Graph Based Algorithms for Word Sense Induction and Disambiguation

Neha R. Kasture
Department of Computer Science and Engineering
Shri Ramdeobaba Kamla Nehru Engineering College
Nagpur, India
nehakasture86@yahoo.co.in

Avinash Agrawal
Department of Computer Science and Engineering
Shri Ramdeobaba Kamla Nehru Engineering College
Nagpur, India
avinashjagrawal@gmail.com

Abstract— **This paper presents a survey of graph based methods for word sense induction and disambiguation. Many areas of Natural Language Processing like Word Sense Disambiguation (WSD), text summarization, keyword extraction make use of Graph based methods. The very idea behind graph based approach is to formulate the problems in graph setting and apply clustering to obtain a set of clusters (senses). The basic aim of this paper is to study various aspects of such graph based approaches in disambiguation of words. The paper also provides an insight into the results obtained by these techniques on standardized evaluation systems.**

Keywords-word sense disambiguation; graph-based;word sense induction; unsupervised methods;

I. INTRODUCTION (HEADING 1)

Many words in natural languages have multiple possible meanings. “Words” have different meanings based on the context of word usage in the sentence. Word Sense Disambiguation (WSD) is the process of determining the correct sense of the word in the given context. For example the word *Can*, can be used as a model verb: You *can* do it, or as a container: She brought a *can* of soda. There exists several examples of such words where the correct sense of the word with multiple meanings is obvious to a human, but developing algorithms to replicate this human ability is a very tough job. Word Sense Disambiguation is very important for a variety of natural language processing tasks: machine translation, information retrieval, grammatical analysis, speech and text processing as given in [1]. WSD techniques are broadly classified into four categories as proposed in [18].

1) *Dictionary and knowledge based methods*: These methods make use of lexical knowledge bases such as dictionaries and thesauri, and hypothesize that context knowledge can be extracted from definitions of the words. For example, Lesk disambiguated two words by finding the pair of senses with the greatest word overlap in their dictionary definitions [4].
2) *Supervised methods*: These methods make use of context to disambiguate the words. Supervised method includes a training phase and a testing phase. In training phase a sense annotated corpus is required, from which syntactic and semantic features are used to create a classifier. In the testing phase the word is classified into senses. These methods often suffer due to data scarcity problem and it is equally hard to acquire sufficient contextual information about senses of large number of words in natural languages.
3) *Semi-supervised methods*: These methods make use of small annotated corpus as seed data in bootstrapping process as proposed in [2].
4) *Unsupervised methods*: These methods gain contextual information directly from un-annotated raw text, and induce the senses from text using some similarity measure. But, it may also be the case that automatically acquired information is noisy or erroneous.

Graph based methods have recently gained a lot of attention in different areas of NLP as can be seen in ([5], [7], [8], [9]).These methods can be employed for word sense induction and disambiguation. In most of the graph-based methods each context word of the target word is represented as a vertex. If two vertices co-occur in one or more instances then they are connected via an edge. After a co-occurrence graph is made, different graph clustering algorithms can be applied to partition the graph. Each cluster then represents the set of words which are semantically related to a particular sense. In unsupervised systems many methods ([1], [10], [11]) construct word co-occurrences graph for a target polysemous word and then apply graph clustering to obtain the possible senses of that word.

The paper is structured as follows. First a detailed overview of the general steps for graph based approaches is given followed by most of the graph based algorithms in Section 2. In Section 3 we present the predominant related work done in this area. This Section forms the core part of this paper. The paper basically contains the survey of five different types of graph-based approaches in this Section 3. The conclusions and review based on work presented in Section 3 is given in Section 4.

II. General Steps for Graph-based Word Sense Disambiguation Methods

Graph-based methods used for Word Sense Disambiguation are mostly unsupervised methods that rely on the lexical KB graph structure for concluding the relevance of word senses in the given context. Usually the reference resource used in majority is WordNet [12].Graph based methods derive nodes from synsets and edges are derived from Semantic relations present between synsets. Following are the set of general steps as presented in [13]:

A. Building the Graph

Since we are considering WordNet as the reference resource, it is mapped into a graph whose nodes are synonym sets i.e., concepts and edges are semantic relationships between concepts (e.g., hyperonymy). A graph G= (V, E) is built, which is derived from the graph of reference lexicon. To be precise given a sentence $\sigma = w_1, w_2, \ldots, w_n$, where w_i is a word, we need the following steps to build G:

- The sense vocabulary $V\sigma$ is derived as $V\sigma := \bigcup_{i=1}^{n} Senses(w_i)$, where senses(w_i) is the set of senses of any of the w_i of the sentence.
- For each node, a visit of the WordNet graph is performed, every time a node is encountered, all intermediate node and edges are added to the graph.
- The constructed graph is the subgraph containing nodes and relations of all relevant vocabulary in the sentence.

B. Sense Ranking

Different ranking models are used with derived graph to find the correct senses of words in the sentence. Correct interpretation of the sentence can be obtained by ranking each vertex in graph G according to its centrality. In [7] different ranking models are described.

C. Disambiguation

The disambiguation is performed by assigning to each word w_i in the source sentence its correct j-th concept i.e., $sense_{ij}$ which is associated with the maximum resulting rank.

In the above mentioned general steps of graph-based methods, the major concern is related to the Sense ranking step. Although complex methods have been proposed for sense ranking, sentence oriented algorithms that build a graph G once per sentence irrespective of number of words present in the sentence are much more efficient.

III. Related Work

A. Random Walk Algorithm

Many natural language processing tasks consists of labeling sequences of words with linguistic annotations, e.g. word sense disambiguation, part-of-speech tagging, named entity recognition, and others. A graph based sequence data labeling algorithm is presented in [6] as solution for such natural language annotation tasks. The algorithm simultaneously annotates all the words in sequence by exploiting relations identified among word labels, using random walks on graphs encoding label dependencies.

The basic idea behind this algorithm is of "recommendation" or "voting". When one vertex links to the other, it is basically casting a vote for that other vertex. More the number of votes cast by the vertex, higher the importance of the vertex. Given a graph G= (V, E), let *In*(V_a) be the set of incoming vertices and *Out*(V_a) be the set of vertices the vertex V_a points to. Page Rank (from [6]) is given by:

$$P(V_a) = 1 - d + d * \sum_{V_b \in In(V_a)} \frac{P(V_b)}{|Out(V_b)|}$$

where parameter d is set between 0 and 1.

Algorithm for Sequence Data Labeling has following steps:

- Construction of label dependencies graph.
- Label scoring using graph based ranking algorithms.
- Label assignment.

The **church bells** no longer **rung** on **Sundays**.

church
1: one of the groups of Christians who have their own beliefs and forms of worship
2: a place for public (especially Christian) worship
3: a service conducted in a church

bell
1: a hollow device made of metal that makes a ringing sound when struck
2: a push button at an outer door that gives a ringing or buzzing signal when pushed
3: the sound of a bell

ring
1: make a ringing sound
2: ring or echo with sound
3: make (bells) ring, often for the purposes of musical edification

Sunday
1: first day of the week; observed as a day of rest and worship by most Christians

Figure1. Senses of words from WordNet (from [6])

Let us consider an example for explaining this algorithm. Consider the task of assigning senses to the words in the text "The church bells no longer rung on Sundays." Let us assume at most three senses of each word as shown in Figure 1. All word senses are added as vertices in the label graph, and weighted edges are drawn as dependencies among word senses, derived using definition based similarity measure. The resulting label graph is an undirected weighted graph as shown in Figure 2.

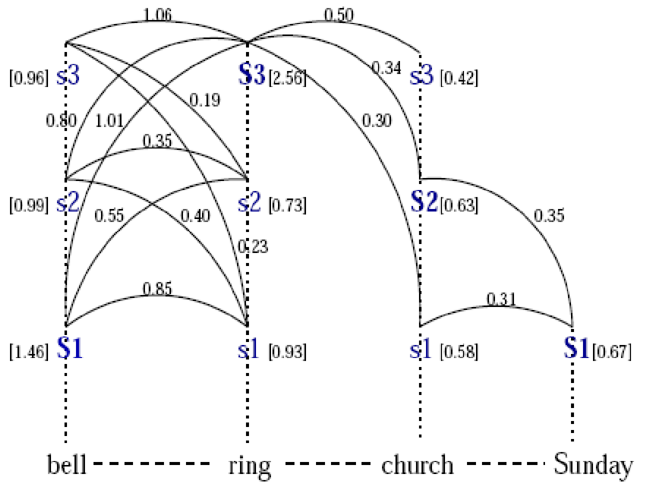

Figure2. The label graph for assigning senses to words in the sentence used as an example. (from [6])

After running the ranking algorithm, scores are identified for each word sense in the in the graph, indicated between brackets next to each node. Selecting for each word the sense with largest score results in the following sense assignment: The church#2 bells#1 no longer rung#3 on Sundays#1, which is correct.

B. A Distributional Extension of PageRank

Diego De Cao et al. (2010) presented an adaptation of PageRank algorithm for Word Sense Disambiguation. It preserves the reachable accuracy while significantly reducing the processing time. They exploit distributional evidence that can be automatically acquired from corpus, to amplify the performance of sentence oriented version. This type of algorithm has achieved improved efficiency and a speed-up of two orders of magnitude at no cost in accuracy. They employed Senseval 2007 coarse WSD dataset to measure the accuracy. Their evaluation is based on two main aspects. First the impact of topical expansion at sentence level on the accuracy reachable by Personalized PageRank PPR. Second analysis of the efficiency of algorithm and its impact on the sentence or word oriented perspective. In order to validate the hypothesis that Latent Semantic Analysis (LSA) is helpful to improve time complexity of WSD, analysis of processing times of different data sets was conducted, to compare methods and resources used.

C. Word Sense Induction using graph of collocations

Klapaftis and Manandhar (2008) proposed a graph-based approach that represents pair of words as vertices instead of single words. The idea behind the approach was that single words can appear with more than one senses of the target word. They hypothesized that the pair of words is unambiguous. This approach achieved good results in both evaluation settings of SemEval-2007 task. If one or more pairs of words representing the induced sense co-occur in the test instance then it is disambiguated towards one of the induced senses. But since cooccurrence of a pair of words is less likely than the occurrence of a single word, this approach creates a data sparsity problem.

D. Graphs of Unambiguous Vertices for Word Sense Induction and Disambiguation

Korkontzelos and Manandhar (2010) presented an unsupervised graph based method for word sense induction and disambiguation. They relaxed the condition imposed by the previous approach by allowing assignment of either a word or word pair to each vertex of the graph. This could be done because in some cases a single word is unambiguous. If the word is found to be unambiguous then it is used as a single word vertex. Otherwise, it is represented as pair-of-word vertex. Word senses are induced by clustering the constructed graph. While disambiguating the word, each induced cluster is scored according to number of its vertices found in the context of the target word. Thus, this system works in two major steps, word sense induction and then disambiguation.

1) Word Sense Induction

It consists of three main components, corpus preprocessing, graph construction and clustering.

- Corpus preprocessing: This step aims to find those words which are conceptually matching with the target word. They also applied certain filtering criteria in this step to eliminate words which occur in stop list, kept only nouns since they are more discriminative and also eliminated nouns whose relative frequency is greater in reference corpus than in the target word corpus.
- Graph construction*:* This step presents as vertices the nouns which were extracted in the previous step. Some vertices may also represent pair of nouns. The aim here is to keep only those pairs which point to the different sense of target word than their component nouns. Edges are then drawn based on co-occurrence of corresponding vertices contents and finally weights are applied.
- Clustering*:* A randomized graph clustering algorithm called Chinese Whispers proposed in [17] was used to cluster the graph. It automatically infers the number of clusters to be produced.

Figure 3 shows the block diagram giving the overview of word sense induction system.

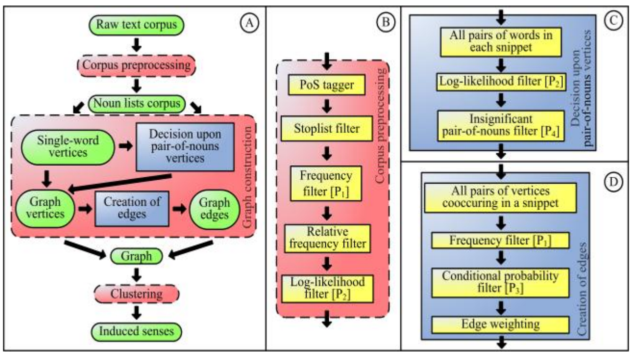

Figure 3. A: Block diagram presenting the system overview. B, C, D: Block diagram analysing the complex components of A. Parameter names are in square brackets. (from[15])

1) Word Sense Disambiguation

After word sense induction is complete, each induced cluster is assigned a score and the instance is assigned to the sense with highest score. This system participated in SemEval-2010 word sense induction and disambiguation task

E. Word Sense Induction and Disambiguation using Hierarchical Random Graphs

Klapaftis and Manandhar (2010) proposed that graphs often exhibit the hierarchical structure that goes beyond simple flat clustering. They present an unsupervised method for inferring the hierarchical grouping of senses of a polysemous word. The inferred hierarchical structures (binary trees) are then applied to word sense disambiguation. In this method the vertices of the graph are contexts of the polysemous word and edges represent the similarity between contexts.

The binary tree produced by this method groups the context of polysemous word at different heights of the tree and thus, induces the word sense at different levels of sense granularity. The Figure 4 shows the stages of this method.

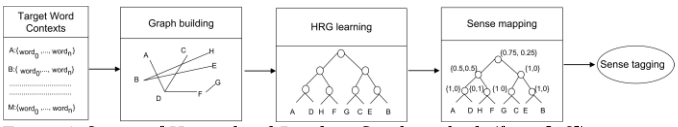

Figure 4. Stages of Hierarchical Random Graph method. (from [16])

IV. CONCLUSIONS

The paper gives the overview of graph based methods for word sense induction and disambiguation. Graph based methods which were dominantly used in social network analysis area are recently applied to Word Sense Disambiguation tasks. A large number of NLP problems can be solved using these graph based approaches. Although the expected accuracy of these methods is positively high, the main drawback of these methods is high computational demands when applied on large scale repositories.

The distributional approach of PageRank is useful in practical applications like query processing or document indexing. This area is open for cross-linguistic applications supported by multilingual lexical sense repositories.

The study shows us that using the pair of words as a vertex results in data sparsity problem and also hard clustering the graph will potentially identify less conflating senses of the target word. The major disadvantage of this system is the large number of induced senses.

It can also be concluded that inferring the Hierarchical structure of the graphs leads to superior performance as compared to simple flat clustering methods and traditional agglomerative clustering. These Hierarchical random graphs can also be applied in the area of taxonomy learning.

REFERENCES

[1] J. Veronis. 2004. Hyperlex: lexical cartography for information retrieval. *Computer Speech & Language*, 18(3):223-252, July.

[2] Hearst, M. (1991) Noun Homograph Disambiguation using Local Context in Large Text Corpora, Proc. 7th Annual Conference of the University of Waterloo Center for the new OED and Text Research, Oxford.

[3] Roberto Navigli, Kenneth C. Litkowski, and Orin Hargraves.2007. Semeval-2007 task 07: Coarse grained English all-words task. In the *Proceedings of Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pages 30-35, Prague, Czech Republic.

[4] Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice-cream cone. In the Proceedings of Fifth Annual International Conference on Systems Documentation (Toronto, Ontario, Canada). V. DeBuys, Ed. SIGDOC '86.

[5] Ravi Sinha and Rada Mihalcea. 2007. Unsupervised graph-based word sense disambiguation using measures of word semantic similarity. *In IEEE ICSC 2007.*

[6] Rada Mihalcea. Unsupervised large-vocabulary word sense disambiguation with graph-based algorithms for sequence data labeling. In *Proceedings of Human Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pages 411-418, Vancouver, British Columbia, Canada, October 2005. Association for Computer Linguistics.

[7] Roberto Navigli and Mirella Lapata. 2007. Graph connectivity measures for unsupervised word sense disambiguation. In the *Proceedings of IJCAI'07*, pages 1683-1688, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

[8] Eneko Agirre and Aitor Soroa. 2008. Using the multilingual central repository for graph based word sense disambiguation. In *Proceedings of LREC'08*, Marrakech, Morocco, May.

[9] E. Agirre and A. Soroa. 2009. Personalizing page rank for word sense disambiguation. In *Proceedings of 12th conference of EACL'09*, Athens, Greece, March 30-April 3.

[10] Dominic Widdows and Beate Dorow. 2002. A Graph Model for Unsupervised Lexical Acquisition. In Proceedings of Coling-2002, pages 1-7, Morristown, NJ, USA.

[11] Eneko Agirre, David Martinez, Oier Lopez de Lacalle, and Aitor Soroa. 2006. Two Graph based Algorithms for state-of-the-art WSD. In *Proceedings of EMNLP-2006*, pages 585-593, Sydney, Australia.

[12] G. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K. Miller. 1990. An on-line lexical database. *International Journal of Lexicography*, 13(4):235-312.

[13] Diego De Cao, Roberto Basili, Matteo Luciani, Francesco Mesiano, Ricardo Rossi. 2010. Robust and Efficient Page Rank for Word Sense Disambiguation. In the proceedings of 2010 Workshop on Graph-based Methods for Natural Language Processing, ACL 2010, pages 24-32, Uppsala, Sweden.

[14] I. Klapaftis and S. Manandhar. 2008. Word sense induction using graphs of collocations. In *proceedings of EACI-2008*, Patras, Greece.

[15] I. Korkontzelos and S. Manandhar. 2010. UoY: Graphs of Unambiguous Vertices for Word Sense Induction and Disambiguation. In Proceedings of 5th International Workshop on Semantic Evaluation, ACL 2010,pages 355-358, Uppsala, Sweden.

[16] I. Klapaftis and S. Manandhar. 2010. Word Sense Induction and Disambiguation Using Hierarchial Random Graphs. In the Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, pages 745-755, MIT, Massachusetts, USA.

[17] C. Biemann. 2006. Chinese Whispers- an efficient graph clustering algorithm and its application to natural language processing problems. In *proceedings of TextGraphs*, New York City. ACL.

[18] Agirre, Eneko, Philip Edmonds (eds.). 2006. Word Sense Disambiguation: Algorithms and Applications, Springer.