Determining the Semantic Orientation of Opinion Words using Typed Dependencies for Opinion Word Senses and Sentiwordnet Scores from Online Product Reviews

Teja Santosh

Assistant Professor in CSE, GITAM University, Hyderabad, India

Abstract
Opinion word orientation is a sub discipline of text classification concerned with positive or negative association of the opinionated terms. It has a broad range of applications from tracking users' opinions about products or about political candidates as conveyed in online forums, to customer relationship management to reduce the churn rate. The synonymy relation graph is used to determine the orientation of the adjectives present in the text available in the product reviews data corpus. It considers the minimum path length along with the connected WordNet synsets. The synonymy relation graph limits in determining the number of orientations of the opinion words present in the synonym graph's minimal path. To evaluate opinion orientation of any adjective from the dataset, the synonymy relation graph of WordNet is replaced with the Sentiwordnet 3.0 scores of the opinion word. These scores are provided to the opinion word by finding the contextual clues surrounding the opinion word to disambiguate its sense. The contextual clues are finalized based on the typed dependencies grammatical relations. The distance between the opinion word and the context insensitive seed term (good/bad) is computed by calculating the difference between these scores. This paper addresses the limitation identified in the synonymy relation graph of WordNet based determination of semantic orientation. This improves the accuracy of the determined opinion word orientations.

Keywords:
Opinion Mining, Text classification, seed terms, opinion word orientation, semantic orientation

1. Introduction

The capability to ascertain the idea of finding the distance in terms of similarity between the words and concepts is the most important task in the semantic analysis research. The distances between the adjectives in the WordNet lexical database [1] has greatly profited from this idea of research. In the task of opinion mining, the adjectives are identified as opinion words [2]. In order to determine the orientation of the opinion word, the synonymy relation graph [1] of WordNet is analyzed. The synonymy relation graph provides the minimal path length [3] between the current opinion word under consideration and the seed terms (good or bad).These seed terms lack sensitivity to the context [4] in the reviews where they are written. The path in the synonym relation graph between the opinion word and seed term is obtained by counting the number of synonym words belonging to same synset connectively till the seed term. This gives the distance between the opinion word and the seed term. The distance between the seed terms is also calculated in the same manner.

The synonymy relation graph limits in determining the number of orientations of the opinion words present in the synonym graph's minimal path. To evaluate opinion orientation of any adctive from the dataset, the synonymy relation graph of WordNet [5] is replaced with the Sentiwordnet 3.0 [6] scores of the opinion word. These scores are provided to the opinion word by finding the contextual clues surrounding the opinion word to disambiguate its sense. The contextual clues are finalized based on the typed dependencies grammatical relations [17]. The distance between the opinion word and the seed term is computed by calculating the difference between these scores. The accuracy of the determined opinion word orientations with the help of Sentiwordnet scores by replacing the synonymy relation graph of WordNet is estimated.

Several subtasks involved in carrying out the semantic orientation of the opinion word. These are the identification of adjective as opinion word with the help of opinion lexicon. Next, assigning Sentiwordnet 3.0 scores to the identified opinion words based on its sense from the adjectives category. Further, using of these scores in the Semantic Orientation (SO) measure for determining the orientation of the opinion word. The distance metric is changed to the difference between the Sentiwordnet scores of the words.

The paper is organized as follows: related work is described in section 2 and the proposed approach is explained in section 3, the experiments conducted to evaluate the proposed approach is presented in section 4 and

Manuscript received February 5, 2026
Manuscript revised February 20, 2026
https://doi.org/10.22937/IJCSNS.2026.26.2.9

the results of opinion orientation of the proposed method compared with the existing approaches is discussed in section 5 and finally the conclusions and the scope for future work are presented in section 6.

2. Related Works

The distance based similarity measures based on WordNet is considered as a major research work for more than a decade. Quite a number of researchers were focused their research on this particular subject. Rada et al. used [7] edge-counting approach on WordNet's taxonomy relations. Hirst and St-Onge extended [8] the idea of Rada et al. [7] by utilizing the path length to all relations in WordNet. Leacock and Chodorow considered [9] the path length of hyponymy relations (IS-A or HAS-PART relation in WordNet), while reducing the distances by the depth in the hierarchy. Resnik extended [10] the work of Leacock and Chodorow [9] by considering the information derived from word frequencies in the Brown corpus and combining the obtained information in the hyponymy relation hierarchy. Kamps et al. defined [1] a distance measure using the concepts of graph theory. They have concentrated on the shortest distance between two nodes (containing words). This is called as minimal path-length. The works on distance measures [7, 8, 9, 10] are only applicable to the hyponymy relations. This restriction makes distance measures applicable to the syntactic categories of noun and verb and not to the syntactic categories of adjectives and adverbs.

The semantic orientation of texts is an age old classical work for more than five decades. Osgood et al. identified [11] several pairs of bipolar adjectives that greatly influence the shift in the orientation of the opinion words. Hatzivassiloglou and McKeown attempted [12] to predict the orientation of the opinion words by analyzing the pairs of adjectives bounded by conjunctions. Turney and Littmann approached [4] the problem by first using a seed set to bootstrap the process of opinion word identification. Once the opinion words are identified, Pointwise Mutual Information (PMI) was calculated on the identified opinion word and the term in the seed set. The work on determining the orientation of the terms in [12] concentrated on pairs of adjectives bounded by conjunctions. The researchers were considered only 657/679 documents (labeled Positive/Negative) in which the adjectives bound by conjunctions are available from the Wall Street Journal (WSJ) corpus.

Kamps et al. focused [1] on the relations between the words defined in the WordNet. They calculated the relative distance from the two seed terms to the identified opinion word to determine the orientation of the opinion word. The work of solving the ambiguity of the terms that appear in both the Positive and Negative categories was never concentrated [1]. They were removed those terms from the sets and experimented on the reduced sets. The number of considered terms after removing the ambiguous entries is 1614/1982. They restricted the adjectives in the analysis to 663 from the total 3596 terms of Turney and Littmann as used in [4]. This is because the synonymy relation graph of WordNet evaluates only those adjectives that are in the path of the graph bounded with the seed terms at the ends of the graph.

All the works never concentrated on the detecting the contextual clues of the opinion word which are the useful indicators of the important facet namely sense of the opinion word. Assigning weights to the opinion words based on the context based sense provides a better way of evaluating the semantic orientation of those words. An alternate way of determining the semantic orientation of the terms is required by using the opinion word senses and the Sentiwordnet 3.0 scores. This allows including all the adjectives in the evaluation process. Also the distance measure is to be modified in such a way that the relative distance which is calculated to determine the semantic orientation uses the Sentiwordnet 3.0 scores.

3. Semantic Orientation of Opinion Words using Typed Dependencies for Opinion Word Senses and Seniwordnet Scores

The determination of semantic orientation of opinion words using typed dependencies for opinion words and sentiwordnet scores is presented in Figure 1. Input to the model is the online reviews. Initially, the incoming product reviews are pre-processed. The steps in pre-processing are namely review tokenization, stopwords removal and Part-of-Speech (PoS) tagging. The process of review tokenization divides the sentence into individual tokens. Then, the stopwords list is applied on the tokens to remove those words which carry no meaning in the analysis. The stop words are compiled from the reviews itself. This compilation is carried out by sorting the terms in the decreasing order of collection frequency and thereby hand-filtering those terms for their semantic content relative to the domain of the product reviews. Finally, Part of Speech (PoS) tagging is carried out on the list of filtered tokens to associate the unambiguous word categories with each of the token. The Stanford log-linear Part of Speech tagger is used [13] for tagging the tokens. The adjectives that are present either before or after to the identified product features are extracted. These adjectives are analyzed for opinions and orientations in the further steps.

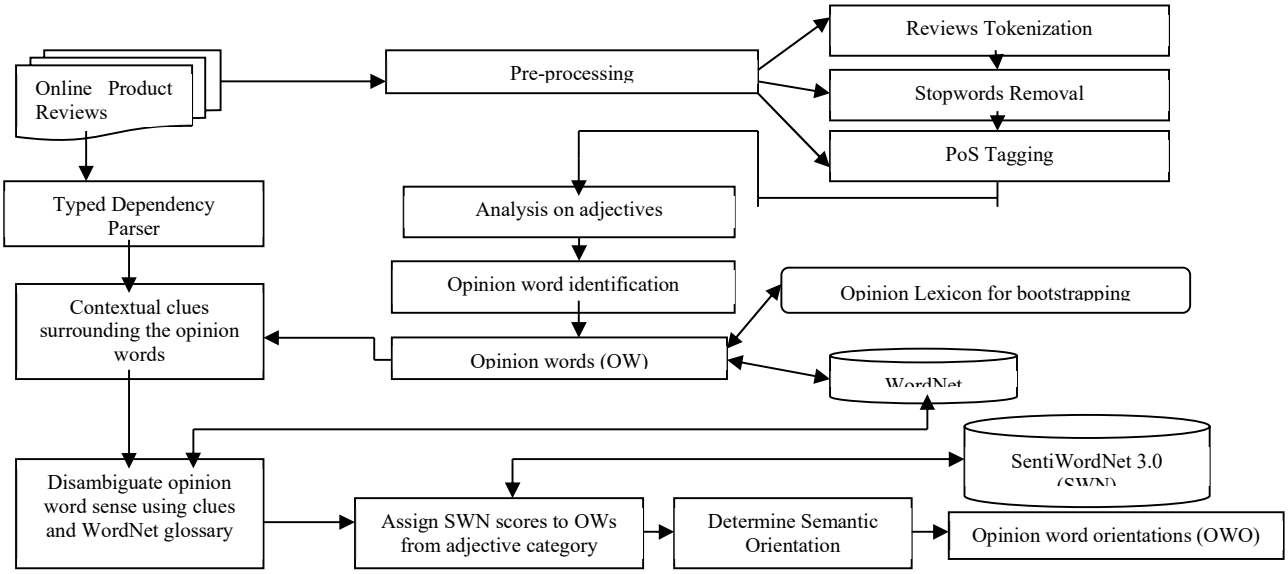

Figure 1. Proposed Model

The variations in the method for determining the orientation of a opinionated term using the opinion word senses and the Sentiwordnet scores were presented. The process is composed of following steps:

1. A standard opinion lexicon in which two sets of adjectives are present is considered as input for bootstrapping. These sets are representative of the two categories namely Positive and Negative. Two seed terms 'good' and 'bad' representative of the two categories are taken into consideration.

2. The sizes of the Positive and Negative adjective sets are increased by adding the synonyms of the available adjectives using WordNet.

3. The increased sizes of Positive and Negative adjective sets is used to compare with the obtained adjectives from the dataset. Once the dataset adjectives are matched with the opinion lexicon adjectives, then the dataset adjectives are considered as opinion words. This completes the identification of opinion words from the dataset.

4. The opinion word and the seed terms are assigned with the sentiment scores available under adjective category from Sentiwordnet. This is carried out by finding the contextual clues surrounding the opinion word. These contextual clues will help to disambiguate the sense of the opinion word. The contextual clues are finalized based on the typed dependency grammatical relations.

5. The distance between the opinion word and the seed term and the distance between the seed terms is calculated as given below.

$$distance(w_i,w_j) = sentiwordnetscore(w_i) - sentiwordnetscore(w_j)$$

where w_i is either the opinion word or the seed term and w_j is the seed term.

6. The semantic orientation (SO) of the opinion word is determined as given below.

$$SO(\text{opinion word}) = \frac{distance(\text{opinion word, bad})distance(\text{opinion word, good})}{distance(\text{good, bad})}$$

7. The opinion word is deemed to be positive if the orientation measurement is greater than zero, and negative otherwise.

Step 2 is based on the premise that the lexical relations used in this expansion task which define a relation of orientation. It is possible that two synonyms may have same orientation and two antonyms have opposite orientation. In step 4, the basic assumption is that the terms with a similar orientation tend to have similar glossaries. The similarity or difference between the opinion word and the seed term is based on identifying the appropriate senses in the context in

which the opinion word is written in the document. The senses of the seed term is going to change based on the context of the opinion word under analysis. The replacement of the number of synonyms in the synonymy graph with the sentiwordnet scores in step 5 enables to determine the orientation of any opinion word with the help of SO measure specified in step 6.

4. Experiments Conducted on Synonymy Relation Graph based so and Opinion Word Senses and the Sentiwordnet Scores based So

A. Expansion method for seed sets

The WordNet version 2.1 (WN) is used as the source of lexical relations because it is easy to use for automatic processing. From many lexical relations available in WN the synonymy, antonymy, hypernymy and hyponymy relations are chosen to explore the concept of orientation. These four lexical relations are restricted to a given Part-Of-Speech (PoS) i.e., adjective. This restriction is possible as WN relations are defined on word senses rather than on words and the WN word senses are PoS tagged.

B. Distance calculation using Minimum Path-Length in Synonymy Relation Graph

Recall the definition of Minimum Path Length (MPL) given by Kamps in [3] that two terms are n-related if there exists (n+1)-long sequence of terms such that the adjacent terms in the sequence belong to the same synset. Based on this definition the distance between the seed terms 'good' and 'bad' from WordNet in its synonymy relation graph is 4. In the same lines for example the opinion word 'honest' is evaluated for its semantic orientation, the distance between 'honest' and 'good' is 2 and the distance between 'honest' and 'bad' is 6. The sequence of terms that are synonymous between {'honest', 'bad'}, {'honest', 'good'} and {'good', 'bad'} analyzed from WN are tabulated in Table 1 below.

Table 1. Sequence of Terms that are Synonymous between Considered Terms from WN

{'honest', 'bad'}	*honest* ----> dependable → *dependable* ----> good → *good* ----> sound → *sound* ----> heavy → *heavy* ----> big → *big* ----> bad
{'honest', 'good'}	*honest* ----> dependable → *dependable* ----> good
{'good', 'bad'}	*good* ---->sound → *sound* ----> heavy → *heavy* ----> big → *big* ----> bad

In the above table, the arrow '---->' represents the relationship between the word w_0 and the same synset word w_i. The arrow ' → ' represents the relationship between the word w_i and the same synset word w_{i+1}.

C. Semantic Orientation of opinion words using the MPL distance values

The semantic orientation of opinion words is carried out by using the calculated MPL distance values in the formula

$$SO(\text{opinion word}) = \frac{d(\text{opinion word, 'bad'}) - d(\text{opinion word, 'good'})}{d(\text{'good', 'bad'})}$$

For the opinion word 'honest' the SO is $\frac{d(\text{'honest', 'bad'}) - d(\text{'honest', 'good'})}{d(\text{'good', 'bad'})} = \frac{6 - 2}{4} = +1.$

The opinion word 'honest' is positive as the orientation measurement is greater than zero.

D. Semantic Orientation of opinion words using typed dependencies for opinion word senses and the Sentiwordnet scores

The semantic orientation of opinion words is carried out by first understanding the sense and the corresponding glossary of the word. Once the glossary is matched with the context, the associated sentiwordnet score corresponding to the positive/negative categories were assigned to the opinion word. For the opinion word 'honest' in the example sentence 'For the first time in their relationship, he had been completely honest with her', the sense is first disambiguated from WordNet. To do this, the contextual clues are identified from the sentence using sentence dependency parsing. The

typed dependency parsing for the given sentence is shown in Figure 2 below.

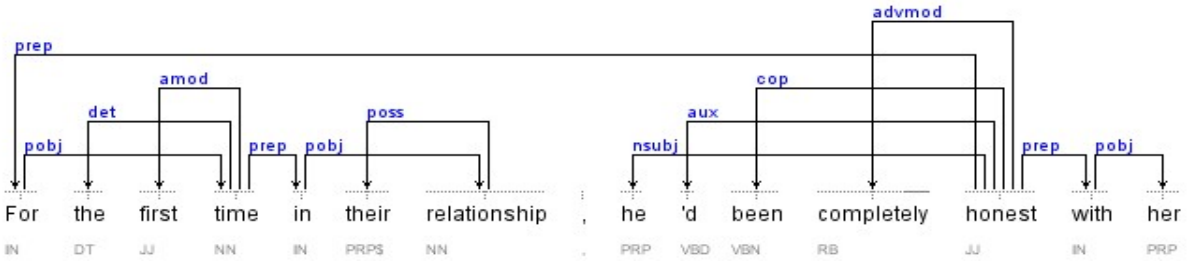

Figure 2. Typed dependencies in the sentence using sentence dependency parsing

From the generated word dependencies in the sentence the grammatical relations advmod(), prep(), pobj() and poss() provide contextual clues for the opinion word 'honest'. The window size of 2n+1 words [14] surrounding the opinion word and including the opinion word is selected as the contextual clues. The contextual clues for the sentence are {completely, with, her, their, relationship}. With these clues the sense value for 'honest' is finalized as 3 from out of 7 different senses by using WordNet sense similarity software package [18]. The gloss is 'worthy of being depended on' confirms the context of the opinion word. The word synonyms are dependable, reliable, true. The sentiwordnet positive score is 0.5. The seed term 'good' has sense 1 for this opinion word. The sentiwordnet positive score is 0.75. The seed term 'bad' has sense 1 for this opinion word. The sentiwordnet negative score is 0.625. The sense, gloss and the sentiwordnet score for the current opinion word and the seed terms are tabulated in Table 2 below.

Table 2 Terms Sense, Gloss from Wordnet and Their Sentiwordnet Scores

Opinion Word/Seed Word	Sense	WordNet Gloss	Pos. Score(p)/Neg. Score(n)
Honest	3	worthy of being depended on	p=0.5
Good	1	having desirable or positive qualities especially those suitable for a thing specified	p=0.75
Bad	1	having undesirable or negative qualities	n=0.625

Now for the opinion word 'honest' the SO is distance('honest', 'bad') – distance('honest', 'good')

distance('good', 'bad')

$$SO(honest) = \frac{(0.5-0.625) - (0.5-0.75)}{(0.75-0.625)} = -0.125 - (-0.25)/0.125 = 0.125/0.125 = +1.$$

The opinion word 'honest' is positive as the orientation measurement is greater than zero.

5. Results

The electronic product reviews corpus which was obtained from Amazon is used for this experiment. This corpus consists of eleven consumer product reviews. Three product reviews were considered for conducting this experiment. Rand Mcnally IntelliRoute TND 700 Truck GPS device, Nook Tablet and LCD mounting arm are the products for which the reviews were considered for analysis. The labels provided to these three datasets are D1, D2 and D3. Table 3 presents the details of the dataset used for this experiment.

Table 3 Dataset Details

Part-of-speech	CC in D2
Noun	{setup, purchase, hdmi port, note, nook, software, adapter, product, price, lieu}
Verb	{recommend, viewing, work}
Adverb	{loosely, just}
Preposition	{like}

Table 4 Percentage of Adjectives in the Datasets

Document attributes	Values
Number of review documents	2000
Minimum sentences per review	1
Maximum sentences per review	26
Minimum number of words per review sentence	24
Maximum number of words per review sentence	32.6

The opinion lexicon used for bootstrapping the process of opinion words orientation is Hu and Liu [15] compiled opinion lexicon base seed sets. Table 4 presents the percentage of adjectives identified from the datasets after removing wrongly tagged nouns, verbs, adverbs as adjectives for this experiment.

Table 5 Some of the Contextual Clues in the Three Datasets

Information	Statistics on D1	Statistics on D2	Statistics on D3
Percentage of adjectives	96.16%	98.62%	98.62%

Although the review sentence consists of a large number of words, only few of these words are useful to disambiguate the sense of the adjective. The major contextual clues surrounding the adjectives are namely Noun, Verb, Adverb and Preposition in the review sentences. The NearestWords algorithm used [14] in this work collects the above mentioned PoS words as contextual clues. The contextual clues surrounding the adjectives in the three datasets is presented in Table 5 given below.

It is clear from the above table data that the meaning of the target adjective is not easily ascertained by only analyzing the adjective itself. The accuracy in the orientation of opinion words that are identified without the expansion of seed sets using WordNet is presented below in Table 6.

Table 6 Opinion Word Orientation % without the Expanssion seed Sets using Wordset

Information	Statistics on D1	Statistics on D2	Statistics on D3
Percentage of adjectives	96.16%	98.62%	98.62%
Opinion word orientation %	79.08%	80.35%	78.04%

The accuracy values in determining the orientation of opinion words obtained using the considered seed sets without expansion is relatively better than those of the accuracies obtained using Naïve Bayes classifier on Kamps and Turney and Littmann seed sets in the work of [16]. The correct adjectives are considered for carrying out this experiment, which the works of [1, 4] never clarified.

The sense of the adjective is finalized by calculating the relatedness of the adjective with the available senses in the WordNet with the surrounding contextual clues using WordNet::Similarity software package [18]. The accuracy in the orientation of opinion words that are identified with the expansion seed sets using synonyms and antonyms from WordNet is presented below in Table 7.

Table 7 Opinion Word Orientation % with the Expansion od Seed Sets using Wordnet

Information	Statistics on D1	Statistics on D2	Statistics on D3
Percentage of adjectives	96.16%	98.62%	98.62%
Opinion word orientation %	89.38%	88.01%	89.67%

The accuracy in determining the orientation of opinion words obtained using the considered seed sets with expansion (best results on D3 dataset) is slightly better than the accuracy obtained using linear SVM classifier on Kamps seed set in the work of [16]. A slight increase of 1.62% is observed. This is because all the correct adjectives from the dataset were considered in this work as opposed to Kamps work. The accuracy in determining the orientation of opinion words obtained using the considered seed sets with expansion (best results on D3 dataset) outperforms the accuracy obtained using probabilistic Term Frequency – Inverse Document Frequency classifier on Turney and Littmann seed set in the work of [16]. An increase of 6.58% is observed. This is because the probabilistic Term Frequency – Inverse Document Frequency classifier has not included the context parameter for learning the orientation.

By achieving an average accuracy of 4.1% across Kamps seed set and Turney and Littmann seed set, it is concluded that the semantic orientation of opinion words using typed dependencies for opinion word senses and the sentiwordnet scores approach performs better than the synonymy relation graph of WordNet for determining the orientation of opinion words in the semantic environment.

6. Coclusion and Future Works.

An attempt to replace the semantic orientation of opinion words using synonymy relation graph of WordNet with typed dependencies for opinion word senses and the Sentiwordnet scores has been carried out successfully. This kind of analysis evaluates the opinion orientation of any adjective from the data corpus.

In future, the obtained opinion words with their orientations of the extracted product features are analyzed for feature specific sentiments. These sentiments are useful in recommending the similar products in a better way than the traditional recommendations when a search for product takes place. These sentiments when annotated with the ontology, the intentions of the reviewers are possible for analysis by debugging the Semantic Web Rule Language (SWRL) rules on the constructs of the ontology. This advanced data model analysis will help the businesses to decrease their customer churn.

References

[1] Kamps, Jaap, et al. "Using wordnet to measure semantic orientations of adjectives." (2004): 1115-1118.

[2] Mukherjee, Subhabrata, and Pushpak Bhattacharyya. "Sentiment analysis: A literature survey." *arXiv preprint arXiv:1304.4520* (2013).

[3] Kamps, Jaap, and Maarten Marx. "Visualizing wordnet structure." *Proc. of the 1st International Conference on Global WordNet*. 2002.;

[4] Turney, Peter D., and Michael L. Littman. "Measuring praise and criticism: Inference of semantic orientation from association." *ACM Transactions on Information Systems (TOIS)* 21.4 (2003): 315-346.

[5] Fellbaum, Christiane. *WordNet*. Blackwell Publishing Ltd, 1998.

[6] Baccianella, Stefano, Andrea Esuli, and Fabrizio Sebastiani. "SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining." *LREC*. Vol. 10. 2010.

[7] Rada, Roy, et al. "Development and application of a metric on semantic nets." *IEEE transactions on systems, man, and cybernetics* 19.1 (1989): 17-30.

[8] Hirst, Graeme, and David St-Onge. "Lexical chains as representations of context for the detection and correction of malapropisms." *WordNet: An electronic lexical database* 305 (1998): 305-332.

[9] Leacock, Claudia, and Martin Chodorow. "Combining local context and WordNet similarity for word sense identification." *WordNet: An electronic lexical database* 49.2 (1998): 265-283.

[10] Resnik, Philip. "Using information content to evaluate semantic similarity in a taxonomy." *arXiv preprint cmp-lg/9511007* (1995).

[11] Osgood, C. E., G. J. Succi, and P. H. Tannenbaum, 1957. The Measurement of Meaning. University of Illinois Press, Urbana IL.

[12] Hatzivassiloglou, Vasileios, and Kathleen R. McKeown. "Predicting the semantic orientation of adjectives." *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, 1997.

[13] Kristina Toutanova and Christopher D. Manning. 2000. Enriching the Knowledge Sources Used in a Maximum Entropy Part-of-Speech Tagger. In Proceedings of the Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora (EMNLP/VLC-2000), pp. 63-70.

[14] Patwardhan, Siddharth, Satanjeev Banerjee, and Ted Pedersen. "Using measures of semantic relatedness for word sense disambiguation." *International Conference on Intelligent Text Processing and Computational Linguistics*. Springer Berlin Heidelberg, 2003.
[15] Hu, Minqing, and Bing Liu. "Mining and summarizing customer reviews." *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2004.
[16] Esuli, Andrea, and Fabrizio Sebastiani. "Determining the semantic orientation of terms through gloss classification." Proceedings of the 14th ACM international conference on Information and knowledge management. ACM, 2005.
[17] De Marneffe, Marie-Catherine, and Christopher D. Manning. *Stanford typed dependencies manual*. Technical report, Stanford University, 2008.
[18] Pedersen, Ted, Siddharth Patwardhan, and Jason Michelizzi. "WordNet:: Similarity: measuring the relatedness of concepts." Demonstration papers at HLT-NAACL 2004. Association for Computational Linguistics, 2004.