

Using a Weighted Semantic Network for Lexical Semantic Relatedness

Reda Siblini

Concordia University
1400 de Maisonneuve Blvd. West
Montreal, Quebec, Canada, H3G 1M8
r_sibl@encs.concordia.ca

Leila Kosseim

Concordia University
1400 de Maisonneuve Blvd. West
Montreal, Quebec, Canada, H3G 1M8
kosseim@encs.concordia.ca

Abstract

The measurement of semantic relatedness between two words is an important metric for many natural language processing applications. In this paper, we present a novel approach for measuring semantic relatedness that is based on a weighted semantic network. This approach explores the use of a lexicon, semantic relation types as weights, and word definitions as a basis to calculate semantic relatedness. Our results show that our approach outperforms many lexicon-based methods to semantic relatedness, especially on the TOEFL synonym test, achieving an accuracy of 91.25%.

1 Introduction

Lexical semantic relatedness is a measurement of how two words are related in meaning. Many natural language processing applications such as textual entailment, question answering, or information retrieval require a robust measurement of lexical semantic relatedness. Current approaches to address this problem can be categorized into three main categories: those that rely on a lexicon and its structure, those that use the distributional hypothesis on a large corpus, and hybrid approaches.

In this paper, we propose a new lexicon-based approach to measure semantic relatedness that is based on a weighted semantic network that includes all 26 semantic relations found in WordNet in addition to information found in the glosses.

2 Related Work

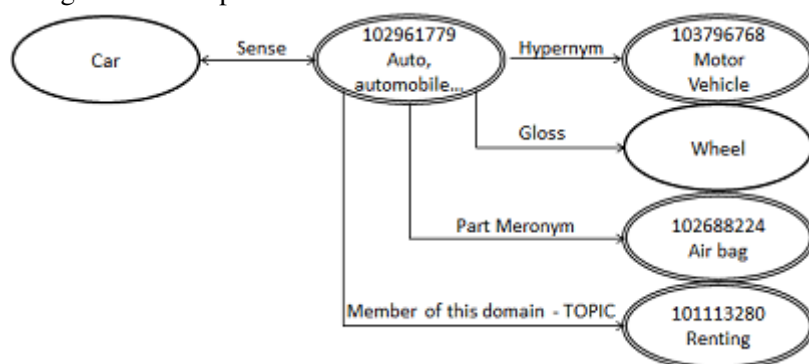
Approaches to computing semantic relatedness can be classified into three broad categories:

lexicon-based, corpus-based, and hybrid approaches.

Lexicon-based methods use the features of a lexicon to measure semantic relatedness. The most frequently used lexicon is Princeton's WordNet (Fellbaum, 1998) which groups words into synonyms sets (called synsets) and includes various semantic relations between those synsets, in addition to their definitions (or glosses). WordNet contains 26 semantic relations that include: hypernymy, hyponymy, meronymy, and entailment.

To measure relatedness, most of the lexicon-based approaches rely on the structure of the lexicon, such as the semantic link path, depth (Leacock and Chodorow, 1998; Wu and Palmer, 1994), direction (Hirst and St-Onge, 1998), or type (Tsatsaronis et al., 2010). Most of these approaches exploit the hypernym/hyponym relations, but a few approaches have also included the use of other semantic relations. Leacock and Chodorow (1998) for example, computed semantic relatedness as the length of the shortest path between synsets over the depth of the taxonomy. Wu and Palmer (1994) also used the hyponym tree to calculate relatedness by using the depth of the words in the taxonomy and the depth of the least common superconcept between the two words. Hirst and St-Onge (1998), on the other hand, used the lexical chains between words based on their synsets and the semantic edges that connect them. In addition to using the hypernym relations, they classified the relations into classes: "extra strong" for identical words, "strong" for synonyms, "medium strong" for when there is a path between the two, and "not related" for no paths at all. The semantic measurement is then based on the path length and the path direction changes. Tsatsaronis et al. (2010) used a combination of semantic path length, node depth in the hierarchy, and the types of the semantic edges that compose the path.

Figure 1: Example of the semantic network around the word *car*.



On the other hand, corpus-based approaches rely mainly on distributional properties of words learned from a large corpus to compute semantic relatedness. Such as the work of Finkelstein et al. (2001) that used Latent Semantic Analysis, and the work of Strube and Ponzetto (2006) and Gabrilovich and Markovitch (2007), which both used the distributional hypothesis on Wikipedia. Finally, hybrid approaches use a combination of corpus-based and lexicon-based methods. For example, the approach proposed by Hughes and Ramage (2007) used a random walk method over a lexicon-based semantic graph supplemented with corpus-based probabilities. Another example is the work of Agirre et al. (2009) that used a supervised machine learning approach to combine three methods: WordNet-based similarity, a bag of word based similarity, and a context window based similarity.

The approach presented in this paper belongs to the lexicon-based category. However, as opposed to the typical lexicon-based approaches described above, our approach uses all 26 semantic relations found in WordNet in addition to information found in glosses. The novelty of this approach is that these relations are used to create an explicit semantic network, where the edges of the network representing the semantic relations are weighted according to the type of the semantic relation. The semantic relatedness is computed as the lowest cost path between a pair of words in the network.

3 Our Approach

Our method to measure semantic relatedness is based on the idea that the types of relations that relate two concepts are a suitable indicator of the semantic relatedness between the two. The type

of relations considered includes not only the hyponym/hypernym relations but also all other available semantic relations found in WordNet in addition to word definitions.

3.1 WordNet's Semantic Network

To implement our idea, we created a weighted and directed semantic network based on the content of WordNet. To build the semantic network, we used WordNet 3.1's words and synsets as the nodes of the network. Each word is connected by an edge to its synsets, and each synset is in turn connected to other synsets based on the semantic relations included in WordNet. In addition each synset is connected to the content words contained in its gloss. For example, Figure 1 shows part of the semantic network created around the word *car*. In this graph, single-line ovals represent words, while double-line ovals represent synsets.

By mining WordNet entirely, we created a network of 265,269 nodes connected through a total of 1,919,329 edges. The nodes include all words and synsets, and the edges correspond to all 26 semantic relations in WordNet in addition to the relation between a synset and every content word of a synset definition.

3.2 Semantic Classes of Relations

To compute the semantic relatedness between nodes in the semantic network, it is necessary to take into consideration the semantic relation involved between two nodes. Indeed, WordNet's 26 semantic relations do not contribute equally to the semantic relatedness between words. The *hypernym* relation (relation #2), for example, is a good indicator of semantic relatedness; while the relation of *member of this domain - topic* (relation #15) is less significant. This can be seen in Fig-

Category	Weight	Semantic Relations in WordNet
<i>Similar</i>	α	antonym, cause, entailment, participle of verb, pertainym, similar to, verb group
<i>Hypernym</i>	$2 \times \alpha$	derivationally related, instance hypernym, hypernym
<i>Sense</i>	$4 \times \alpha + \beta$	lemma-synset
<i>Gloss</i>	$6 \times \alpha$	lemma-gloss content words
<i>Part</i>	$8 \times \alpha$	holonym (part, member, substance), inverse gloss, meronym (part, member, substance)
<i>Instance</i>	$10 \times \alpha$	instance hyponym, hyponym
<i>Other</i>	$12 \times \alpha$	also see, attribute, domain of synset (topic, region, usage), member of this domain (topic, region, usage)

Table 1: Relations Categories and Corresponding Weights.

ure 1, for example, where the word *car* is more closely related to *Motor vehicle* than to *Renting*. In order to determine the contribution of each relation, we compared a manually created set of 210 semantic relations for their degree of relatedness. For example, for the concept *car* we have compared the sense of *automobile* with the hypernym *motor vehicle*, the gloss word *wheel*, the part meronym *air bag*, the member of this topic *renting*, and another sense of *car* such as a *cable car*. This comparison has led us to classify the relations into seven categories, and rank these categories from the most related category to the least related one as follows: *Similar* (highest contribution), *Hypernym*, *Sense*, *Gloss*, *Part*, *Instance*, and *Other* (lowest contribution). By classifying WordNet’s relations into these classes, we are able to weight the contribution of a relation based on the class it belongs to, as opposed to assigning a contributory weight to each relations. For example, all relations of type *Similar* will contribute equally to the semantic relatedness of words, and will contribute more than any relations of the class *Hypernym*. Table 1 shows the seven semantic categories that we defined, their corresponding weight, and the WordNet relations they include. The weights¹ were simply assigned as a multiple of a small value α , representing the lowest weight, and an addition of 2 for each multiplier in the list in order to represent a higher cost of the less related categories. Let us describe each category in detail.

The category *Similar* includes WordNet’s relations of *antonym*, *cause*, *entailment*, *similar to*, *participle of verb*, *pertainym* and *verb group*. This

¹The weight can be seen as the cost of traversing an edge; hence a lower weight is assigned to a highly contributory relation.

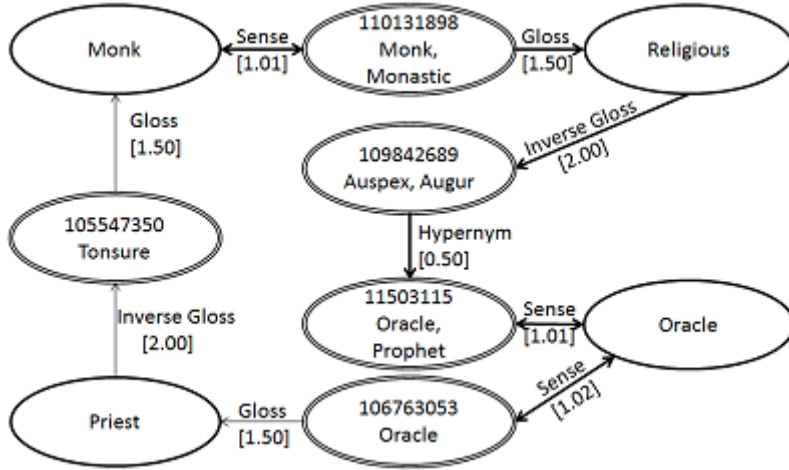
class of relations includes relations that are the most useful to compute semantic relatedness as per our manual corpus analysis and are the rarest available relations in the semantic network and hence was assigned the lowest weight of all categories of relations: α .

The second category of semantic relations is the *Hypernym* which includes WordNet’s relations of *hypernym*, *instance hypernym* and *derivationally related*. Being less important than the *similar* relations to compute relatedness, as shown in Table 1, the *Hypernym* category was assigned a weight of $(2 \times \alpha)$.

The *Sense* category represents the relationship between a word and its synset. Because a word can belong to several synsets, in order to favor the most frequent senses as opposed to the infrequent ones, the weight of this category is modulated by a factor β . Specifically, we use $(4 \times \alpha + \beta)$, where β is computed as the ratio of the frequency of the sense number in WordNet over the maximum number of senses for that word.

The fourth category of semantic relations is the *Gloss* that covers relations between synsets and their glosses. A synset gloss contains a brief definition of the synset, which usually consists of a genus (or type) and one or more differentia (or what distinguishes the term from the genus). The genus relations is explicitly defined in WordNet as a hypernym relation, however the differentia is most of the time not defined. The differentia includes essential attributes to the term being defined, that makes it more semantically related to the main term than other attributes. For this reason, we explicitly included those relations in the semantic network. For example, the gloss of the synset #102961779 *car, auto, automobile ... is a*

Figure 2: Lowest Cost Path Between the Words *Monk* and *Oracle*.



motor vehicle with four wheels, the hypernym of this synset is a *motor vehicle*, and the differentia is *four wheel*. There is no semantic relation explicitly defined in WordNet between *car* and *four wheel*, nor is there a relation with *wheel*. Even if a meronymy relation existed with *wheel* existed in WordNet, it also should be more related to it than the rest of the meronymy relations as it is a defining attribute. To include such relations to the semantic network, we create an edge between every content word in the gloss and the synset, but only consider words that have an entry in the lexicon. As this is a simplistic approach of adding the gloss relations, we gave it a high weight of $(6 \times \alpha)$, but less than the next category covering meronymy relations. The inverse of this edge (from a gloss word to a synset) is also included, but is considered to be less related and thus included in the next category.

The fifth category is the *Part* category that includes *holonymy*, *meronymy*, and *inverse gloss* relations which are all weighted as $(8 \times \alpha)$.

The sixth category, the *Instance* category, only includes the *hyponymy* and *instance of hyponymy* relations that are weighted as $(10 \times \alpha)$.

Finally, all others relations available in WordNet are grouped under the last category *Other* and given the maximum weight of $(12 \times \alpha)$.

3.3 Calculation of Semantic Relatedness

Given the weighted semantic network, the semantic relatedness, $S(w_1, w_2)$, between two words w_1 and w_2 is computed essentially as the weight of the

lowest cost path² between the two words. However, because the network is directed, the lowest cost from w_1 to w_2 , $P_{min}(w_1, w_2)$, may be different than from w_2 to w_1 , $P_{min}(w_2, w_1)$. To account for this, we therefore consider the semantic relatedness $S(w_1, w_2)$ to be equal to the highest relatedness score in either direction. More formally, the semantic relatedness between w_1 and w_2 is defined as:

$$S(w_1, w_2) = \max \left(\frac{M - (P_{min}(w_1, w_2) - K)}{M}, \frac{M - (P_{min}(w_2, w_1) - K)}{M} \right)$$

Where, M is a constant representing the weight after which two words are considered unrelated, and K is constant representing the weight of true synonyms. In our implementation, we have set $M = 2 \times (12 \times \alpha)$ corresponding to the maximum of traveling twice the relation with the highest weight, and $K = 2 \times (4 \times \alpha)$ corresponding to the minimum of traveling from a word to its sense and back to the word itself.

3.4 An Example

Figure 2 shows an extract of the network involving the words *Monk* and *Oracle*. The lowest cost path from *Monk* to *Oracle* is highlighted in bold. As the figure shows, the word *Monk* is connected with a *Sense* relation to the synset #110131898 [*Monk, Monastic*]. As indicated in Table 1, the weight of this relation is computed as $(4 \times \alpha + \beta)$. Because

²The lowest cost path is based on an implementation of Dijkstras graph search algorithm (Dijkstra, 1959)

Approach	Category	Pearson
(Gabrilovich and Markovitch, 2007)	Corpus	0.72
(Hirst and St-Onge, 1998)	Lexicon	0.74
(Wu and Palmer, 1994)	Lexicon	0.78
(Resnik, 1995)	Hybrid	0.80
(Leacock and Chodorow, 1998)	Lexicon	0.82
(Lin, 1998)	Hybrid	0.83
(Bollegala et al., 2007)	Corpus	0.83
(Jiang and Conrath, 1997)	Hybrid	0.85
(Tsatsaronis et al., 2010)	Lexicon	0.86
(Jarmasz and Szpakowicz, 2003)	Lexicon	0.87
(Hughes and Ramage, 2007)	Lexicon	0.90
(Alvarez and Lim, 2007)	Lexicon	0.91
(Yang and Powers, 2005)	Lexicon	0.92
(Agirre et al., 2009)	Hybrid	0.93
Our approach	Lexicon	0.93

Table 2: Pearson Correlation of Various Approaches on the Miller and Charles Data Set.

this synset is the first sense (the most frequent sense given by WordNet) for the word *Monk*, then $(\beta = 1/75 = 0.01$, where 75 is the maximum number of senses for a word in WordNet. If α is set to 0.25, then, as shown in Figure 2, the weight of this edge is computed ($4 \times 0.25 + 0.01 = 1.01$). The synset #11013898 [*Monk, Monastic*] is connected to the word *Religious* through a *Gloss* relation type. In WordNet, the gloss of this synset is: *a male religious living in a cloister and devoting himself to contemplation and prayer and work*. The content words are: *male, religious, live, cloister, devote, contemplation, prayer, and work*, which are each related to this synset with the weight set to $(6 \times \alpha = 1.5)$.

Overall, the weight of the lowest cost path $P_{min}(Monk, Oracle)$ is hence equal to the sum of the edges shown in Figure 1 ($1.01+1.50+2.00+0.50+1.01 = 6.02$). As the figure shows, in this example, $P_{min}(Monk, Oracle)$ is identical to $P_{min}(Oracle, Monk)$. With the constants M set to 6 and K to 2, $S(Monk, Oracle)$ will therefore be $(6-(6.02-2))/6 = 0.33$.

4 Evaluation

To evaluate our approach, we used two types of benchmarks: using human ratings and using synonym tests.

4.1 Evaluation using Human Ratings

In their study on semantic similarity, Miller and Charles (1991) (M&C) gave 38 undergraduate students 30 pairs of nouns to be rated from 0, for no similarity, to 4, for perfect synonymy. The noun pairs were chosen to cover high, intermediate, and low level of similarity and are part of an earlier

study Rubenstein and Goodenough (1965) (R&G) which contained 65 pairs of nouns. The M&C test gained popularity among the research community for the evaluation of semantic relatedness. The evaluation is accomplished by calculating the correlation between the average student’s ratings and one’s approach. The commonly used correlation measurement for this test is the Pearson correlation measurement (Pearson, 1900), but some have also used the Spearman ranking coefficient (Spearman, 1904) as an evaluation measurement. Our approach achieved a Pearson correlation of 0.93 and a Spearman correlation of 0.87 with the M&C data set. In addition, it achieved 0.91 Pearson correlation and 0.92 Spearman correlation on the R&G data set.

For comparative purposes, Table 2 shows the Pearson correlation of several previous approaches to semantic relatedness measures against the same data set, as reported in their respective papers. For information, the table indicates the type of approach used: lexicon-based method, corpus-based method, or hybrid. As Table 2 shows, most other approaches achieve a correlation around 85%, while a few achieve a correlation above 90%. These results do not seem to be influenced by the type approach. Our approach compares favorably to the state of the art in the field on the Miller and Charles data set, with a high correlation of 93%. Our result is higher than any other lexicon based approach, however it must be noted that the Miller and Charles Data Set is quite small for empirical analysis.

WordSimilarity-353 is another set of human ratings that was introduced by Finkelstein et al.

Approach	Category	Spearman
(Strube and Ponzetto, 2006)	Corpus	0.48
(Jarmasz and Szpakowicz, 2003)	Lexicon	0.55
(Hughes and Ramage, 2007)	Lexicon	0.55
(Finkelstein et al., 2001)	Hybrid	0.56
(Gabrilovich and Markovitch, 2007)	Corpus	0.75
(Agirre et al., 2009)	Hybrid	0.78
Our approach	Lexicon	0.50

Table 3: Spearman Correlation of Various Approaches on WordSimilarity-353 Data Set.

Approach	Category	Accuracy
(Resnik, 1995)	Hybrid	32.66%
(Leacock and Chodorow, 1998)	Lexicon	36.00%
(Lin, 1998)	Hybrid	36.00%
(Jiang and Conrath, 1997)	Hybrid	36.00%
(Hirst and St-Onge, 1998)	Lexicon	62.00%
(Turney, 2001)	Corpus	74.00%
(Terra and Clarke, 2003)	Corpus	80.00%
(Jarmasz and Szpakowicz, 2003)	Lexicon	82.00%
(Tsatsaronis et al., 2010)	Lexicon	82.00%
Our Approach	Lexicon	84.00%

Table 4: Results with the ESL Data Set.

(2001). The data set is much larger than the Miller and Charles Data Set and includes 353 pairs of words, each rated by 13 to 16 subjects who were asked to estimate the relatedness of the words on a scale of 0 for “totally unrelated words” to 10 for “very much related or identical words”. The common practice with this data set is to use the Spearman coefficient.

Table 3 shows various approaches and their corresponding Spearman correlation as described in the literature. On this data set, our approach achieved a correlation of 0.50, which is quite lower than the current state of the art. After analysing our results, we identified several reasons why our approach did not perform as expected. First, all lexicon based methods seem to perform poorly on this data set because it includes a number of named entities that are typically not available in a lexicon. For example, in the word pair: (*Maradona – football*), the word *Maradona* does not appear in WordNet, hence favoring corpus-based and hybrid approaches. Another difficulty is the high variance of human ratings for some word pairs, which could be due to the subjectivity required for this task, or the fact that the subjects who rated the data set were not native English speakers. That being said, perhaps the most important factors for the poor performance is that most of the pairs in that data set require general world knowledge that is not usually available in a lexicon. Nevertheless, other approaches were able to achieve a high correlation

with this data set such as the machine learning approach of Agirre et al. (2009) that achieved a high correlation of 0.78.

4.2 Evaluation using Synonym Tests

To test the approach further, we also evaluated it on synonym identification tests. This type of test includes an initial word and a set of options from which the most synonymous word must be selected.

The first synonym test that we experimented with is the English as a Second Language (ESL) test. The test set was first used by Turney (2001) as an evaluation of algorithms measuring the degree of similarity between words. The ESL test includes 50 synonym questions and each having four choices. The following is an example question taken from ESL data set:

```
Text: A rusty nail is not as strong as
a clean, new one.
Stem: rusty
Choices:
(a) corroded
(b) black
(c) dirty
(d) painted
Solution: (a) corroded
```

The results of our approach, along with other approaches, on the 50 ESL questions are shown

Approach	Category	Accuracy
(Resnik, 1995)	Corpus	20.31%
(Leacock and Chodorow, 1998)	Lexicon	21.88%
(Lin, 1998)	Hybrid	24.06%
(Jiang and Conrath, 1997)	Hybrid	25.00%
(Landauer and Dumais, 1997)	Corpus	64.38%
<i>Average non-English US college applicant</i>	<i>Human</i>	<i>64.50%</i>
(Padó and Lapata, 2007)	Corpus	73.00%
(Hirst and St-Onge, 1998)	Lexicon	77.91%
(Jarmasz and Szpakowicz, 2003)	Lexicon	78.75%
(Terra and Clarke, 2003)	Corpus	81.25%
(Ruiz-Casado et al., 2005)	Corpus	82.55%
(MaTveeva et al., 2007)	Corpus	86.25%
(Tsatsaronis et al., 2010)	Lexicon	87.50%
(Rapp, 2003)	Corpus	92.50%
(Turney et al., 2003)	Hybrid	97.50%
(Bullinaria and Levy, 2012)	Corpus	100.00%
Our Approach	Lexicon	91.25%

Table 5: Results with the TOEFL Data Set.

in Table 4. The results are measured in terms of accuracy - the percentage of correct responses by each approach. Our approach has achieved an accuracy of 84% on the ESL test, which is slightly better than the reported approaches in the literature. It should be noted that sometimes the difference between two approaches belonging to the same category are merely a difference in the data set used (Corpus or Lexicon) rather than a difference in the algorithms. Also, the ESL question set includes a sentence to give a context for the word, which some approaches (e.g. (Turney, 2001)) have used as an additional information source; we on the other hand, did not make use of the context information in our approach.

The second synonym test that we used is the Test of English as a Foreign Language (TOEFL) test. The test was first used by Landauer and Dumais (1997) as an evaluation for the algorithm measuring the degree of similarity between words. The TOEFL test includes 80 synonym questions each having four choices. The following is an example TOEFL question:

```

Stem: levied
Choices:
(a) imposed
(b) believed
(c) requested
(d) correlated
Solution: (a) imposed

```

The results on the 80 TOEFL questions are

shown in Table 5, which also includes the results of other approaches for comparative purposes. Here again, the results are reported in terms of accuracy. As with the previous experiments, the category of the approach does not seem to have an impact on the results. It should be noted, however, that some of the approaches have been tuned specifically for the TOEFL questions.

Table 5 also includes an entry for the “Average non-English US college applicant” of 64.5%. The score that was originally reported in Landauer and Dumais (1997) is 52.5% for college applicants, however this figure penalizes random guessing by subtracting a penalty of 1/3. To provide a more fair comparison, this penalty has been removed leading to a score of 64.5%. Our approach has achieved an accuracy of 91.25% on the TOEFL test, which is better than any of the reported lexicon based approaches.

5 Conclusion

In this paper we have presented a state of the art semantic relatedness approach that is based on a weighted semantic network. The novelty of the approach is that it uses all 26 relations available in WordNet, along with information found in glosses, and the contribution of each relation to compute the semantic relatedness between pairs of words. This information was mined from WordNet to create a large semantic network consisting of 265,269 concepts connected through a total of 1,919,329 relations. To account for the different contribution of each semantic relation, each edge of the semantic network is assigned a weight according to the category of its semantic

relation. All 26 of WordNet's semantic relations and the glosses have been categorised into seven categories, each carrying a weight. Computing the semantic relatedness between two words is now seen as computing the weight of the lowest cost path between the two words in the semantic network. However, because the semantic network is directed, we take the maximum weight among both directions that link the two words.

We evaluated the approach with several benchmarks and achieved interesting results, often among the best systems. Specifically, the approach achieved a Pearson correlation of 0.93 with the M&C human ratings, a Spearman correlation of 0.50 on the Word Similarity353 data set, an accuracy of 84% on the ESL synonym test, and an accuracy of 91.25% on the TOEFL synonym test. Future work includes performing additional experiments to find the best values for the parameters α , β , and the class weights. Currently, the value of these parameters have been set empirically over several small experiments, but a more formal training to find the best combination of these parameters is necessary. In addition, the semantic information that we tried to include from the gloss have all been categorized into one single category with a unique weight. However, this should be modified to categorize the gloss relations further. For example, extensional types of definitions that specify extensions in the definition are usually less related than differentiating attributes. For example, in the *glow* definition: *have a complexion with a strong bright color, such as red or pink*, the extensions *red or pink* should have a lower relatedness than the attribute *bright* to the concept *glow*. Finally, some important issues in computing lexicon based semantic similarity in general must still be addressed. In particular, all words that are related to another word by the same path will have the same semantic relatedness. For example, a *take out* will have the same semantic relatedness to its sister terms *impulse-buy* and *buy out* by most of the lexicon based approaches as they all have the same path length and depth, however a *take out* can be more of an *impulse buy* than a *buy out* and thus should be more related. In addition, most lexicons do not have pragmatic relations that are important for calculating semantic relatedness, for example the pair *movie* and *popcorn* from the WordSimilarity data set has an average semantic relatedness by 13 different annotators of

6.19/10. however, the lowest cost path between the two in WordNet is through the *physical entity* concept, which means that a *movie* will have a shorter path to a *popcorn* through the *product* concept than to *popcorn*.

Acknowledgments

The authors would like to thank the anonymous reviewers for their comments on a previous version of this paper.

References

- Eneko Agirre, Enrique Alfonseca, Keith Hall, Jana Kravalova, Marius Pasca, and Aitor Soroa. 2009. A study on similarity and relatedness using distributional and wordnet-based approaches. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 19–27, Boulder, June.
- Marco A. Alvarez and SeungJin Lim. 2007. A graph modeling of semantic similarity between words. In *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)*, pages 355–362, Irvine, September.
- Danushka Bollegala, Yutaka Matsuo, and Mitsuru Ishizuka. 2007. Measuring semantic similarity between words using web search engines. In *Proceedings of the Sixteenth International World Wide Web Conference (WWW2007)*, volume 7, pages 757–786, Banff, May.
- John A. Bullinaria and Joseph P. Levy. 2012. Extracting semantic representations from word co-occurrence statistics: stop-lists, stemming, and SVD. *Behavior Research Methods*, 44:890–907, September.
- Edsger W. Dijkstra. 1959. A note on two problems in connexion with graphs. *Numerische mathematik*, 1(1):269–271.
- Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
- Lev Finkelstein, Evgeniy Gabrilovich, Yossi Matias, Ehud Rivlin, Zach Solan, Gadi Wolfman, and Eytan Ruppin. 2001. Placing search in context: The concept revisited. In *WWW '01: Proceedings of the 10th international conference on World Wide Web*, pages 406–414, New York, May.
- Evgeniy Gabrilovich and Shaul Markovitch. 2007. Computing semantic relatedness using wikipedia-based explicit semantic analysis. In *Proceedings of the 20th international joint conference on Artificial intelligence (IJCAI 2007)*, pages 1606–1611, Hyderabad, January.

- Graeme Hirst and David St-Onge. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. *WordNet An electronic lexical database*, pages 305–332, April.
- Thad Hughes and Daniel Ramage. 2007. Lexical semantic relatedness with random graph walks. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing - Conference on Computational Natural Language Learning (EMNLP-CoNLL)*, pages 581–589, Prague, June.
- Mario Jarmasz and Stan Szpakowicz. 2003. Roget's thesaurus and semantic similarity. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2003)*, pages 212–219, Borovets, September.
- Jay J Jiang and David W Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proceedings of International Conference on Research in Computational Linguistics*, pages 19–33, Taipei, Taiwan, August.
- Thomas K. Landauer and Susan T. Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, 104(2):211–240.
- Claudia Leacock and Martin Chodorow. 1998. Combining local context and wordnet similarity for word sense identification. *WordNet: An electronic lexical database*, 49(2):265–283.
- Dekang Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th international conference on Machine Learning (ICML 1998)*, volume 1, pages 296–304, Madison, July.
- Irina MaTveeva, Gina-Anne Levow, and Ayman Farahat. 2007. Term representation with generalized latent semantic analysis. *Recent Advances in Natural Language Processing IV*, 292:45.
- George A. Miller and Walter G. Charles. 1991. Contextual correlates of semantic similarity. *Language & Cognitive Processes*, 6(1):1–28.
- Sebastian Padó and Mirella Lapata. 2007. Dependency-Based Construction of Semantic Space Models. *Computational Linguistics*, 33(2):161–199.
- Karl Pearson. 1900. Mathematical contributions to the theory of evolution. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 195:1–405.
- Reinhard Rapp. 2003. Word Sense Discovery Based on Sense Descriptor Dissimilarity. In *Proceedings of the Ninth Machine Translation Summit*, pages 315–322, New Orleans, September.
- Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *International Joint Conference for Artificial Intelligence (IJCAI-95)*, pages 448–453, Montreal, August.
- Herbert Rubenstein and John B. Goodenough. 1965. Contextual correlates of synonymy. *Communications of the ACM*, 8(10):627–633.
- Maria Ruiz-Casado, Enrique Alfonseca, and Pablo Castells. 2005. Using context-window overlapping in synonym discovery and ontology extension. In *Proceedings of Recent Advances in Natural Language Processing (RANLP 2005)*, Borovets, September.
- Charles Spearman. 1904. The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.
- Michael Strube and Simone Paolo Ponzetto. 2006. WikiRelate! Computing semantic relatedness using Wikipedia. In *Proceedings of the National Conference on Artificial Intelligence*, volume 21, page 1419, Boston, July.
- Egidio Terra and Charles LA Clarke. 2003. Frequency estimates for statistical word similarity measures. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (HLT-NAACL '03)*, volume 21, pages 165–172, Edmonton, May.
- George Tsatsaronis, Iraklis Varlamis, and Michalis Vazirgiannis. 2010. Text relatedness based on a word thesaurus. *Journal of Artificial Intelligence Research*, 37(1):1–40.
- Peter D. Turney, Michael L. Littman, Jeffrey Bigham, and Victor Shnayder. 2003. Combining Independent Modules to Solve Multiple-choice Synonym and Analogy Problems. In *Recent Advances in Natural Language Processing (RANLP 2003)*, pages 101–110, Borovets, September.
- Peter Turney. 2001. Mining the Web for Synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings of the Twelfth European Conference on Machine Learning (ECML-2001)*, pages 491–502, Freiburg, Germany, September.
- Zhibiao Wu and Martha Palmer. 1994. Verbs semantics and lexical selection. In *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138, New Mexico, June.
- Dongqiang Yang and David MW. Powers. 2005. Measuring semantic similarity in the taxonomy of wordnet. In *Proceedings of the Twenty-eighth Australasian conference on Computer Science*, volume 38, pages 315–322, Newcastle, January.