International Journal of Advances in Computer Science & Its Applications– IJCSIA
Volume 5: Issue 1 [ISSN : 2250-3765]
Publication Date : 30 April, 2015

Bokhari-WSD: Context Based Multimodal Word Sense Disambiguation

[Mohammad Ubaidullah Bokhari[1], Syed Hamid Hasan[2], Faraz Hasan[1]]

***Abstract*— Word Sense Disambiguation (WSD) is one of the core challenging area for researchers since several decades and it plays a crucial role in all natural language processing (NLP) applications viz. Information Retrieval, Information Extraction, Question Answering, Text Mining, Machine Translation etc. Researchers defined WSD as to identify the actual meaning of a word based on the context in which it occurs. Whereas in linguistic, context is defined as the text in which a word or passage appears and which helps ascertain its meaning. Hence, context of a word depends on different part of speech (POS) of a sentence i.e. Noun, Verb, pronoun, adjective and adverb. This paper proposes a novel approach for context based word sense disambiguation using soft sense disambiguation, map-reduce, knowledge based multimodal algorithm and WordNet.**

***Keywords*— Word Sense Disambiguation; Multimodal Word Sense Disambiguation, Soft Sense Disambiguation, Natural Language Processing; Computational Linguistic; Knowledge Based Measure; WordNet**

I. Introduction

Although WSD is a complicated task, it plays a crucial role in various NLP applications viz. Information Retrieval, Information Extraction, Question Answering, Text Mining, Machine Translation etc. [1][2][3][4]. Factually, sense is the core of intelligence whether it is natural or artificial. Therefore, WSD is the core artificial intelligence problem that plays a crucial role in all textual information systems.

WSD is defined as a task of resolving the ambiguity from a polysemous word to its appropriate sense within the context of a word [7]. Something is ambiguous when it can be perceive in various possible ways, in other words when it has more than one significant meaning. Ambiguity is broadly categorizes into two types i.e. structural ambiguity and lexical ambiguity. Ambiguity in a sentence or phrase is called structural ambiguity whereas ambiguity in a word is known as lexical ambiguity.

Lexical semantic ambiguity occurs when a single word is associated with multiple senses. In fact, almost every word has more than one meaning. For example, consider the noun "bank". In WordNet the word "bank" has 10 senses such as "river bank", "financial institution", "blood bank" etc. Humans can easily understand the context of the language whereas it's a very complicated task for the machines to understand.

Lexical disambiguation is defined as to identify the actual meaning of a word based on the context in which it occurs. Whereas in linguistic, context is defined as the text in which a word or passage appears and which helps ascertain its meaning. Hence, context of a word depends on different part of speech (POS) of a sentence i.e. Noun, Verb, pronoun, adjective and adverb. Although, most of the researchers uses only four POS (i.e. Noun, Verb, Adjective and Adverb) for resolving the disambiguation in context based WSD. This paper focuses on lexical disambiguation that uses all the five POS, as this is the crucial issue for most of the applications.

For Lexical semantic ambiguity, various knowledge-based measures were proposed viz. Resnik measure [8], Lesk measure [9], Lin measure [10], Wu & Palmer measure [11] etc. However, combination of these measures performs better than individual measures [12].

The paper is organized as follows. Section II provides an overview about various approaches for word sense disambiguation. In section III, knowledge based WSD measures were discussed followed by types of WSD tasks. Next section, is about a proposed Map-Reduce based WSD Framework. Section VI describes the proposed Bokhari-WSD model. Finally, conclude the paper.

II. Approaches For Word Sense Disambiguation

WSD measures were broadly divided into three categories i.e. unsupervised, supervised and Knowledge based measures based on their level of knowledge requisite [5][6].

A. *Unsupervised WSD*

Unsupervised methods are designed for unlabeled data, and do not use any manually sense-tagged corpus to provide a sense choice for a word in context. These methods do not require any training by the expert of a domain, and hence have a wider scope for implementation. But, due to the lack of knowledge source these methods generally lacks in accuracy.

B. *Supervised WSD*

Supervised approaches requires a training to learn a classifier from labeled corpora, that is, prior to actual

1
Department of Computer Science, Aligarh Muslim University, Aligarh INDIA

2
Faculty of Computing and Information Technology,
Department of Information Systems,
King Abdulaziz University, Jeddah, Saudi Arabia

disambiguation process a number of features were encoded together with their appropriate sense label. These methods usually perform better than other methods but requirement for training and manually tagged data restricted their scope for implementation.

C. Knowledge Base WSD

Knowledge based approaches require a generic knowledge source or dictionary like WordNet for the disambiguation process. These methods are the intermediary methods between unsupervised and supervised methods in terms of knowledge requisite. Hence, these methods were attaining a wider reach together with intermediary level of accuracy.

III. Knowledge Based WSD Measures

In order to quantify the relatedness of two words to which extent they are semantically related using knowledge based semantics, there are various measures that were presented [4].

While selecting a measure, we concentrate on those which has better accuracy on word sense disambiguation as well as those that uses the common subsumer, gloss overlap and multimodal knowledge based measures. Consequently, the six measures explained below were selected. All of these measures presume as input a pair of concepts, and return a value of their semantic relatedness.

A. Resnik Measure

Resnik [8] describes a measure for semantic relatedness between two concepts based on their lowest common subsumer (LCS) that evaluates the information content (IC) of two concepts using the formula:

$$\mathrm{Sim}\ (C_1, C_2) = -\log\ (P\ (IC\ (LCS\ (C_1, C_2)))) \qquad (1)$$

Where P represents the probability, IC is information content and LCS represents lowest common subsumer.

B. Jiang-Conrath Measure

Jiang and Conrath's [13] is a variant of Resnik's definition that evaluates the difference in the information content instead of calculating the probability of the two concepts to indicate their similarity.

$$\mathrm{Sim}(C1,C2)=2-IC(LCS(C1,C2))-(IC(C1)+IC(C2)$$

C. Lin Measure

Lin [10] states that the similarity between two concepts is based on the ratio of information content of their least common subsumer to the information content of individual concepts. This measure is a close variant of the Jiang–Conrath measure.

$$\mathrm{Sim}(C_1,C_2)=(2-IC(LCS(C_1,C_2)))/(IC(C_1)+IC(C_2))$$

D. Lesk Measure

Lesk [9] proposed a measure of semantic relatedness between two concepts that defined as the measure of word overlap between the definitions or glosses of two concepts, as provided by the knowledge base. The benefit of the Lesk similarity measure is that it can be used with any knowledge base or dictionary that provides word definitions. Lesk measure performs well for verb, adverb and adjectives.

E. Wu and Palmer Measure

Wu-Palmer [11] states that the similarity between two concepts depends on the closeness between them in the hierarchy of WordNet, i.e., the similarity between two concepts is measured as:

$$\mathrm{Sim}\ (C1, C2) = (2*M3)/(M1+M2+2*M3)$$

Where M1 is the number of nodes between C1 and C3 (least common super concept of C1 and C2), similarly M2 represents the number of nodes between C2 and C3, and M3 is the number of nodes between C3 and the root of the concept hierarchy. Wu and Palmer perform well for noun and verb POS.

F. Combination of Similarity Measures

Sinha and Mihalcea [12] implemented a multimodal similarity measure, which combines the Jiang–Conrath and Lesk measures for getting the benefits of each individual metric. They perform a graph-based similarity, where jcn is used to draw semantic network between nouns and the similarity metric lch is used to draw similarity between verbs. Whereas, lesk measure is used to draw the semantic network for other parts of speech.

G. JIGSAW

JIGSAW [15] is a multimodal WSD algorithm that integrates three modes of disambiguation viz. JIGSAW_{noun}, JIGSAW_{verb} and JIGSAW_{others} for different POS provided by WordNet. JIGSAW_{noun} uses Leacock-Chodorow [16] measure and JIGSAW_{others} uses Adapted Lesk algorithm [17]. Multilevel disambiguation procedures were used to disambiguate word senses. First, JIGSAW_{noun} disambiguate the word senses. Next, these disambiguated word senses were used by the JIGSAW_{verb} to further disambiguate the verbs. Finally, JIGSAW_{others} disambiguate the other POS.

International Journal of Advances in Computer Science & Its Applications– IJCSIA
Volume 5: Issue 1 [ISSN : 2250-3765]
Publication Date : 30 April, 2015

IV. Types of WSD Task

Generally, there are two types of task related with the WSD either disambiguate the selected words in a sentence or disambiguate all words that are polysemous in nature.

A. *Targeted WSD*

Targeted WSD is a process of disambiguating a selected set of target words occurring in a sentence. Typically, supervised systems performs better in this task, as this task have a restricted or closed set of ambiguous words that becomes easy to train a system using a number of manually tagged training set.

B. *All Words WSD*

This is a task of disambiguating all polysemous POS words in a sentence and therefore requires more system resources. Also, systems in this setting consume more execution time than targeted WSD.

V. Proposed WSD Framework based on Map-Reduce

The basic idea behind implementation of a multimodal similarity measure is that the "benefits" of various measures integrates into one model, which then enhances the performance of word sense disambiguation.

Sense of a target word depends on similarity between the synsets of two words and the context of a word. Whereas, the supporting or other POS words in a sentence decides the context of a word. Also, processing time plays a crucial role for all information systems. Keeping both issues in mind, Map-Reduce and context based WSD Framework is proposed.

Fig.1 illustrate the flow of proposed framework.

A. *Preprocessing*

Preprocessing in all natural language processing (NLP) applications generally consists of five steps i.e. tokenization, part-of-speech tagging, lemmatization, chunking and parsing.

1) Tokenization

Tokenization is the process of splitting the text into set of tokens. Tokens are generally set of words used in a sentence.

2) Part-of-speech tagging

POS tagging is the process of assigning a grammatical category to each token or word according to its lexical appearance in the sentence.

3) Lemmatization

The process of reducing a word in its base form is known as lemmatization.

4) Chunking

It consists of partitioning of a text according to syntactically correlated parts (e.g., the noun phrase and the verb phrase).

5) Parsing

It is the process of generating a syntactic structure of a sentence also known as parse tree.

B. *SenseMapper*

This module is responsible for parallel distribution of each

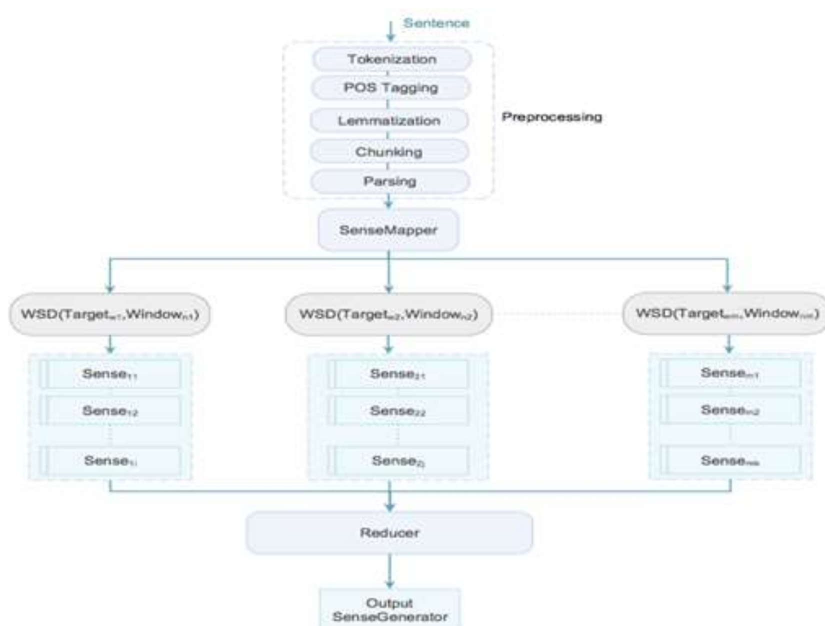

Figure 1. Map-Reduce Based Word Sense Disambiguation Framework

targeted word with their supporting list of words for the sake of finding the actual sense of a word instead of finding the max weighted sense. In other words, in this module the key-value pairs have been created containing target word as a key and list of window words as a value. These maps are further executed in a hadoop map-reduce framework in an independent manner for finding the actual sense by the Bokhari-WSD module.

C. WSD Module

This module takes the key-value pair as input and retrieves all the senses for the target word from the WordNet and calculates the similarity for all the senses of target word. The generated similarity vectors for each targeted words further accumulated by the reducer module for taking collective decision.

D. Reducer

The module receives sense similarity vector for each targeted word from Bokhari-WSD module and assign most appropriate sense to the target word according to its maximum similarity. In case, there is any collision between two senses then the module resolves that collision by providing higher priority to the most frequent used sense according to the WordNet structure. Finally, the module performs sense tagging to their respective target words.

VI. Proposed Bokhari-WSD Model

The proposed model is a multimodal approach, which combines multiple similarity measures according to their POS. The main idea is different POS words have different perceptional view or in other words different POS words affects the context of a word in different manner. Hence, a proposed approach uses different approaches for different POS. Also, the context of a word decides in which sense the word is in use. Therefore, proposed model uses a soft sense disambiguation technique to disambiguate a target word using all context words of POS noun, verb, adverb and adjective. Besides these, sometimes pronoun changes the sense of a word in a sentence. E.g.

1. What a cat?
2. What a cat she is.

In above said sentences, the sense of word “cat” changes just due to the presence of pronoun “she”. Therefore, in proposed model pronouns were also used to disambiguation process. Keeping in mind that pronoun only have a direct relation with nouns, proposed algorithm computes the similarity of a pronoun only with the noun POS. Moreover, in proposed model pronouns were replaced by its noun variant like “he” replaced by “man” and “she” replaced by “women” etc. Working of algorithm for different POS are described below.

A. Similarity calculation for both Noun

The algorithm extracts all the senses containing noun synsets of word from the WordNet and computes the similarity between target word senses and the context word senses using the equation 4. The algorithm differs from the original Wu & Palmer approach in the use of the similarity measure. Instead of using the maximum weighted sense, our approach produces n×1 stochastic similarity matrix where n is the number of target word senses. Wu & Palmer measure generates n×m similarity matrix that further reduces into n×1 similarity matrix by taking the maximum of each row.

B. Similarity Calculation for Noun and Verb

The algorithm extracts all the senses containing noun synsets for the word having noun POS and all verb synsets for the word having verb POS from the WordNet and computes the similarity between target word senses and the context word senses using the concept of Wu & Palmer approach and $\text{JIGSAW}_{\text{verb}}$ [15] separately. The algorithm first convert these similarity matrices into two n×1 matrices by taking the maximum of each row and then fuse them into one n×1 stochastic matrix by taking their mean and then by dividing each element of a vector by its column sum. Similarly, process the verb and noun combination.

C. Similarity Calculation for Noun and Other

The algorithm extracts all the senses containing noun synsets for the word having noun POS and all synsets for the word having respective POS from the WordNet and computes the similarity between target word senses and the context word senses using the Lesk gloss overlap and $\text{JIGSAW}_{\text{verb}}$ approach separately. These approaches generate matrices of order n×m that further converted into matrices of order n×1 by taking the maximum of each row. In last, these matrices fused into one matrix of order n×1 by taking the maximum of each sense from the two matrices.

D. Similarity Calculation for others

The algorithm extracts all the senses for the first word according to its POS from the WordNet and computes the similarity between target word senses and the context word senses using the concept of Lesk gloss overlap. Firstly, a similarity matrix of order n×m converted into order of n×1 and then an n×1 stochastic similarity matrix would be generated.

E. Soft Sense Disambiguation

This module recieves all the similarity vectors of order n×1 from similarity calculators of different POS and compute a fuzzy membership score for each target word sense according to the given membership function.

$$MF = \frac{1}{m} \times \sum_{i=0}^{m} S_i$$

where m is the number of context words and Si is the similarity score for i^{th} sense received from similarity calculators.

Conclusion & Future Work

In this paper we discussed about various lexical semantic similarity measures for word sense disambiguation that provides a along with a proposed framework based on map-reduce for enhancing the efficiency and minimize execution time. Also, a multimodal context based WSD model was proposed that exploits pronoun for disambiguating noun POS. Soft sense disambiguation technique was used to choose the appropriate sense of a word with in the context of a target word. As a future work, we plan to perform experiments on standard datasets using WordNet database to compare our proposed model with some of the baseline WSD methods.

References

[1] E. Agirre, and P. G. Edmonds, “Word sense disambiguation: Algorithms and applications,” Springer, Vol. 33, 2007.

[2] E. Altintas, E. Karsligil, and V. Coskun, "A new semantic similarity measure evaluated in word sense disambiguation," In Proceedings of 15th NODALIBA conference, Joensuu, 2005.

[3] S. Banerjee, and T. Pedersen, "An adapted Lesk algorithm for word sense disambiguation using WordNet," Computational linguistics and intelligent text processing, Springer Berlin Heidelberg, pp. 136-145, 2002.

[4] B. Alexander, and G. Hirst, "Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures," Workshop on WordNet and Other Lexical Resources, Vol. 2, 2001.

[5] J. Sreedhar, S. V. Raju, A. V. Babu, A. Shaik, and P. P. Kumar, "Word Sense Disambiguation: An Empirical Survey," International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, 2012.

[6] K. Kaye, and W. T. Aung, "Word Sense Disambiguation: A Briefly Survey," International Journal of Computer & Communication Engineering Research (IJCCER), Vol. 1, No. 4, pp. 118-123, 2013.

[7] S. Torres, and A. Gelbukh, "Comparing similarity measures for original WSD lesk algorithm," Research in Computing Science, Citeseer, No. 43, pp. 155-166, 2009.

[8] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," arXiv preprint cmp-lg/9511007, 1995.

[9] M. Lesk, “Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from a ice cream cone,” In Proceedings of SIGDOC ’86, 1986.

[10] Lin, “An information-theoretic definition of similarity,” In Proceedings of the 15th International Conference on Machine Learning, Madison, WI, 1998.

[11] W. Zhibiao, and M. Palmer, "Verbs semantics and lexical selection," In Proceedings of the 32nd annual meeting on Association for Computational Linguistics, Association for Computational Linguistics, 1994.

[12] R. Sinha, and R. Mihalcea, “Unsupervised Graph-based Word Sense Disambiguation Using Measures of Word Semantic Similarity,” In Proceedings of the IEEE International Conference on Semantic Computing (ICSC), Irvine, CA, September 2007.

[13] J. Jiang, and D. Conrath, “Semantic similarity based on corpus statistics and lexical taxonomy,” In Proceedings of the International Conference on Research in Computational Linguistics, Taiwan, 1997.

[14] Navigli, Roberto. "Word sense disambiguation: A survey." *ACM Computing Surveys (CSUR)* 41.2 (2009): 10.

[15] Semeraro, Giovanni, et al. "Combining Learning and Word Sense Disambiguation for Intelligent User Profiling." *IJCAI*. Vol. 7. 2007.

[16] C. Leacock and M. Chodorow. Combining Local Context and WordNet Similarity for Word Sense Identification. In C. Fellbaum, editor, *WordNet: An Electronic Lexical Database*, pages 266–283. MIT Press, 1998.

[17] S. Banerjee and T. Pedersen. An Adapted Lesk Algorithm for Word Sense Disambigua- tion using WordNet. In *CICLing ’02: Proceedings of the Third International Conference on Computational Lin- guistics and Intelligent Text Processing*, pages 136–145, London, UK, 2002. Springer-Verlag.

About Author (s):

Dr. Mohammad Ubaidullah Bokhari is currently working as a Professor and Chairman, Department of Computer Science, AMU, Aligarh. He has published more than 100 research papers in different reputed journals and conference proceedings. He has also authored 5 books on different fields of Computer Science. His current research interests are Requirement Engineering, cryptography, Software Reliability, Wireless Network Security, information retrieval, soft computing and Database.

Dr. Syed Hamid Hasan has been working as Professor in Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia. He obtained his PhD degree in Computer Science and has a teaching and research experience of more than thirty years. His current research interests include Information Security, Cryptography and e-learning. Dr. Hasan is member of IEEE and has published many research papers.

Faraz Hasan has been working as a research scholar in Department of Computer Science, Aligarh Muslim University, Aligarh, INDIA. He obtained his Masters degree in Computer Science and has an industry experience of four years. His current research interests include Multimodal Information Retrieval, Word Sense Disambiguation, Soft Computing and Unsupervised Learning Techniques. Mr. Hasan is member of various reputed journals and has published many research papers.