Predicting the Polarity Strength of Adjectives Using WordNet

Gbolahan K. Williams, Sarabjot S. Anand
Department of Computer Science, University of Warwick,
United Kingdom
{williamg, ssanand}@dcs.warwick.ac.uk

Abstract

A key element of any sentiment analysis system is the ability to assign a polarity strength value to words appearing within the documents. In this paper we present a novel approach to polarity strength assignment. The approach is knowledge based in that it uses WordNet to build an adjective graph which is used to measure semantic distance between words of known polarity (reference or seed words) and the target word, which is then used to assign a polarity to the target word. We extend previous work in this area by using a small training data set to learn an optimal predictor of polarity strength and to dampen polarity assigned to non-polar adjectives. We also extend the coverage of previous approaches by exploring additional lexical relations not studied previously. The method has been evaluated on a validation set and shows excellent potential in reducing the assignment of spurious polarity and accurately predicting polarity values for polar adjectives.

Introduction

The last decade has seen an exponential increase in the amount of sentiment expressed online by Internet users though blogs and review web sites. While information retrieval and text mining algorithms have traditionally focused on indexing and classifying texts based on the topics they are about, the automatic analysis of these user opinions has recently gained attention. Known as Sentiment Analysis or Opinion Mining, the aim is to identify the polarity of opinion expressed in these documents at a document, sentence, item or item feature level (collectively referred to as the $subject$). The polarity of the opinion may simply be at the level of classifying it into a set of pre-defined classes or may actually be on a continuous scale. In the latter case each opinion is assigned a numerical *polarity strength value* which discloses the degree of positivity or negativity associated with the subject.

The process for analyzing the sentiment expressed within text consists of a number of well defined steps. The first of these is the identification of opinion vs. factual statement/phrases. The second step is to assign a polarity strength to each of the phrases identified in the first step. This requires robust techniques for identifying the polarity strength value of words. Finally, a method is required to integrate phrase level polarity to provide a sentiment strength value/classification for the subject.

In this paper, we present a novel approach to assigning polarity strength values to words. In general, previous work in this area can be classified into Knowledge-based approaches (Kamps *et al.* 2004; Godbole, Srinivasaiah, & Skiena 2007) and Corpus-based methods (Hatzivassiloglou & McKeown 1997; Turney 2002). Knowledge-based approaches tend to use a lexicon such as WordNet (Fellbaum 1998) along with a seed set of adjectives with known polarity and assign polarity to other words based on their semantic distance within the graph generated from WordNet using lexical relations. Corpus-based methods instead use part-of-speech taggers to assign syntactic class labels to words and then use the corpus of texts along with a seed set of positive and negative reference words to assign individual phrases or words with a polarity strength value.

The method presented here is a knowledge based approach to polarity strength assignment. It extends previous work in the area in a number of ways:

1. We explore the use of multiple seed pairs rather than a single arbitrary pair within the polarity strength assignment function. Using a small training set we learn an optimal set of adjective pairs to be used within the function
2. We extend the coverage of these methods by exploring additional lexical relations within WordNet
3. We present a method for filtering out/ dampening the polarity assigned to non-polar adjectives
4. We evaluate our method through a user validation of the predicted polarity values. Our method achieved a correlation of 0.79 with user assigned polarity strength values
5. We explore the use of edge weights to dissuade paths associated with less common word senses to form part of the calculation of polarity strength values.

Polarity Strength Assignment

In our method, we begin with a small set of reference positive and negative terms to build an adjective graph, G^a, by recursively querying the lexical relations defined in WordNet for the set of seed words and adding edges between the words resulting from the query and the query word. Then using WordNet, adjectives are identified and extracted from the review document texts. It is to these adjectives, a_i, that we are interested in assigning polarity strength values.

Measuring Polarity Strength

Given a reference pair (w^+, w^-), the polarity strength of an adjective a_i is calculated as in (Kamps *et al.* 2004):

$$SO_{(w_r^+,w_r^-)}(a_i) = \frac{G^a_{path}(a_i, w_r^-) - G^a_{path}(a_i, w_r^+)}{G^a_{path}(w_r^+, w_r^-)} \quad (1)$$

where $G^a_{path}(w_i, w_j)$ is the shortest path between w_i and w_j within the graph G^a. The value of $SO_{(w_r^+,w_r^-)}(a_i)$, assuming G^a is an undirected graph, belongs to the range $[-1, 1]$.

Our methods however, can result in a directed graph or may use multiple reference paies. For a directed graph, we use the equivalent measure:

$$SO_{(w_r^+,w_r^-)}(a_i) = \frac{G^a_{path}(a_i, w_r^-) - G^a_{path}(a_i, w_r^+)}{max(G^a_{path}(w_r^+, w_r^-), G^a_{path}(w_r^-, w_r^+))} \quad (2)$$

If a set of reference pairs, R^α_{set}, is used, we compute the polarity strength of a word a_i as:

$$SO_{R^\alpha_{set}}(a_i) = \frac{1}{|R^\alpha_{set}|} \sum_{(w_r^+,w_r^-) \in R^\alpha_{set}} SO_{(w_r^+,w_r^-)}(a_i) \quad (3)$$

The Data

For performance evaluation of the methods proposed in this paper, 15 independent individuals were asked to rate a set of 20 adjectives (hereafter, *the GS term set*) in the range [-1,1]. A Validation Set of 28 adjectives rated by 4 independent individuals. Table 1 shows the list of adjectives in the *GS term set* along with the average rating by human subjects (u_{ave}) and the standard deviation (σ) for each adjective. The difficulty of the task at hand is obvious from the large standard deviation associated with the human ratings of a number of adjectives.

Table 1: GS Term Set showing the average rating by 15 human judges; u_{ave} = average rating; σ = standard deviation

#	a_i	u_{ave}	σ	#	a_i	u_{ave}	σ
1	excellent	+0.87	0.1275	11	despicable	-0.69	0.3510
2	splendid	+0.74	0.1570	12	honest	+0.51	0.4770
3	awesome	+0.82	0.1579	13	naughty	-0.17	0.3538
4	nice	+0.40	0.1972	14	awful	-0.76	0.2004
5	dainty	+0.22	0.4526	15	interesting	+0.46	0.3292
6	annoying	-0.48	0.3540	16	poor	-0.54	0.3780
7	rude	-0.57	0.3134	17	good	+0.52	0.2290
8	gracious	+0.49	0.2751	18	amusing	+0.57	0.2620
9	terrible	-0.63	0.3516	19	bad	-0.62	0.3017
10	adverse	-0.42	0.2817	20	amazing	+0.83	0.2097

The correlation coefficient, $CORREL(u_{ave}, p_r)$, of the human ratings u_{ave} and the predicted ratings p_r, and the mean squared error (MSE) was used to evaluate each of the methods proposed using the validation set.

Methods for Polarity Strength Assignment

M1: Using the Path Length of Chains of Synonyms An undirected adjective graph G^a is generated by recursively querying WordNet's synonym relation for words in the seed set. The semantic orientation of an adjective a_i is computed using equation 1.

We experimented with various reference pairs (w^+, w^-) used in literature (Kamps *et al.* 2004; Turney 2002) to study their impact on polarity strength assigned. For example, using *(good,bad)* and *(excellent, poor)* resulted in correlation coefficients of 0.41 and 0.48 respectively. We also chose the pair of words with maximum average user assigned positive and negative polarity *(excellent, awful)* and achieved a correlation coefficient of 0.44. An exhaustive search of all pairs of positive and negative adjectives within *GS term set* resulted in the reference pair $(excellent, annoying)$ being chosen with correlation coefficient 0.53.

The drawbacks of **M1** are inherent in the fact that, firstly, it is only possible to compute $SO_{(w^+,w^-)}(a_i)$ if a path exists between the adjective a_i and the reference pair (w^+, w^-). Secondly, this method is dependent on a good choice for the reference pair (w^+, w^-) being used in the computation of $SO_{(w^+,w^-)}(a_i)$. Thirdly, the method is prone to allowing synonym links between paths of unlikely senses of a word being traversed. Finally, there is no rationale for limiting the method to using only one reference pair.

M2: Using M1 and weighted edges between synonym links Godbole et al (Godbole, Srinivasaiah, & Skiena 2007) suggest that paths that use less common senses are less reliable and hence they use a threshold on the number of senses used when generating the adjective graph. Instead of a hard threshold, we investigated the use of increased edge weights for edges associated with less common word senses, hence allowing the traversal of these paths but assigning a penalty for doing so. As WordNet provides the ordering of every *Word Sense* (or *concept*) based on how commonly used they are, the rank based on this ordering of concepts is utilized. That is, given a word w_i and it's synonym sets $ws_{i1}, ws_{i2}, ..., ws_{it}$ where the synset ws_{ir} corresponds to the r-th most common sense of w_i, we assign a weight to all edges from w_i to words in the synset ws_{ir} of $e^{((r-1)\times\alpha)}$, where α is a constant that controls the severity of the penalty assigned to lower ranked senses. The smaller the value of α the lower is the penalty for less common word senses.

For example, the 6th word sense of *awful* according to WordNet has the adjectives *awesome* and *amazing* as synonyms. Penalizing such word senses should lead to more accurate polarity assignments as more *valid* paths would be used in the computation.

The new adjective graph is now a directed, weighted graph as the edge weights of the edges (w_i, w_j) and (w_j, w_i) may be different. Hence we use Equation 2 to assign polarity strength to an adjective. Note that **M1** is a special case of this method where α = 0.

At α = 0.7 the algorithm performs best with a correlation of 0.57 using the optimal reference pair $(excellent, splendid)$.

M3: Using M2 in conjunction with a set of reference pairs Some experimentation with the adjective polarity values generated using Methods **M1** and **M2** with different reference pairs led to the belief that taking the average semantic orientation of an adjective a_i in the computation of $SO(a_i)$ across *multiple* reference pairs improves the estimation of the semantic orientation of the adjective.

For example, using the reference pairs $(excellent, poor)$ and $(good, bad)$ achieved a correlation of 0.65. This lead us to investigate a more principled approach to choosing reference pair sets.

We developed an algorithm that takes an adjective seed list of positive and negative terms (Adj^+ and Adj^- respectively) derived from the *GS term set* and performs a *greedy search* of the space of subsets of pairs (a_i, a_j) where $a_i \in Adj^+$ and $a_j \in Adj^-$. Once the optimal set of reference pairs has been learnt, the polarity strength is calculated using Equation 3.

The algorithm was run on a range of values of $\alpha = \{0.1, 0.2, ..., 1.0\}$. The best correlation of 0.92 is achieved when $\alpha = 0.0$, when using the reference pair set $\{(excellent, annoying), (amazing, awful), (good, bad), (splendid, rude), (awesome, awful), (good, despicable)\}$.

Investigating additional Lexical Relations in WordNet

Although **M3** shows strong correlation with the human ratings set, it still suffer from very poor coverage of adjectives. The adjective graph only contains 4425 adjective vertices and is not connected to strongly polar adjectives such as *'interesting'*. For all such disconnected adjectives, a_i, $SO_{R^{\alpha}_{set}}(a_i) = 0$.

A solution to this problem is to expand the adjective graph by exploring additional lexical word relations, for example: antonymy, similarity, relatedness. Esuli and Sebastiani (Esuli & Sebastiani 2005) and Godbole et al. (Godbole, Srinivasaiah, & Skiena 2007) have also suggested the use of additional lexical relations for expanding the graph but as far as we are aware the similarity and relatedness relations have not been explore previously. As edge weights may differ for each lexical relation explored, for now we assign an edge weight of one to all edges in the graph.

M4: Expansion based on Antonym Relation: The adjective graph, $G^{a\prime}$, is expanded by creating edges between synonyms and antonyms. This resulted in a graph with 5822 adjective vertices. We assume that the antonym of a word has opposite orientation (Kim & Hovy 2004; Esuli & Sebastiani 2005). Hence, in the computation of the path distance $G^a_{path}(x, y)$ between any two adjective vertices x and y the orientation is flipped when an odd number of antonym links are part of the path.

The optimal set of reference pairs discovered for $G^{a\prime}$ was $\{(splendid, despicable), (dainty, bad), (amazing, awful), (honest, awful), (awesome, awful), (good, annoying)\}$ with a correlation of 0.83.

M5: Expansion based on Similar Words Relation and Related Words Relation: Finally the adjective graph G^a was expanded through the addition of the similar words relation and related word relation. This expanded graph, $G^{a\prime\prime}$, increased coverage significantly from 4425 adjective vertices to over 16000 adjective vertices.

Based on $G^{a\prime\prime}$, the optimal reference pair set (with correlation coefficient 0.95) discovered was: $\{(amazing, awful), (splendid, awful), (interesting, rude), (awesome, awful), (interesting, annoying), (excellent, naughty)\}$.

Evaluation of Methods and Results

Analysis of a set of theatre reviews gathered from ten different theatre related web sites showed a possible downside to the increased coverage. Not all adjectives are polar i.e. have a positive or negative sentiment associated with them. This has been observed in previous work too. For example Turney's method based on PMI (Turney 2002) assigns a strong positive orientation to phrases such as "online experience" and "online service". The only other work that we are aware of that specifically tries to dampen or filter out such spurious polarity assignment is that by Hatzivassiloglou and Wiebe (Hatzivassiloglou & Wiebe 2000). To deal with this problem we decided to add a number of adjectives to the training data with a polarity strength value of 0. The hope was that the function learned from the training data that included such non-polar adjectives would learn to filter out/dampen the polarity assigned to other non-polar words. We studied the effect of adding 10 through to 50 non-polar adjectives to the training set. We also added 30 non-polar adjectives to the validation set to evaluate the effect of learning from the new training data on unseen examples of non-polar adjectives. Table 2 shows the results obtained. As can be seen, the correlation coefficient and MSE reduce for the training data as the size of the non-polar sample in the training set increases. The reduction in correlation coefficient is caused by the large number of words with zero polarity being added to the training data. However, as we can see the MSE reduces by a statistically significant amount at the 95% confidence level. To separate out the effect on non-polar and polar adjectives within the validation data, we show the correlation coefficient and MSE for the polar adjectives in the validation data and the MSE for the non-polar adjectives separately. For the polar adjectives, the correlation coefficient reduces as the sample size increases by a small amount and the MSE increases by a non significant amount. The effect on the non-polar adjectives is, on the other hand, statistically significant. Hence clearly adding more non-polar words to the training data is having a positive effect on our desire to dampen polarity strength. The resulting model is referred to as **M5′**.

Table 3 summarizes the methods described in the previous section and the results obtained on the validation data set. It would appear to be the case that methods **M1-M4** suffer from overfitting of the training data. In fact a large part of the reduction in performance of these models on the validation data set is a result of poor coverage. Of the 28 polar adjectives in the validation set, 9 adjectives were not connected to the reference pairs used by the model. While

Table 3: Method Summary and Performance

Method	Lexical Relations	Reference Pair	Data	Coverage	Training Data Corr	Validation Data Corr	Validation Data MSE
M1	Synonym	Single	20 polar	4425	0.535	0.476	0.26±0.08
M2	Synonym with Weights	Single	20 polar	4425	0.572	0.445	0.27±0.08
M3	Synonym	Multiple	20 polar	4425	0.92	0.335	0.38±0.18
M4	Synonym & Antonyms	Multiple	20 polar	5822	0.834	0.3598	0.4±0.13
M5	Synonym, Similar Words & Related Words	Multiple	20 polar	16000	0.956	0.78	0.13±0.08
M5′	Synonym, Similar Words & Related Words	Multiple	20 polar & 50 non-polar	16000	0.708	0.77	0.185±0.02

Table 2: Effect of Size of Non-polar Sample in Training and Validation Sets

Sample Size	Training Data Corr	Training Data MSE	Validation Data Corr	Validation Data MSE	Non Polar MSE
0	0.95	0.064±0.026	0.78	0.13±0.08	0.08±0.03
10	0.85	0.062±0.026	0.79	0.14±0.067	0.068±0.028
20	0.81	0.06±0.026	0.79	0.15±0.05	0.036±0.016
30	0.76	0.058±0.026	0.77	0.17±0.057	0.027±0.01
40	0.74	0.052±0.021	0.78	0.17±0.055	0.025±0.01
50	0.71	0.05±0.019	0.77	0.18±0.059	0.02±0.008

M3 showed a very significant improvement in the correlation coefficient on the training sample, it performed rather poorly on the validation data set. This would suggest that the model did actually overfit the training data set.

While extending the graph with the antonym relation **M4** did improve coverage, 7 of the adjectives in the validation set were still not connected to the resulting graph affecting performance on the validation set. **M5** and **M5′** display good coverage and hence a much better correlation coefficient and MSE value on the validation set than the other models. **M5′** shows a statistically insignificant increase in MSE and decrease in correlation for polar adjectives, as compared with **M5**, on the validation set. However, it displays good potential in dampening the polarity of non-polar adjectives. As can be seen from Table 2, on non-polar adjectives **M5′** shows a statistically significant decrease in MSE as compared to **M5** (Sample Size = 0).

Conclusions and Future Work

In this paper we have investigated a number of approaches to learning a function to predict the polarity strength of adjectives. The methods use a small seed set of adjectives to build adjective graphs using various lexical relations defined in WordNet. Evaluation of these methods show that they can indeed extend the coverage of previously proposed methods and achieve a good accuracy on unseen, validation data while substantially reducing the polarity assigned to non-polar words. The best results were achieved when using the lexical relations of related words and similar words in addition to the standard synonym relation commonly used. The optimal method also used a set of reference pairs learned by our system with the help of a training set of 10 adjectives with positive polarity, 10 adjectives of negative polarity and 50 non-polar adjectives. The method achieved a correlation of 0.77 between the predicted polarity strength values and the average polarity strength assigned by human subjects to words in a validation set.

The research presented in this paper has also raised a number of issues that need to be investigated further. In particular we would like to investigate the issue of edge weighting for each reference pair within the reference pair set. The issue of how edges must be weighted in graphs generated using multiple lexical relations is also an open question. We would also like to extend the work to assign polarity strength values to phrases. Words from syntactic classes other than adjectives can also have a polarity associated with it e.g. the noun "masterpiece" has a strong positive polarity. We would also like to carry out a comparative evaluation of our approach with some of the Corpus based techniques. Finally, we would also like to evaluate the effect of improved polarity strength prediction on classification of sentences into the two classes of opinion and factual sentences and the classification, at a document level, of reviews to a predefined set of categories of polarity.

References

Esuli, A., and Sebastiani, F. 2005. Determining the semantic orientation of terms through gloss classification. In *In Proc. CIKM-05*, 617–624.

Fellbaum, C., ed. 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.

Godbole, N.; Srinivasaiah, M.; and Skiena, S. 2007. Large-scale sentiment analysis for news and blogs. In *Proceedings of the ICWSM*.

Hatzivassiloglou, V., and McKeown, K. R. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*, 174–181.

Hatzivassiloglou, V., and Wiebe, J. 2000. Effects of adjective orientation and gradability on sentence subjectivity. In *COLING*.

Kamps, J.; Marx, M.; Mokken, R. J.; and Rijke, M. D. 2004. Using wordnet to measure semantic orientation of adjectives. In *LREC*, 1115–1118.

Kim, S.-M., and Hovy, E. 2004. Determining the sentiment of opinions. In *COLING '04*, 1367. Morristown, NJ, USA: Association for Computational Linguistics.

Turney, P. D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In *Proceedings of ACL '02*, 417–424. Morristown, NJ, USA: Association for Computational Linguistics.