%**
Crane	63.15%	60.1%	65%	66.9%	80.43%	**87%**
Motion	47.6%	40%	49.1%	56.4%	66%	**74.32%**
Palm	68.12%	62.5%	70%	75%	81.25%	**84.21%**
Plant	53.57%	60.53%	55%	57.89%	62.75%	**67%**
Thank	60%	50%	62.12%	64.05%	65.71%	**71.03%**

In Figure 1 is demonstrated our proposed method for six word ambiguous.

[1] http://www.cse.unt.edu/rada/downloads.html#twa

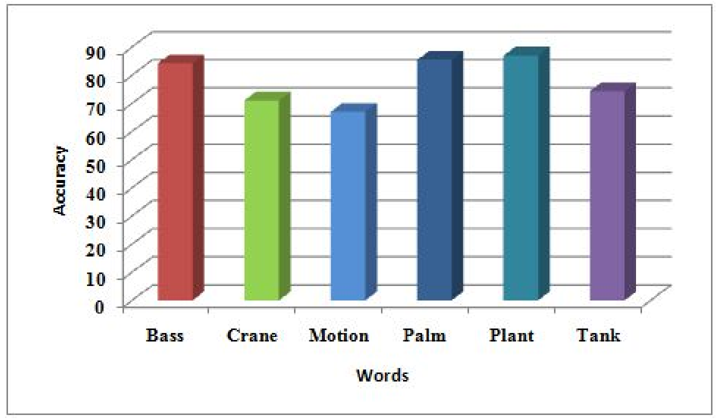

Fig 1. Accuracy of the proposed Method

If only WordNet is used, the below result obtained. However it’s not possible to predict all of the data therefore we need to use feature combination includes WordNet, keyword extraction and the words around the ambiguous word.

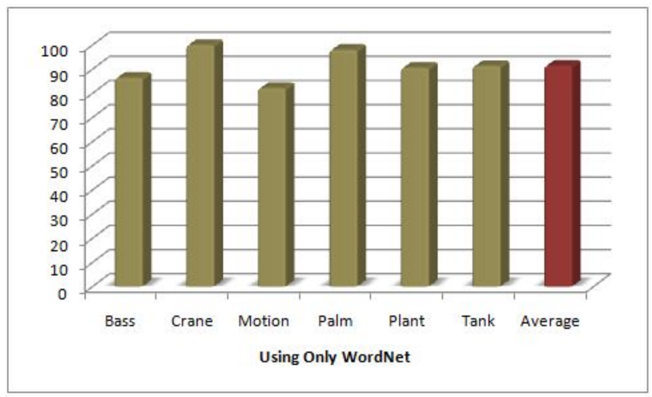

Fig 2. Using Only WordNet

Figure 3 is showed value of accuracy that we achieved 78.2% accuracy by perform this method to compared previous methods.

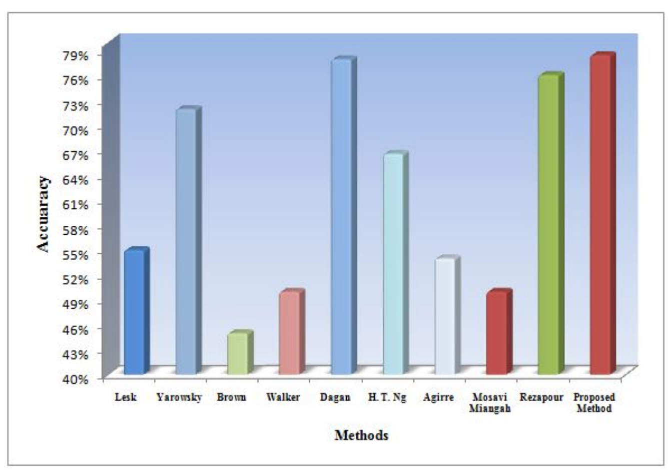

Fig 3. Accuracy Values Compared to Previous Methods

V. Conclusion

In this paper, we proposed and assessed several supervised methods to resolve the semantic and structural ambiguities of words in the text. One of the best results was obtained from the combination of WordNet, the keywords extraction algorithm, and surrounding words of an ambiguous word in the text. The combination is employed in order to find the features for disambiguation. Then, a model is created according to the extracted features and finally the Naive Bayes algorithm is applied to classify all data based on the model. In this paper, we used the TWA dataset as a benchmark for evaluating our method. The average accuracy of the proposed algorithm is 78.2%. The experimental results demonstrate that our algorithm is very encouraging.

VI. Future Works

One of the important factor of this study which can be considered as the future work is to determine the part of speech tagging of the word sense in the sentence, i.e. verb, noun, adj, adv and etc. and then this part of speech is used for word sense disambiguation.

VII. Refrences

[1] Daelemans, W., Zavrel, J., Berck, P., Gillis, S. (1996). MBT: A Memory-Based Part of Speech Tagger-Generator. In E. Ejerhed I. Dagan (Eds.), Proceedings of the Fourth Workshop on Very Large Corpora (pp. 14-27).

[2] Wilks, Y., Stevenson, M. (1998). Word Sense Disambiguation using Optimized Combination of Knowledge Sources. ACL 36COLING 17 (pp. 1398-1402).

[3] Kit, C., Wilks, Y. (1999). Unsupervised Learning of Word Boundary with Description Length Gain. Language, 1-6. Citeseer.

[4] De Pauw, G., Daelemans, W. (2000). The role of algorithm bias vs information source in learning algorithms for Morphosyntactic Disambiguation. Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning - Volume 7 (pp. 19-24). Stroudsburg, PA, USA: Association for Computational Linguistics. doi:10.3115/1117601.1117607

[5] Yarowsky, D. (1992). Word-sense disambiguation using statistical models of Rogets categories trained on large corpora. Proceedings of the 14th conference on Computational linguistics (Vol. 2, pp. 454-460). Association for Computational Linguistics.

[6] Lee, L. (2000). Foundations of Statistical Natural Language Processing Christopher D. Manning and Hinrich Schtze (Stanford University and Xerox Palo Alto Research Center) Cambridge, MA : The MIT Press , 1999 , xxxvii + 680 pp; hardbound, ISBN 0-262-13360-1 , 60.00. Computational Linguistics, 26(2), 277-279. MIT Press 238 Main St., Suite 500, Cambridge, MA 02142-1046 USA journals-info@mit.edu.

[7] Navigli, R. (2009) Word sense disambiguation: A survey. ACM Computing Survevs, 41, 2, Article 10.

[8] Gale, W.A., K. Church and D. Yarowsky (1992) A Method for Disambiguating Word Senses in a Large Corpus, Computers and Humanities, 26, 415-439.

[9] Yarowsky, D., (1994) Decision Lists for Lexical Ambiguity Resolution, Application to Accent Restoration in Spanish and French, In Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, 88-95.

[10] Ng, T. and H.B. Lee (1996) Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-based Approach, In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, 40-47.

[11] Mangu, L., and E. Brill (1997) Automatic rule acquisition for spelling correction. In Proceedings of the 14th International Conference on Machine Learning, 187-194.

[12] Golding, R. and D. Roth (1999) A Winnow-Based Approach to Context-Sensitive Spelling Correction, Machine Learning, 34, 107-130.

[13] Escudero, G., L. Mrquez and G. Rigau (2000) Boosting applied to word sense disambiguation. In proceedings of the 12th European Conference on Machine Learning (ECML), Barcelona, Spain, 129-141.

[14] MURATA, M., M. UTIYAMA, K. UCHIMOTO, Q. MA and H. ISAHARA (2001) Japanese word sense disambiguation using the simple Bayes and support vector machine methods, In Proceedings of the 2nd International Workshop on Evaluating Word Sense Disambiguation Systems (Senseval-2, Toulouse, France), 135138.

[15] Pedersen, T. (2000). A Simple Approach to Building Ensembles of Naive Bayesian Classifiers for Word Sense Disambiguation. Proceedings of the 1st North American chapter of the Association for Computational Linguistics conference, 7. Morgan Kaufmann Publishers Inc.

[16] Brown, P. F., S.A. DellaPietra, V.J. DellaPietra and R.L. Mercer (1991) Word sense disambiguation using statistical methods, In proceedings of Annual Meeting of the Association for Computational Linguistics, 264270.

[17] Yarowsky, D. (1992), Word Sense Disambiguation Using Statistical Models of Rogets Categories Trained on Large Corpora, In proceedings of 15th International Conference on Computational Linguistics, 454460.

[18] Dagan, I. and A. Itai (1994), Word sense disambiguation using a second language monolingual corpus, Association for Computational Linguistics, 20(4), 56396.

[19] Justeson, J. and S.M. Katz, (1995), Principled disambiguation: discriminating adjective senses with modified nouns, Computational Linguistics, 21(1), 128.

[20] Ng, T. and H.B. Lee (1996) Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-based Approach, In Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics, 40-47.

[21] Mosavi Miangah, T. and A. D. Khalafi (2005) Word Sense Disambiguation Using Target Language Corpus in a Machine Translation Systems, Literary and Linguistic Computing, 20 (2).

[22] Reddy, S. and A. Inumella (2010a) WSD as a Distributed Constraint Optimization Problem, In Proceedings of the ACL2010 Student Research Workshop, Uppsala, Sweden, 13-18.

[23] WIDDOWS, D., and B. DOROW (2002) A graph model for unsupervised lexical acquisition, In Proceedings of the 19th International Conference on Computational Linguistics, Taipei, Taiwan, 17.

[24] Veronis, J. (2004). HyperLex: lexical cartography for information retrieval. (J. Preiss M. Stevenson, Eds.)COMPUTER SPEECH AND LANGUAGE, 18(3), 223-252.

[25] SINHA, R. and R. MIHALCEA (2007) Unsupervised graph-based word sense disambiguation using measures of word semantic similarity, In Proceedings of the International Conference on Semantic Computing, Washington, DC, USA. IEEE Computer Society, 363369.

[26] Reddy, S., A. Inumella, D. McCarthy, M. Stevenson (2010b) IIITH: Domain Specific Word Sense Disambiguation, In proceedings of the 5th international workshop on semantic evaluation, Uppsala, Sweden, 387-391.

[27] SCHUTZE, H. (1998) Automatic word sense discrimination, Computational Linguistics, 24(1), 97124.

[28] Litkowski, K.C. (2000) Senseval: The CL research experience, Computers and the Humanities, 34(1,2), 153-158.

[29] LIN, D. (2000) Word sense disambiguation with a similarity based smoothed library, Computers and the Humanities: Special Issue on Senseval, 34, 147-152.

[30] Nigam, K., A.K. McCallum, S. Thrun and T. Mitchell (2000) Text Classification from Labeled and Unlabeled Documents using EM. Machine Learning, 39(2-3), 103134.

[31] Shinnou, H., M. Sasaki (2003) Unsupervised learning of word sense disambiguation rules by estimating an optimum iteration number in the EM algorithm, In Proceedings of the 7th conference on Natural language learning at HLT-NAACL, Edmonton, Canada, 41-48.

[32] Agirre, E., J. Atserias, L. Padr and G. Rigau (2000) Combining supervised and unsupervised lexical knowledge methods for word sense disambiguation, Computers and the Humanities, Special Issue on SensEval, 34 (1,2), 103-108.

[33] Yarowsky, D. (1995) Unsupervised word sense disambiguation rivaling supervised methods, In Meeting of the Association for Computational Linguistics, 189.196.

[34] Towell, G. and E. Voothees (1998) Disambiguating Highly Ambiguous Words, Computational Linguistics, 24 (1), 125-146.

[35] GALLEY, M. and K. MCKEOWN (2003) Improving word sense disambiguation in lexical chaining, In Proceedings of the 18th International Joint Conference on Artificial Intelligence, Acapulco, Mexico, 14861488.

[36] Mihalcea, R. and D. I. Moldovan (2000) An Iterative Approach to Word Sense Disambiguation, In: FLAIRS Conference, AAAI Press, 219–223.

[37] Rezapour, A. R., Fakhrahmad, S. M., Sadreddini, M. H. (2011). Applying Weighted KNN to Word Sense Disambiguation. Proceedings of the World Congress on Engineering (Vol. III).

[38] Hulth, A. (2003). Improved automatic keyword extraction given more linguistic knowledge. (M. Collins M. Steedman, Eds.)Proceedings of the 2003 conference on Empirical methods in natural language processing, 10(2000), 216-223. Association for Computational Linguistics.

[39] Fakhrahmad, S.M., M.H. Sadreddini and M. Zolghadri Jahromi (2007) Mining Frequent Itemsets in Large Data Warehouses: A Novel Approach Proposed for Sparse Data Sets, In Proceedings of the 8th International Conference on Intelligent Data Engineering and Automated Learning, Birmingham, U.K , 517-526.

[40] Abdul-Rauf, S. and H. Schwenk (2009) On the use of Comparable Corpora to improve SMT performance, In proceedings of the 12th Conference of the European Chapter of the ACL, Athens, Greece, 16-23.

[41] Agirre, E. and A. Soroa (2009) Personalizing pagerank for word sense disambiguation, In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics, Morristown, NJ, USA. Association for Computational Linguistics, 3341.

[42] Ahmed, F. and A. Nurnberger (2008) Arabic/English Word Translation Disambiguation using Parallel Corpora and Matching Schemes, In proceedings of 12th EAMT conference, Hamburg, Germany, 6-11.

[43] Brin, S., R. Motwani, C. Silverstein (1997) Beyond market baskets: generalizing association rules to correlations, In Proceedings of the 1997 ACM SIGMOD international conference on Management of data, Tucson, Arizona, United States, 265-276.

[44] Lagarda, A.L., V. Alabau, F. Casacuberta, R. Silva, E. Daz and D. Liao (2009) Statistical Post-Editing of a Rule-Based Machine Translation System, In proceedings of NAACL HLT 2009: Short Papers, 217220.

[45] LIN, D. (1998) Automatic retrieval and clustering of similar words, In Proceedings of the 17th International Conference on Computational linguistics, Canada, 768774.

[46] Mihalcea, R. and L. Yang (2003), TWA Sense Tagged Data, University of North Texas. WWW.cse.unt.edu/ rada/downloads.html.

[47] VERONIS, J. (2004) Hyperlex: Lexical cartography for information retrieval, Computer Speech and Language, 18 (3), 223252.