

Semantic Relatedness for Biomedical Word Sense Disambiguation

Kiem-Hieu Nguyen and Cheol-Young Ock

School of Electrical Engineering

University of Ulsan

93, Daehakro, Nam-gu, Ulsan 680-749, Korea

{hieunk, okcy}@mail.ulsan.ac.kr

Abstract

This paper presents a graph-based method for all-word word sense disambiguation of biomedical texts using semantic relatedness as edge weight. Semantic relatedness is derived from a term-topic co-occurrence matrix. The sense inventory is generated by the MetaMap program. Word sense disambiguation is performed on a disambiguation graph via a vertex centrality measure. The proposed method achieves competitive performance on a benchmark dataset.

1 Introduction

Word Sense Disambiguation (WSD) has been an open problem in Computational Linguistics (Navigli, 2009). It aims at identifying the correct meaning of an ambiguous word in a given context, e.g., ‘adjustment’ could refer to *individual adjustment* or *adjustment action* in “*marital adjustment*” and “*dietary adjustment*”, respectively.

Supervised methods outperform unsupervised and knowledge-based methods (McInnes, 2009; Nguyen and Ock, 2010). However, they require expensive manual annotations and only the words of which training data are available could be disambiguated. On the other hand, knowledge-based and unsupervised methods overcome the two shortcomings by using knowledge sources or untagged raw texts (McInnes, 2008; Agirre et al., 2010; Ponzetto and Navigli, 2010).

Among knowledge-based methods, the graph-based method using semantic relatedness achieves state-of-the-art performance (Sinha and Mihalcea,

Concept	Semantic Type
<i>Individual adjustment</i>	<i>Individual Behavior</i>
<i>Adjustment action</i>	<i>Functional Concept</i>
<i>Psychological adjustment</i>	<i>Mental Process</i>

Table 1: The UMLS concepts and appropriate Semantic Types of the term ‘adjustment’.

2007; Nguyen and Ock, 2011). This work aims at applying the method to the biomedical domain.

This paper proposes calculating semantic relatedness between Semantic Types based on the Semantic Type Indexing (Humphrey et al., 2006) algorithm. WSD is then performed via a state-of-the-art graph-based method (Sinha and Mihalcea, 2007). The proposed method achieves competitive performance on a benchmark dataset.

The paper is organized as follows: Section 2 presents the graph-based WSD method; Section 3 describes the calculation of semantic relatedness; Experimental results are shown in Section 4; The paper ends with conclusions in Section 5.

2 The WSD Method

Sense inventory is essential for WSD. In the biomedical domain, the MetaMap program (Aronson, 2001) has been used to generate concept candidates in the Unified Medical Language System (Bodenreider, 2004) (UMLS) for ambiguous terms.

The concepts in the UMLS are assigned to predefined topics called Semantic Types (ST) (Table 1). Hence, STs could be efficiently used for disambiguation of biomedical terms (Humphrey et al., 2006). For instance, if the term ‘adjustment’ is

mapped to the ST *Mental Process* then it is disambiguated as *psychological adjustment*.

The WSD method used in this work is derived from Sinha and Mihalcea (2007) with an additional postprocessing step (Humphrey et al., 2006). The method consists of three steps:

- A disambiguation graph is generated for each context. The vertices are STs. The edge weight is semantic relatedness between STs (Section 3).
- Each ambiguous term is mapped to the ST with the highest rank based on vertex centrality.
- The term is disambiguated as the appropriate concept of the selected ST.

On the one hand, *i*) the method achieves state-of-the-art performance for WSD on general texts using WordNet (Miller, 1995) as sense inventory and the source to calculate semantic relatedness (Sinha and Mihalcea, 2007; Nguyen and Ock, 2011). By far, semantic relatedness between biomedical concepts has been studied on the UMLS meta-thesaurus (Pedersen et al., 2007) but there has been no work on applying semantic relatedness (particularly between STs) to biomedical WSD. On the other hand, *ii*) the method is effective in terms of implementation, comprehension, and computational complexity.

2.1 Disambiguation Graph

The **Algorithm 1** generates an undirected fully connected disambiguation graph for a context $C = \{w_0, w_1, \dots, w_n\}$. The dictionary D maps from an ambiguous term w_i to its ST candidates $D(w_i)$. D is generated by MetaMap. Given a term w_i , MetaMap generates a list of its UMLS concept candidates. In UMLS, each concept is, in turn, assigned to one or several STs. From that, we can create a list of ST candidates for w_i . The resulted disambiguation graph G contains vertices as STs and weight edge as semantic relatedness between STs.

From line 1 to line 8, the algorithm generates the vertices of G from the dictionary. From line 9 to line 15, the algorithm calculates the edge weight as semantic relatedness between STs (3).

Algorithm 1 Disambiguation graph creation

Input: Context $\{w_0, w_1, \dots, w_n\}$.

Input: Dictionary D

Output: Disambiguation graph $G = (V, E)$

```

1:  $V \leftarrow \phi$  # Initialize graph vertices.
2: for all  $w_i \in w$  do
3:   for all  $ST_i \in D(w_i)$  do
4:     if  $ST_i \notin V$  then
5:        $V \leftarrow V \cup \{ST_i\}$ 
6:     end if
7:   end for
8: end for
9:  $E \leftarrow \phi$  # Initialize the edges of the graph.
10: for all  $ST_i \in V$  do
11:   for all  $ST_j \in V \setminus \{\{ST_i\} \cup D^{-1}(ST_i)\}$  do
12:      $e_{ST_i, ST_j} \leftarrow sr(ST_i, ST_j)$ 
13:      $E \leftarrow E \cup \{e_{ST_i, ST_j}\}$ 
14:   end for
15: end for
16: return  $G = (V, E)$ 

```

2.2 Disambiguation based on Vertex Centrality

Given the disambiguation graph G , the rank of a vertex ST_i is defined as its weighted node-degree (shortly as *degree*):

$$degree(ST_i) = \sum_{ST_j \in V, e_{ST_i, ST_j} \in E} e_{ST_i, ST_j}, \quad (1)$$

For each ambiguous term w_i , the ST with the highest rank is selected among its ST candidates:

$$\underset{ST_k \in D(w_i)}{\operatorname{argmax}} degree(ST_k) \quad (2)$$

While there are alternative vertex centrality measures such as *betweenness*, *closeness*, and *eigenvector centrality*, empirical evidences show that *degree* achieves state-of-the-art performance on several benchmark datasets (Sinha and Mihalcea, 2007; Ponzetto and Navigli, 2010; Nguyen and Ock, 2011).

Given a sentence “*Clinically, these four patients had mild symptoms which improved with dietary adjustment*”, the terms not existing in the sense inventory are ignored, the rest are mapped to ST candidates as (‘four’: *Quantitative Concept*, ‘patients’: *Patient or Disabled Group*; ‘mild’: *Qualitative Concept*; ‘symptoms’: *Functional Concept*, *Sign or*

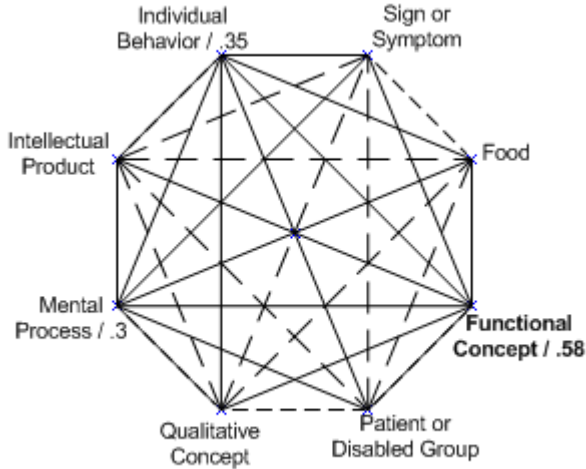


Figure 1: The disambiguation graph for “Clinically, these four patients had mild symptoms which improved with dietary adjustment”. The edges containing the ST candidates of ‘adjustment’ are solid lines, the rest are dot lines.

Symptom; ‘improved’: *Qualitative Concept, Intellectual Product*; ‘dietary’: *Food*; ‘adjustment’: *Individual Behavior, Functional Concept, Mental Process*). The disambiguation graph hence contains eight STs (Fig. 1).

If we want to disambiguate, for instance, ‘adjustment’, we could compare the degree of its ST candidates. As seen in Fig. 1, *Functional Concept* is the highest rank ST. Consequently, ‘adjustment’ is disambiguated as *adjustment action* (not *individual adjustment* or *psychological adjustment*).

3 Semantic Relatedness between STs

3.1 Motivations

Pedersen et al. (2007) show that there is no general-purpose measure among the six state-of-the-art semantic relatedness measures calculated based on the UMLS ontology and medical corpora. For instance, the corpus-based measure is close to physician judgments while the path-based and information content based measures are close to medical coders judgments. This is one of the main obstacles that prevent the use of semantic relatedness between concepts for biomedical WSD.

In another direction, Humphrey et al. (2006) induce a term-ST matrix from medical corpora. The WSD method proposed in that work is similar to

the *Lesk* algorithm (Lesk, 1986) where each ST *profile* is compared with the context using the term-ST matrix to select the highest rank ST. Nguyen and Ock (2011) show that using the same *synset profiles* in WordNet, the Lesk-based method achieves higher *precision* but lower *recall* than the semantic relatedness-based method.

3.2 The Proposed Measure

Semantic relatedness between STs is calculated from a term-ST matrix $A_{m,n}$ proposed in the Semantic Type Indexing algorithm (Humphrey et al., 2006) where m is the number of STs and n is the size of vocabulary. $A_{i,j}$ is the normalized frequency that the i^{th} ST and the j^{th} term co-occur. Hence, each row of the matrix is an ST *profile* that can be used to calculate context-sensitive semantic relatedness between STs as follows:

- Given a set of terms $\{w_0, w_1, w_2, \dots, w_n\}$ in a context C and the term-ST matrix A .
- The static vector of the i^{th} ST is A_i , the i^{th} row of A . $A(i)$ contains all the terms in the vocabulary. $A_i(C)$ is generated from A_i by assigning *zero* to all the terms not in C .
- The context-sensitive semantic relatedness of the i^{th} and j^{th} STs is defined as the dot product of the two context-sensitive vectors:

$$sr(ST_i, ST_j) = A_i(C) \cdot A_j(C) \quad (3)$$

4 Experiments

4.1 Test Dataset

The NLM-WSD dataset contains 5,000 contexts of 50 frequent ambiguous biomedical terms from the paper abstracts of the 1998 MEDLINE database (Weeber et al., 2001). Each ambiguous term has 100 contexts including the surrounding sentence, paper title and abstract. The average number of senses per term is 3.28.

4.2 Experimental Setups

The MetaMap program was used to generate ST candidates of ambiguous terms.

The *most frequent sense (MFS)* heuristic was used as the baseline system: For each ambiguous term,

System	A	P	R	F
<i>SR</i>	93.2	74.8	69.7	72.2
<i>Static-SR</i>	95.0	66.2	62.9	64.5
<i>STI</i>	93.2	74.1	69.0	71.5
<i>PPR</i>	100.0	68.1	68.1	68.1
<i>MFS</i>	100.0	85.5	85.5	85.5

Table 2: Experimental results on the NLM-WSD dataset.

the most frequent concept calculated based on the NLM-WSD dataset is simply selected.

The experimental results were compared using *attempted*, *precision*, *recall* and *F-measure* (A, P, R, and F, respectively).

4.3 Experimental Results

The proposed method, namely *SR*, was compared with three knowledge-based systems:

- *Static-SR*: The system uses static ST vectors, i.e., A_i , instead of context-sensitive ST vectors, i.e., $A_i(C)$ as described in Section 3.
- *STI*¹ (Humphrey et al., 2006): For a context, the rank of an ST candidate is the average ranks across all words in the context, e.g., for the context *Clinically, these four patients had mild symptoms which improved with dietary adjustment*, the rank of *Functional Concept* is [.5314 (‘symptoms’) + .4714 (‘adjustment’) + .7149 (‘patients’) + .1804 (‘dietary’) + .7226 (‘mild’) + .7282 (‘improved’) + .7457 (‘four’)] / 7 = .5849.
- *PPR*² (Agirre et al., 2010): The system uses the UMLS metathesaurus as a lexical knowledge graph and executes the Personalized PageRank, a state-of-the-art graph-based method, on the graph (Agirre and Soroa, 2009).

The performance of the *MFS* baseline is remarkably high, i.e., 85.5% of *F-measure* (Table 2). This shows that the sense distribution in the biomedical domain is highly skewed. Hence, this simple supervised heuristic outperformed all the investigated knowledge-based systems.

¹Available as a component of the MetaMap program.

²<http://ixa2.si.ehu.es/ukb>

Because *STI* and *SR* performed WSD via the disambiguation of STs, the two systems failed when the ST with the highest rank was assigned to at least two concepts. For instance, given the term ‘cold’, if *Disease or Syndrome* scores the highest rank, the two systems cannot decide whether *common cold* or *chronic obstructive airway disease* is the correct concept. Hence, the *attempted* status of *SR* and *STI* didn’t reach 100%.

SR was remarkably superior to *Static-SR* which empirically supports the context-sensitive ST vectors over static ones. Overall, *SR* and *STI* achieved the best performance.

5 Conclusions

In our experiments, the ST *profiles* were induced from the term-ST co-occurrence matrix. On the other hand, semantic relations and textual definitions in WordNet are useful for word sense disambiguation (Ponzetto and Navigli, 2010; Nguyen and Ock, 2011). Hence, the semantic relations between STs and the textual definitions of ST in the Unified Medical Language System could be potential resources for the disambiguation of biomedical texts.

The paper presents a graph-based method to biomedical word sense disambiguation using semantic relatedness between pre-defined biomedical topics. The proposed method achieves competitive performance on the NLM-WSD dataset. Because the achieved performance is significantly inferior to the performance of the most frequent sense heuristic, there is still more ground for improvement.

Acknowledgments

This work is partially supported by a scholarship of the Korean Ministry of Information and Communication. The first author receives a tuition fellowship of PhD course from the University of Ulsan, Korea. The authors would like to thank the United States National Library of Medicine for providing the NLM-WSD dataset and the MetaMap program, and the anonymous reviewers for their suggestions.

References

- E. Agirre and A. Soroa. 2009. Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the*

- 12th Conference of the European Chapter of the ACL (EACL 2009), pages 33–41, Athens, Greece, March. Association for Computational Linguistics.
- E. Agirre, A. Soroa, and M. Stevenson. 2010. Graph-based word sense disambiguation of biomedical documents. *Bioinformatics*, 26:2889–2896, November.
- AR Aronson. 2001. Effective mapping of biomedical text to the UMLS metathesaurus: the MetaMap program. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 17–21.
- O. Bodenreider. 2004. The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(Database issue):D267–D270, January. PMID: 14681409 PMID: 308795.
- S. M. Humphrey, W. J. Rogers, H. Kilicoglu, D. Demner-Fushman, and T. C. Rindfleisch. 2006. Word sense disambiguation by selecting the best semantic type based on journal descriptor indexing: Preliminary experiment. *J. Am. Soc. Inf. Sci. Technol.*, 57:96–113, January.
- Michael Lesk. 1986. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings of the 5th annual international conference on Systems documentation*, SIGDOC '86, pages 24–26, New York, NY, USA. ACM.
- B. T. McInnes. 2008. An unsupervised vector approach to biomedical term disambiguation: integrating umls and medline. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Student Research Workshop*, HLT-SRWS '08, pages 49–54, Stroudsburg, PA, USA. Association for Computational Linguistics.
- B. T. McInnes. 2009. Supervised and knowledge-based methods for disambiguating terms in biomedical text using the umls and metamap.
- G. A. Miller. 1995. Wordnet: a lexical database for english. *Commun. ACM*, 38:39–41, November.
- R. Navigli. 2009. Word sense disambiguation: A survey. *ACM Comput. Surv.*, 41:10:1–10:69, February.
- K. H. Nguyen and Ch. Y. Ock. 2010. Margin perceptron for word sense disambiguation. In *Proceedings of the 2010 Symposium on Information and Communication Technology*, SoICT '10, pages 64–70, New York, NY, USA. ACM.
- K. H. Nguyen and Ch. Y. Ock. 2011. Word sense disambiguation as a traveling salesman problem. *Artificial Intelligence Review*, December.
- T. Pedersen, S. V. S. Pakhomov, S. Patwardhan, and Ch. G. Chute. 2007. Measures of semantic similarity and relatedness in the biomedical domain. *J. of Biomedical Informatics*, 40:288–299, June.
- S. P. Ponzetto and R. Navigli. 2010. Knowledge-rich word sense disambiguation rivaling supervised systems. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, ACL '10, pages 1522–1531, Stroudsburg, PA, USA. Association for Computational Linguistics.
- R. Sinha 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*, pages 363–369, Washington, DC, USA. IEEE Computer Society.
- M. Weeber, J. G. Mork, and A. R. Aronson. 2001. Developing a test collection for biomedical word sense disambiguation. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 746–750. PMID: 11825285.