Word Sense Disambiguation – A Context based Multimode WSD Model

Syed Hamid Hasan
Department of Information Systems, Faculty of Computing & Information Technolgy
King Abdulaziz University-Jeddah
Saudi Arabia

Shaji K.A. Theodore
Department of Information Technology
Musanna College of Technology
Sultanate of Oman

Faraz Hasan
Department of Computer Science
Aligarh Muslim University
Aligarh-India

ABSTRACT
For many decades researchers in the domain of NLP (Natural Language Processing) and its applications like Machine Translation, Text Mining, Question Answering, Information Extraction and Information retrieval etc. have been posed with a challenging area of research i.e. WSD (Word Sense Disambiguation) WSD can be defined as the ability to correctly ascertain the meaning of a word, with reference to the context in which the word was used. Linguistics, has defined context as the passage, sentence or text in which the word appears that is used for ascertaining its meaning. Thus, context is dependent on the POS (Part Of Speech) where the word is used e.g. Adverb, Adjective, Pronoun, Verb and Noun. In the following study we recommend a unique way of WSD based on Context through WordNet, multimodal algorithm that is knowledge based, map-reduce and soft sense WSD.

Keywords
WordNet, Knowledge Based Measures, Computational Linguistic, NLP; Soft Sense WSD, Multimodal WSD

1. INTRODUCTION
Various NLP application like Machine Translation, Text Mining, Question Answering, Information Extraction and Information retrieval etc. are critically supported by WSD. Even though WSD is quite a complex task. [1,2,3,4]. In fact 'sense' is the essence of intelligence be it artificial or natural. Thus, all the major textual information system are crucially challenged by the issue of WSD.

In [7] we find WSD is described as the activity of clearing ambiguity around a polysemous word with reference to the word's context in order to reveal its appropriate sense/meaning. We can term a thing to be ambiguous if there are various ways of perceiving it, or better described asa situation where something that has multiple significant meanings. We can classify Ambiguity into two broad categories: Lexical and Structural. When a word is ambiguous it's called Lexical Ambiguity and when a phrase of sentence is ambiguous it's called Structural Ambiguity.

When a word has multiple significant senses associated with it we term it as Lexical semantic ambiguity. To be honest very few words are free from Lexical semantic ambiguity, and have a single meaning associated with it. E.g. "Bank" which is a noun and WordNet presents to us as many as 10 meanings like "blood bank", "financial institution" and "river bank" etc. It is fairly easy for a human to understand the sense of words

in their contexts, while the same simple task of understanding is very complex for a machines.

Identifying the appropriate meaning of a word, based on the context in which it is used, is called Lexical disambiguation.. Linguistics, has defined context as the passage, sentence or text in which the word appears that is used for ascertaining its meaning. Thus, context is dependent on the POS (Part Of Speech) where the word is used e.g. Adverb, Adjective, Pronoun, Verb and Noun. Even though we find that most researchers consider usage on only 4 POS (Adverb, Adjective, Verb and Noun) in order to resolve disambiguation while modeling a contextual WSD model. We would focus on using all the 5 POSs, since we consider failure to do so is proves crucial for all such applications. This paper focuses on lexical disambiguation that uses all the five POS, as this is the crucial issue for most of the applications.<Done>

We find that a number of measures that are knowledge based were suggested to tackle "Lexical Semantic Ambiguity", e.g. Wu & Palmer [8], Lin measure [9], Lesk measure [10], and Resnik measure [11] etc. However, it is observed that a combination of various measures works better than any single measure alone [12].

We have the following sections in this paper. Section II discusses the current day approaches available for WSD. Section III, discusses the different WSD measures that are knowledge based, along with the different kinds of WSD task in Section IV. Section V describes the suggested WSD framework that is "Map-Reduce" based. Whereas, in Section VI we present the Context based Multimodal WSD Model. Post which we conclude the paper in last Section.

2. WSD APPROACHES
The various measures and approaches on WSD can be majorly categorized as Knowledge Based, Supervised and Unsupervised on the basis of the knowledge level required for their application [5,6].

2.1 Knowledge based WSD
This measure has pre-requisite of a dictionary or knowledge source like WordNet etc. to carry out the disambiguation. The level of knowledge required in conducting these methods make them to be termed as mediatory method between the supervised and unsupervised methods. Thus, these methods are gaining wider popularity with their accuracy level being intermediary.

2.2 Supervised WSD

Despite being pretty accurate in the results, these methods are not so widely used owing to the level of prerequisite knowledge and training required for application of these methods. The training enables classification based on labeled corpora even before the actual process of disambiguation is done. The measures have different feature encoded along with appropriate sense labels to be used for classification.

2.3 Unsupervised WSD

These measures works best when the data is unlabeled with complete absence of "sense tagged" repository for enabling sense choices for the words in their context. As evident, these measures have no knowledge and training prerequisites thus widely utilized, yet devoid if credible knowledge source the accuracy can't be depended upon.

3. WSD MEASURES THAT ARE KNOWLEDGE BASED

In [4] a number of measures were presented for quantifying the relationship of 2 words and the extent to which these words are semantically related to each other.

The selection of the measure is based on WSD accuracy along with usage of Gloss Overlap, common subsume and which were based on multimodal knowledge. Based on the above criterion we found the 6 measures described below. They have the commonality of presuming concept pairs as input and outputting the level of semantic relatedness.

3.1 Wu and Palmer

According to Wu-Palmer at [11], proximity in hierarchy of two concepts in WordNet defines their level of relatedness; i.e. measurement of similarity amongst 2 concepts is derived from :

$$Sim\,(C_1, C_2) = (2*M3)/(M1+M2+2*M3)$$

Here M1 intermediatory node count between C3(which is LCS-'Least Common Super Concept' of C2 & C1) and C1. M2 is node count between C3 & C2, while M3 is node count between concept hierarchy root and C3. This measure worked well for Verb and Noun POS.

3.2 Lesk

In [9] Lesk suggested measuring of semantic similarity amongst 2 concepts which provided the level of words overlap between the glosses or definition of these 2 concepts, that was defined by the knowledge base. Advantage offered by this measure is that any dictionary or knowledge base that provide word definition can be used for this measure. Adjectives, Adverbs and Verb POS work well with this measure.

3.3 Lin

According to Lin [10] the ratio of "Information contents of the individual concepts" TO "Information content of the LCS (Least Common Subsumer) of these individual concepts" decide the level of relatedness of these 2 concepts. We find a close link between Jiang–Conrath & Lin's measures.

$$Sim(C_1,C_2)=(2-IC(LCS(C_1,C_2)))/(IC(C_1)+IC(C_2))$$

3.4 Jiang-Conrath

In [13] we find Jiang and Conrath's measure to be a Resnik measure variant as it focuses on evaluating the 'Dissimilarity' of the IC of 2 concepts in place of finding the relatedness of these concepts in order to ascertain their similarity.

$$Sim(C1,C2)=2-IC(LCS(C1,C2))-(IC(C1)+IC(C2)$$

3.5 Resnik

We find Resnik [8] describing the measure of semantic similarity amongst 2 concepts through their LCS, which ascertains the IC of the 2 concept via the following calculations:

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

Here, probability is denoted by P.

3.6 JIGSAW

In [15] we find WSD algorithm that is multimodal, called JIGSAW that integrate the 3 disambiguation modes i.e JIGSAWothers, JIGSAWverb and JIGSAWnoun, for various WordNet provided POS. Adapted Lesk algorithms is used by JIGSAWothers [2] while Leacock-Chodorow measure [16] is used by JIGSAWnoun. WSD was achieved through processes of Multilevel disambiguation. The word sense is first disambiguated by JIGSAWnoun, next verb disambiguation is done by JIGSAWverb on these disambiguated words which were finally worked upon by JIGSAWothers for other POS disambiguation.

3.7 Similarity Measures - Combined

In [12] we find a similarity measure that is multimodal implemented by Sinha & Mihalcea, it combines the Lesk and Jiang–Conrath's measures in order to achieve advantages presented by each one of them. It performs a similarity measure that is graph-based. Here, verb similarity is drawn by LCH metric while JCN draws semantic network amongst noun. Semantic Network of other POS is drawn by the Lesk measure.

4. WSD TASKS TYPE

In General, we find two task types in relation to WSD which either disambiguate all polysemous words or any two selected words present in the specific sentence that is being worked upon.

4.1 All Words

In this WSD task type the system disambiguates all POS polysemous words that are present in the target sentence, the time and resources required for the completion of this task is quite a lot in comparison to the other task type.

4.2 Targeted

In this process or WSD task, the target is a selected set of words in the specific sentence being worked upon. We find that performance of Supervised systems is better in this sort of a process owing to the closed set or restricted ambiguous target words, since it is easier training a system to disambiguate a finite training set of words.

5. PROPOSED – 'MAP-REDUCE' WSD FRAMEWORK

The The aim for implementing, a similarity model that is, multimodal is to avail the advantages of all the participating measures all rolled up into one. This approach gives as an enhanced WSD.

The target word's sense is dependent upon word's context and Synsets' similarity of the two target words. Additionally the context of the word is decided by the other POS or supporting words. We also cannot ignore the role played by the processing time taken by the system. We need to pay attention to both the issues. Thus we proposed a 'Map-reduce' WSD framework that is based on context. We have described the proposed framework workflow in Figure.1

5.1 Preprocessing

All NLP apps have the following commonly found steps in preprocessing:

5.1.1 Tokenization

This involves the text being split into set of words called Tokens that are further processed upon. A sentence or text can be split into multiple tokens.

5.1.2 POS Tagging

Parts Of Speech Tagging means to assign each word or token a grammatical category based on the lexical appearance of that token or word in the target sentence or text.

5.1.3 Lemmatization

Here every word is reduced to its base form and it is thus called lemmatization. Where Lemma means the base or dictionary form of the word.

5.1.4 Chunking

Chunking means to partition the text or words into syntactical groups or correlated parts based on syntax (e.g., verb phrases or the noun phrases etc).

5.1.5 Parsing

Here a Parse Tree is developed which is actually the structure of the target sentences based on syntax.

5.2 SenseMapper

The Sense Mapper is aimed at parallely distributing each targeted word with its supportive lists of words so that instead of maximum weighted senses of the word, the actual sense of the target word is found. Alternatively described as a module where a Key-Value pair is formed that contains the targeted words as Key and the supportive list of words as values. The mapping is additionally executed through a 'Map-reduce' hadoop framework independently to find the actual sense of these maps by the Proposed WSD Module.

5.3 WSD Module

The Key-Value pairs are taken as input to this module for retrieving all possible senses that the target words could have, which are then compared with the available sense for the target words in WordNet for calculating the similarity. The resultant vectors of similarity of each word targeted, is again accumulated through the Reducer Module that enables collective decisions.

5.4 Reducer Module

This module is responsible for receiving the sense similarity vectors generated by the Proposed WSD module and then assigning the target words the most suitable sense based on the maximum similarity. If there happens to be clash, then the WordNet structure is used for assigning priority to the most commonly and frequently used sense of the targeted words. In the end the target words are tagged with their respective sense tags.

6. THE PROPOSED MULTIMODAL MODEL

We employ a multimodal approach in the proposed WSD model, where various similarity measures are combined based on the respective POS. The core concept is that each POS word has a unique perceptual view, and each POS word affects contexts in a unique fashion. Thus, in the proposed model we suggest using different measures for different Part-Of-Speech. Additionally, the sense in which a word is used is decided by its context. Hence, 'Soft Sense' disambiguation techniques are used by the proposed model for disambiguating the targeted words through usage of all context word of the Part-Of-Speech like Adjective, adverb, verb and noun. Alternatively, the sense of words in the sentences are changes by pronouns occasionally. For example:

    i.   What a dog?

    ii.   What a dog he is.

This example illustrates the change in the sense of "dog" occurs only because of the pronoun "he". Thus, pronouns found a place in the disambiguation process in the proposed model. Since, nouns have direct relations with pronoun only, in the proposed measure similarity of the pronoun is only calculated with noun Part-Of-Speech. Additionally, the model replaces the pronouns with its related nouns like 'Woman replaces 'She' and 'man' replaces 'he'. The algorithms used for working with different POS are as follows:

6.1 Noun & Noun Similarity calculations

All possible senses that contain the noun synset of the target words are extracted from WorNet by the algorithm, after which equation A (Wu & Palmer) is utilized for computing the similarity between context word sense and target word sense. There is a difference of approach in similarity measure usage in the proposed model with the Wu & Palmer model.. The proposed model generates nx1 stochastic similarity matrices, where 'n' denotes count of all possible word senses in place of the standard maximum weighted senses. Whereas Wu & Palmer model produces nxm similarity matrices which are further reduced into nx1 similarity matrices through maximum of every row being taken.

6.2 Noun & Verb - Similarity Calculations

All possible senses that contain the Verb synset of the targeted word with verb POS and noun synset of the target words with noun POS are extracted from WorNet by the algorithm. Similarity between context word senses and target word sense are calculated separately using the JIGSAWverb and Wu & Palmer model equations. The similarity matrices are first converted into 2 nx1 matrix through maximum of every row being taken and then both of them are fused into a single nx1 stochastic matrix through the mean being taken and further division of the element's every vector by the sum of the respective column. Noun and Verb combinations are processed in the same way.

6.3 Noun and Other - Similarity Calculations

All possible senses that contain the Noun synset of the targeted word with noun POS and all synset of the target words with respective POS are extracted from WorNet by the algorithm. Similarity between context word senses and target word sense are calculated separately using the JIGSAWverb and 'Lesk' Gloss overlap model equations. Matrix of nxm order are generated by these methods which are reduced to

nx1 order through maximum of every row being taken. Finally all the matrix are combined to yield an nx1 order matrix where maximum of every senses of the 2 matrix is taken.

6.4 Others - Similarity Calculations

All possible senses that contain the first words synset of the targeted word with its respective POS are extracted from WorNet by the algorithm. Similarity between context word senses and target word sense are calculated using the 'Lesk' Gloss overlap model equations. Initially similarity matrices of nxm order are converted into nx1 order post which a single stochastic similarity matrix of nx1 order is produced.

6.5 "Soft Sense" Disambiguation

All similarity vector which are of nx1 order are received by the module from different POS similarity calculators. After which fuzzy membership scores are computed for every targeted word sense based on the assigned membership function.

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

Here context word count is m and Si denotes similarity scores of the ith sense that the similarity calculators provides.

7. CONCLUSION

This paper discuss the number of lexically semantic similarity models for WSD and proposes framework that has its basis on "Map-reduce" so that it minimizes time of execution and enhances efficiency. It also proposes a Multimodal WSD model that is context based which utilizes pronouns for disambiguation of noun POS. Suitable word sense was chosen through the technique of "Soft sense" disambiguation, that within the target word's context.

8. REFERENCES

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

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

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

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

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

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

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

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

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

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

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

[12] Sinha, R. and Mihalcea, R. 2007. 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.

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

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

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

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