Word Sense Disambiguation with Semantic Networks

George Tsatsaronis*, Iraklis Varlamis, and Michalis Vazirgiannis

Department of Informatics, Athens University of Economics and Business, Athens, Greece
{gbt,varlamis,mvazirg}@aueb.gr

Abstract. Word sense disambiguation (WSD) methods evolve towards exploring all of the available semantic information that word thesauri provide. In this scope, the use of semantic graphs and new measures of semantic relatedness may offer better WSD solutions. In this paper we propose a new measure of semantic relatedness between any pair of terms for the English language, using WordNet as our knowledge base. Furthermore, we introduce a new WSD method based on the proposed measure. Experimental evaluation of the proposed method in benchmark data shows that our method matches or surpasses state of the art results. Moreover, we evaluate the proposed measure of semantic relatedness in pairs of terms ranked by human subjects. Results reveal that our measure of semantic relatedness produces a ranking that is more similar to the human generated one, compared to rankings generated by other related measures of semantic relatedness proposed in the past.

Key words: Word Sense Disambiguation, Semantic Networks, WordNet

1 Introduction

Word Sense Disambiguation (WSD) is the task of selecting the most appropriate meaning for any given word with respect to its context. The candidate word meanings, also referred to as senses, are usually selected from a machine readable dictionary (MRD) or a word thesaurus. Several approaches have been proposed in the past and are classified depending on the resources they employ for the WSD task. Knowledge-based or dictionary-based approaches usually utilize knowledge sources like MRDs or thesauri in order to address the task. Corpus-based approaches include the use of large corpora. An alternative classification may consider the use of a training mechanism that builds a decision model (i.e. a classifier) trained on manually annotated data in order to predict the correct sense of each given word. Such approaches are considered as supervised WSD approaches. The main distinction between a supervised WSD method and an unsupervised one is in whether they use manually labelled data or not. An extensive presentation of the state of the art in WSD can be found in [1].

In this paper we propose a new knowledge-based WSD approach that does not require training. The approach considers semantic networks generated from the WordNet

* Funded by the 03ED_850 research project, implemented within the Reinforcement Programme of Human Research Manpower (PENED) and co-financed by National and Community Funds (25% from the Greek Ministry of Development- General Secretariat of Research and Technology and 75% from E.U.-European Social Fund).

thesaurus [2] and introduces a new measure of semantic relatedness for a pair of thesaurus' concepts [1]. Experimental evaluation of the semantic relatedness measure in 65 word pairs ranked by human subjects according to their semantic relatedness shows that our measure produces a ranking that is more similar to the human generated one, compared to other related measures of semantic relatedness proposed in the past. Furthermore, we evaluate our approach in a benchmark WSD data set, namely Senseval 2 [3], and show that it surpasses or matches previous unsupervised WSD approaches. The rest of the paper is organized as follows: Some preliminary elements, as well as related work, are discussed in Section 2. Section 3 introduces the new measure of semantic relatedness and the new WSD method. Section 4 presents the experimental evaluation and Section 5 concludes and points to future work.

2 Background and Related Work

The idea of using semantic networks to perform WSD is not new. In fact, recent research has employed the construction of rich semantic networks that utilize WordNet fully. In this section we present preliminary information concerning WordNet and semantic networks.

2.1 WordNet

WordNet is a lexical database containing English nouns, verbs, adjectives and adverbs, organized in synonym sets (synsets). Synsets can be regarded as concepts. They are connected with various edges that represent different semantic relations (see Figure 1) and sometimes cross parts of speech (POS). The proposed measure of semantic relatedness and the introduced WSD approach utilize the full range of WordNet 2.0 semantic relations. Any other thesaurus could be used as long as it provides a similar graph structure, and semantic relations like the aforementioned, that can also cross POS.

2.2 Generating Semantic Networks from WordNet

The expansion of WordNet with semantic relations that cross POS has widened the possibilities of semantic network construction from text. Early approaches [4], were based on the gloss words existing in the terms' definitions in order to build semantic networks from text. More recent approaches in semantic network construction from word thesauri [5, 6] utilized the semantic relations of WordNet. These methods outperformed previous methods that use semantic networks in the *all words* WSD tasks of Senseval 2 and 3 for the English language. The evaluation in [6] revealed that the performance boost of the WSD task was mainly due to the use of the rich semantic links that WordNet offers. In this work we adopt the same semantic network construction method.

[1] *Concept* and *sense* will be used interchangeably for the remaining of the paper.

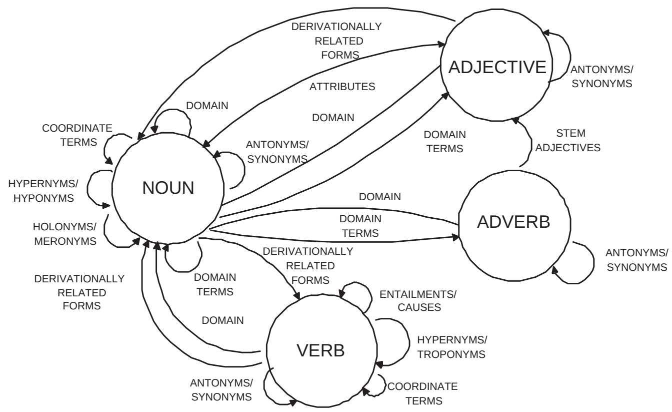

Fig. 1. Semantic relations in WordNet.

2.3 Semantic Relatedness Measures in Word Sense Disambiguation

Agirre and Rigau in [7] base their measure for sets of concepts on the individuals' density and depth and on the length of the shortest path that connects them. Resnik [8] measure for pairs of concepts is based on the information content of the deepest concept that can subsume both. Measures proposed by Jiang and Conrath [9], Hirst and St-Onge [10], Leacock and Chodorrow [11], and Lin [12], were based on similar ideas. Due to space limitations we suggest the reader to consult the analysis of Budanitsky and Hirst [13] for the majority of the aforementioned measures. All these measures are based on the noun hierarchy, whereas our measure defines the semantic relatedness between any two concepts, independently of their POS. The proposed WSD approach is based on a new measure of semantic relatedness for concept pairs, which combines in tandem the length of the semantic path connecting them, the type of the semantic edges and the depth of the nodes in the thesaurus. Experimental evaluation shows that this measure familiarizes human understanding of semantic similarity better than all other measures.

In this work we focus only to the unsupervised approaches that are based on semantic networks. Patwardhan et al. [14] modified the Lesk method to allow for the use of any measure of semantic relatedness. Mihalcea et al. [5] constructed semantic networks from WordNet and ran an adaptation of the PageRank algorithm on them, in order to address the *all words* task for the English language. Tsatsaronis et al. [6] constructed richer semantic networks and surpassed or matched the performance of the PageRank semantic networks, with a constraint spreading activation technique. We compare our work to the WSD approaches mentioned above in section 4.2. Results show that our method surpasses or matches state of the art results for the English *all words* task in Senseval 2.

3 WSD Based on Semantic Relatedness of Terms

In this section we propose a new measure of semantic relatedness and a new WSD approach based on that measure.

3.1 Semantic Relatedness

The proposed measure of semantic relatedness for a pair of concepts considers in tandem three factors: a) the semantic path length that connects the two concepts, captured by *semantic compactness*, b) the path depth, captured by *semantic path elaboration*, and c) the importance of the edges comprising the path. A measure for WSD based on the idea of *compactness* was initially proposed in [15], but it only used nouns and the hypernym relation. We enhanced that measure by considering all of WordNet's relations and POS.

Definition 1. *Given a word thesaurus O, a weighting scheme for the edges that assigns a weight $e \in (0,1)$ for each edge, a pair of senses $S = (s_1, s_2)$, and a path of length l connecting the two senses, the semantic compactness of S ($SCM(S,O)$) is defined as $\prod_{i=1}^{l} e_i$, where $e_1, e_2, ..., e_l$ are the path's edges. If $s_1 = s_2$ $SCM(S,O) = 1$. If there is no path between s_1 and s_2 $SCM(S,O) = 0$.*

Note that *semantic compactness* considers the path length and has values in [0, 1]. Higher *semantic compactness* between senses means higher semantic relatedness. Also, larger weights are assigned to stronger edge types. The intuition behind the assumption of edges' weighting is the fact that some edges provide stronger semantic connections than others. A standard way od obtaining the edges' weights can be the measurement of edges' distribution in WordNet. The frequency of occurrence of each edge type can act as its weight. The *semantic compactness* of two senses s_1 and s_2, can take different values for all the different paths that connect the two senses. Another parameter that affects term relatedness is the depth of the sense nodes comprising the path. A standard means of measuring depth in a word thesaurus is the hypernym/hyponym hierarchical relation for the noun and adjective POS and hypernym/troponym for the verb POS. A path with shallow sense nodes is more general compared to a path with deep nodes. This parameter of semantic relatedness between senses is captured by the measure of *semantic path elaboration* introduced in the following definition.

Definition 2. *Given a word thesaurus O and a pair of senses $S = (s_1, s_2)$, where s_1,$s_2 \in O$ and $s1 \neq s2$, and a path between the two senses of length l, the semantic path elaboration of the path (SPE(S,O)) is defined as $\prod_{i=1}^{l} \frac{2d_i d_{i+1}}{d_i + d_{i+1}} \cdot \frac{1}{d_{max}}$, where d_i is the depth of sense s_i according to O, and d_{max} the maximum depth of O. If $s_1 = s_2$, and $d = d_1 = d_2$ $SPE(S,O) = \frac{d}{d_{max}}$. If there is no path from s_1 to s_2, $SPE(S,O) = 0$.*

SPE is in fact the harmonic mean of the two depths normalized to the maximum thesaurus depth. The harmonic mean offers a lower upper bound than the average of depths and we think is a more realistic estimation of the path's depth. *Compactness* and *Semantic Path Elaboration* measures capture the two most important parameters of measuring semantic relatedness between terms [13], namely path length and senses depth in the used thesaurus. We combine these two measures in the definition of *Semantic Relatedness* between two senses.

Definition 3. *Given a word thesaurus O and a pair of senses $S = (s_1, s_2)$ the semantic relatedness of S (SR(S,O)) is defined as $max\{SCM(S,O) \cdot SPE(S,O)\}$.*

Note that definition 3 can be expanded to measure the semantic relatedness for a pair of terms $T = (t_1, t_2)$, namely $SR(T, O)$. For all the pair combinations of senses that t_1 and t_2 may be assigned, the maximum value of semantic relatedness between any two senses found is defined as the semantic relatedness of the pair of terms. In case $t_1 \equiv t_2 \equiv t$ and $t \notin O$ then semantic relatedness can be considered as 1. The semantic relatedness can only take real values in $[0, 1]$.

Algorithm 1 Word-Sense-Disambiguation(T,w,O,Θ)

Require: A set of POS-tagged terms T to be disambiguated, a word thesaurus O, a weighting scheme $w : E \rightarrow (0..1)$ for the edges of the used thesaurus and an upper threshold Θ for the maximum number of combinations examined in simulated annealing.
Ensure: A mapping of terms to senses that disambiguate them.

Word-Sense-Disambiguation(T,w,O,Θ)
for all terms $t \in T$ **do**
    senses$[t]$ =number of possible senses of t
    correct-sense$[t]$ =random(senses$[t]$)
end for
Minimum-MST-Weight=compute-SCH(T,correct-sense,O)
while iterations $i \leq \Theta$ **do**
    Transit randomly to a neighboring assignment of senses
    Temp-MST-Weight=compute-SCH(T,correct-sense,O)
    ΔE=Minimum-MST-Weight - Temp-MST-Weight
    if $\Delta E > 0$ **then**
        Transit to neighboring state with probability $e^{\frac{\Delta E}{i}}$
    end if
end while
return correct-sense

3.2 Word Sense Disambiguation Based on Semantic Relatedness

We expand the measure of semantic relatedness between a pair of terms introduced in the previous section, to a measure of semantic coherence between a set of terms, and we use this measure to form a new knowledge-based WSD algorithm that does not require training.

Definition 4. *Given a word thesaurus O and a set of n terms $T = (t_1, t_2, ..., t_n)$, where for each t_i, $i = 1..n$, it holds that $t_i \in O$, let $S = (s_1, s_2, ..., s_n)$ be a possible assignment of senses to the terms in T. The semantic coherence of T (SCH(T,O)) is then defined as the weight of the Minimum Spanning Tree (MST) computed on the weighted undirected graph having: 1) a node for each sense in S, 2) an edge for each pair of senses (s_i, s_j) that are semantically related ($SR((s_i, s_j), O) > 0$), with an edge weight $w_{i,j} = \frac{1}{SR((s_i,s_j),O)}$.*

Based on this definition, the WSD algorithm operates as follows: From all MSTs produced for the set of terms we choose the one with the maximum semantic coherence.

To alleviate the computational burden occurring by examining all possible MSTs, we use simulated annealing [16].

4 Experimental Evaluation

The experimental evaluation is bi-fold. Firstly, we compare the measure of semantic relatedness against state of the art measures, using a set of term pairs weighted by humans as a benchmark. Secondly, we evaluate the performance of the proposed WSD algorithm in the Senseval 2 benchmark collection.

4.1 Semantic Relatedness Measure Evaluation

The relatedness measures that we compare to are: Hirst and St-Onge (HS), Jiang and Conrath (JC), Leackock and Chodorow (LC), Lin (L) and Resnik (R), which are thoroughly discussed in [13]. We use the test set of 65 term pairs initially proposed by Rubenstein and Goodenough [17] and rank the term pairs using the semantic relatedness scores given by each measure and by the 51 human subjects. We measure the correlation of all rankings, including ours (SRel), using the Kendall's Tau distance measure [18]. The results are:

Table 1. Kendall's Tau distance from human rankings.

	HC	**JC**	**LC**	**L**	**R**	**SRel**
Kendall's Tau	0.371	0.250	0.247	0.242	0.260	0.169

4.2 Word Sense Disambiguation Evaluation

The proposed WSD method is evaluated in Senseval 2, for the English *all words* task. The computation of the semantic relatedness between any pair of concepts requires a weighting scheme that assigns values in $(0, 1)$. A standard weighting scheme can be the distribution of edge types in the used thesaurus. The frequency of occurrence can be the respective edge's weight. We followed that scheme, and the edge weights we produced (hypernym/hyponym edges obtained 0.57, nominalization edges 0.14, etc.) are in accordance to those stated Song et al. in [19], and the same with the ones obtained if semantic networks are constructed for each sentence in the SemCor data set. Thus, no effort for training in order to learn the edges' weights is required. We compare our approach (MST) with the standard unsupervised baseline, which randomly assigns a sense to a given word, the best reported unsupervised method in the Senseval 2 competition [3], an unsupervised approach utilizing spreading of activation on semantic networks (SANs)[6] and an unsupervised approach executing PageRank [5] (PRSN). Reported accuracy is shown in Table 2. Results show that the proposed WSD method surpasses

Table 2. Overall and per file accuracy on the Senseval 2 data set.

	Words		MST	Baseline	SANs	Best Unsup. Senseval 2	PRSN
	Mono	Poly					
File 1 (d00)	**103**	**552**	**0.436**	**0.365**	**0.459**	unavailable	**0.439**
File 2 (d01)	**232**	**724**	**0.498**	**0.421**	**0.468**	unavailable	**0.544**
File 3 (d02)	**129**	**563**	**0.511**	**0.430**	**0.557**	unavailable	**0.542**
Overall	**464**	**1839**	**0.485**	**0.407**	**0.492**	**0.451**	**0.508**

the baseline and the best unsupervised method of the Senseval 2 competition. Furthermore, it matches the performance of SANs and PageRank methods, which are, to the best of our knowledge, the approaches with the best ever reported performance in unsupervised WSD overall for all POS. The difference with these two methods is in the order of magnitude of 10^{-3} and 10^{-2} respectively. The statistical significance of our results is calculated using 0.95 confidence intervals for all methods' accuracies (figure 2).

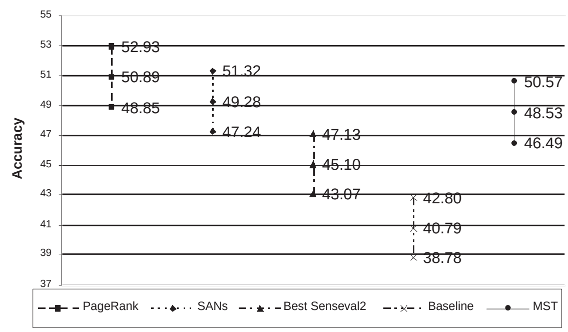

Fig. 2. Methods' accuracies with 0.95 confidence intervals.

5 Conclusions and Future Work

In this paper we introduce a new measure of semantic relatedness between senses and expand it to compute relatedness for a pair of terms. The measure combines in tandem the concepts' depth in the used thesaurus, the semantic path length that connects the two concepts and the importance of the semantic edges comprising the path. Experimental evaluation in ranking term pairs according to their relatedness, shows that our measure produces a ranking that is more similar to the human generated one, compared to rankings generated by other related measures of semantic relatedness proposed in the past. Finally, we embedded this measure into a new WSD approach that is based on measuring the weight of the minimum spanning tree connecting candidate senses.

Experimental evaluation shows that our method surpasses or matches state of the art results. In the future we will investigate the impact of each of the three factors comprising our measure of semantic relatedness in the WSD task, and embed it into text retrieval and text classification models.

References

1. Agirre, E., Edmonds, P.: Word Sense Disambiguation: Algorithms and Applications. Springer (2006).
2. Fellbaum, C.: WordNet, An Electronic Lexical Database. The MIT Press (1998).
3. Palmer, M., Fellbaum, C., Cotton, S.: English tasks: All-words and verb lexical sample. In: Proc. of Senseval-2., Toulouse, France (2001) 21–24.
4. Veronis, J., Ide, N.: Word sense disambiguation with very large neural networks extracted from machine readable dictionaries. In: Proc. of the 13th International Conference on Computational Linguistics, Finland (1990) 389–394.
5. Mihalcea, R., Tarau, P., Figa, E.: Pagerank on semantic networks with application to word sense disambiguation. In: Proc. of the 20th CoLing, Switzerland (2004).
6. Tsatsaronis, G., Vazirgiannis, M., Androutsopoulos, I.: Word sense disambiguation with spreading activation networks generated from thesauri. In: Proc. of the 20th International Joint Conference on Artificial Intelligence, India, AAAI Press (2007) 1725–1730.
7. Agirre, E., Rigau, G.: A proposal for word sense disambiguation using conceptual distance. In: Proc. of the 1st International Conference on Recent Advances in NLP. (1995) 258–264.
8. Resnik, P.: Using information content to evaluate semantic similarity. In: Proc. of the 14th International Joint Conference on Artificial Intelligence, Canada (1995) 448–453.
9. Jiang, J., Conrath, D.: Semantic similarity based on corpus statistics and lexical taxonomy. In: Proc. of ROCLING X, Taiwan (1997) 19–33.
10. Hirst, G., St-Onge, D.: Lexical chains as representations of context for the detection and correction of malapropisms. In: WordNet: An Electronic Lexical Database, chapter 13, Cambridge, The MIT Press (1998) 305–332.
11. Leacock, C., Chodorow, M.: Combining local context and wordnet similarity for word sense identification. In: WordNet: An Electronic Lexical Database, chapter 11, Cambridge, The MIT Press (1998) 265–283.
12. Lin, D.: An information-theoretic definition of similarity. In: Proc. of the 15th International Conference on Machine Learning. (1998) 296–304.
13. Budanitsky, A., Hirst, G.: Evaluating wordnet-based measures of lexical semantic relatedness. Computational Linguistics **32**(1) (2006) 13–47.
14. Patwardhan, S., Banerjee, S., Pedersen, T.: Using measures of semantic relatedness for word sense disambiguation. In: Proc. of the 4th International Conference on Inbtelligent Text Processing and Computational Linguistics, Springer Verlag (2003) 241–257.
15. Mavroeidis, D., Tsatsaronis, G., Vazirgiannis, M., Theobald, M., Weikum, G.: Word sense disambiguation for exploiting hierarchical thesauri in text classification. In: Proc. of the 9th PKDD, Portugal, Springer Verlag (2005) 181–192.
16. Cowie, J., Guthrie, J., L., G.: Lexical disambiguation using simulated annealing. In: Proc. of the 14th CoLing, France (1992) 359–365.
17. H., R., Goodenough, J.: Contextual correlates of synonymy. Communications of the ACM **8**(10) (1965) 627–633.
18. Fagin, R., Kumar, R., D., S.: Comparing top k lists. SIAM Journal on Discrete Mathematics **17**(1) (2003) 134–160.
19. Song, Y., Han, K., Rim, H.: A term weighting method based on lexical chain for automatic summarization. In: Proc. of the 5th CICLing Conference, USA (2004) 636–639.