Generalized Sentiment-Bearing Expression Features for Sentiment Analysis

Shizhu Liu   Gady Agam   David Grossman

Computer Science Department, Illinois Institute of Technology

`sliu28, agam, grossman@iit.edu`

ABSTRACT

In this work, we propose a novel approach to extract sentiment-bearing expression features derived from dependency structures. Rather than directly use dependency relations generated by a parser, we propose a set of heuristic rules to detect both explicit and implicit negations in the text. Then, three patterns are defined to support generalized sentiment-bearing expressions. By altering existing dependency features with detected negations and generalized sentiment-bearing expressions we are able to achieve more accurate sentiment polarity classification. We evaluate the proposed approach on three labeled collections of different lengths, and measure the gain from the generalized dependency features when used in addition to the bag-of-words features. Our results demonstrate that generalized dependency-based features are more effective when compared to standard features. Using these we are able to surpass the state-of-the-art in sentiment classification.

KEYWORDS: Sentiment analysis, Natural Language Processing, Classification.

---

1 Introduction

With the proliferation of Web 2.0 tools and applications on the Internet, there is an exponential increase in the number of online postings submitted by web users on their opinions, experiences, etc. This trend drawn the attention of organizations, companies and researchers who are interested in opinions expressed by people on various topics. Sentiment analysis, the task of identifying sentimental aspect of a text, has been a popular direction in the field of language technologies.

Recent work in supervised sentiment analysis has focused on innovative approaches to feature creation, which aim to improve the performance with features that capture the essence of linguistic constructs used to express sentiment. A straightforward way to extend the traditional bag-of-words representation is to heuristically add new types of features, such as fixed-length n-gram (e.g., bigram or trigram) or pairwise syntactic relations (e.g., typed dependencies).

However, the performance of joint features is still far from satisfactory. N-grams which cover only co-occurrences of N continuous words in a sentence has problems with capturing long dependencies, and the performance of a dependency relation feature set is reported to be inferior to N-grams. We conjecture that this reduced performance is in part due to the following two reasons: 1) pairwise dependency features sometimes fail to reflect the correct sentiment polarity by neglecting to consider the influence of other terms, especially negations, in the given sentence; 2) in dependency relation features, features lack sentiment oriented generalizations.

The main contribution of this paper is in the construction of more accurate generalized sentiment-bearing expression features for the sentiment classification. We propose a set of heuristic rules to detect implicit negation relations and propose three patterns as the basis for generalized dependency-based sentiment oriented features:

Explicit patterns Many terms directly reflect the sentiment. e.g. “great camera”, “love this movie”. The parsed dependency relations amod(camera, great), and obj(love, movie) can already capture these explicit sentiment expressions.

Range patterns In some cases, there is an assumed standard and sentiment is indicated by describing the distance from the standard. For example, in the sentence "The quality of this product is above average", "above average" indicates distance from the standard.

Trend patterns In some cases, sentiment is conveyed by describing the trend of how the object changes. For example, in the sentence "The popularity of this band have continuously decreased from their peak in 2000", "decreased from" indicates a trend.

For each type of sentiment expression, we propose a corresponding generalization strategy. We show that when trained on such revised generalized features, machine learning classification algorithms achieve better sentiment classification accuracy.

The remainder of this paper is organized as follows. Section 2 presents related work on the topic of sentiment analysis. Section 3 introduces the proposed approach using some motivating examples and a set of heuristic rules with generalization strategies. Experimental results are discussed and compared to known techniques in Section 4. Section 5 concludes the paper and outlines future directions.

2 Related Work

Sentiment has been studied at three different levels: word, sentence, and document level. On document level, previous work (Pang et al., 2002) (Pang and Lee, 2004) have shown that traditional text classification approaches can be quite effective when applied to sentiment analysis. On word level, Wilson et al.(Wilson et al., 2005) extract phrase-level clues by identifying polarity shifter words to adjust the polarity of opinion phrases. Kim et al.(Kim et al., 2009) shows how various term weighting schemes improve the performance of sentiment analysis systems. Choi et al.(Choi et al., 2009) validated that topic-specific features would enhance existing sentiment classifiers. On sentence level, linguistic approaches are used to discover the interaction between words that may switch a sentence's sentiment polarity (Wilson et al., 2004) (Choi et al., 2005).

A prominent polarity shifter clue in sentences is negation. Pang et al. (Pang et al., 2002) employ the technique of Das and Chen (Das and Chen, 2001) to add the tag "NOT_" to every word between a negation word and the first following punctuation mark. Negation and its scope in the context of sentiment analysis has been studied in (Moilanen and Pulman, 2007). Choi and Cardie (Choi and Cardie, 2008) combine different kinds of negations with lexical polarity items through various compositional semantic models to improve phrasal sentiment analysis. A recent study by Danescu-Niculescu-Mizil et al. (Danescu-Niculescu-Mizil et al., 2009) looked at the problem of finding downward-entailing operators that include a wider range of lexical items, including soft negation such as adverbs "rarely" and "hardly". Councill et al. (Councill et al., 2010) focus on explicit negation mentions and investigate how to identify the scope of negation in free text.

There have been some attempts at using features for polarity classification from dependency parses. Dave et al. (Dave et al., 2003) found that adding adjective-noun dependency relationships as features does not provide any benefit over a simple bag-of-words based feature space. Arora et al.(Arora et al., 2010) use a subgraph mining algorithm to automatically derive frequent subgraph features in addition to the bag-of-words features. Moilanen et al.(Moilanen and Pulman, 2007) discuss sentiment propagation, polarity reversal, and polarity conflict resolution within various linguistic constituent types. Ng et al. (Ng et al., 2006) proposed that the subjective-verb and verb-object relationships should also be considered for polarity classification. However, they observed that the addition of these dependency relationships does not improve performance over a feature space that includes unigrams, bigrams.

To solve the sparse-data problem for machine learning classifiers, there were attempts at finding better generalized dependency features. (Gamon, 2004) back off words in N-gram (and semantic relations) to their respective POS tags. (Joshi and Penstein-Rose, 2009) proposed a method by only backing off head word in dependency relation pairs to its POS tag. Xia and Zong (Xia and Zong, 2010) further propose to back off the word in each word relation pairs to its corresponding POS cluster to make the feature space smarter and more effective.

3 Methodology

In this section, we first motivate our approach using sample sentences. We then demonstrate the application of heuristic rules for negation and pattern detection. Finally, we describe how to generalize the extracted sentiment-bearing expressions.

3.1 Motivation for our Approach

To facilitate the discussion, consider the following examples:

1. *Avatar is a great movie!*
2. *This is not a great movie.*
3. *No one likes these extra functions.*
4. *This news is too good to be true.*
5. *The leading actors' sterling performances raise this far above the level of the usual maudlin disease movie .*
6. *The lack of training exposed truck drivers to an increased risk of injury.*
7. *This accessory can abate the damage.*
8. *New regulations increase accountability and boost quality in head start.*

By applying the dependency parser to the first two sentences, the extracted dependency relations in both sentences contain the dependency relation *amod(movie, great)* which is used to express both positive (in the first sentence) and negative (in the second sentence) sentiments. If all pairwise dependency relations are directly appended to unigram features, *amod-movie-great* becomes a common feature for positive and negative examples and the sentiment classifier cannot benefit from it. We propose to keep all negated word as negation indicator terms and present them in their negated status as composite dependency features (e.g. not-amod-movie-great for the second sentence).

Besides explicit negation relations that can be detected by a dependency parser directly, implicit negation which does not use negation terms is hard to detect. For example, "no one" in the third sentence shifts the polarity of the verb "likes", and "too" in the fourth sentence shows the implicit negation for the term after the word "to". To construct accurate dependency features for sentiment classification, we propose a set of heuristics for the detection of implicit negation relations.

Sentiment may be expressed implicitly by referring to an assumed standard. For example, consider the fifth and sixth sentences where sentiment information is expressed by describing the target object as being above or below an ordinary level. In the seventh and eighth sentences, sentiment may also be expressed by describing how an object changed. For the construction of composite back off features for the range and trend patterns, related indicator terms are backed off as status info instead of its POS tags (e.g. "prep(lack, training)" backed off as "prep-blw-training" and backed off as "dobj(abate, damage)" as "dobj-dec-damage").

3.2 Heuristics-Based Sentiment Detection Methods

This section describes a set of heuristic rules for detecting sentiment-bearing expressions. For a given sentence, we first parse it and get its corresponding dependency tree represented as a list of dependency relation list. We then attempt to detect negation, range, and trend indicator terms. These are used for generalized sentiment expression construction in the next step.

WordNet [1] is used to construct range and trend pattern indicator term synset. e.g. all the synonyms of "above" will be included in the range indicator synset and all the synonyms of "increase" will be in the trend pattern indicator synset.

Table 1 shows the definition of sentiment indicator detection rules along with motivating examples. In order to apply a rule, we first detect the a dependency relation and then apply the *Detect* function as defined in Table 2. The *Detect* function first checks whether the first argument is a negation indicator term, and if so, insert a negation dependency relation for the second argument. If the first argument is a range or trend indicator term, we keep it in the indicator

[1] http://wordnet.princeton.edu/

term list for the next step of generalized feature construction.

	Rules	Examples
1	$neg(arg1, not) = \neg(arg1)$	not $[bad]_{arg1}$
2	$subj(V,N) = Detect(N,V,subj)$	$[Nobody]_N[likes]_V$ this product
3	$obj(V,N) = Detect(N,V,obj)$	He is $[supported]_V$ by $[none]_N$.
4	$advmod(V,R) = Detect(V,R,advmod)$	PM2.5 $[rarely]_R[decreased]_V$ recently.
5	$ccomp(J,V) = Detect(J,V,ccomp)$	It is $[impossible]_J$ to $[overrate]_V$ it.
6	$xcomp(J,V) = Detect(J,V,xcomp)$	This news is too $[good]_J$ to $[believe]_V$.
7	$amod(N,J) = Detect(N,J,amod)$	$[high]_J[interestrate]_N$.
8	$advmod(J,R) = Detect(J,R,advmod)$	$[too]_R[fast]_N$.
9	$prep(N_1,N_2) = Detect(N_1,N_2,prep)$	$[lack]_{N_1}$ of $[training]_{N_2}$
10	$obj(V,N) = Detect(N,V,obj)$	This accessory can $[abate]_V[damage]_N$.

Table 1: Sentiment indicator term detection rules

if($arg3 == subjANDarg1 \in negatedsubject$)	then insert neg(arg2, not)
else if($arg3 == objANDarg1 \in negatedsubjects$)	then insert neg(arg2, not)
else if($arg3 == advmodANDarg2 \in negatedadv$)	then insert neg(arg1, not)
else if($arg3 == ccompANDarg1 \in negatedadj$)	then insert neg(arg2, not)
else if($arg3 == xcompANDarg1existinadvmod(arg1, too)$)	then insert neg(arg2, not)
else if($arg3 == amodANDarg2 \in abovesynset$)	then label arg1 as abv
else if($arg3 == amodANDarg2 \in belowsynset$)	then label arg1 as blw
else if($arg3 == advmodANDarg2 \in abovesynset$)	then label arg1 as abv
else if($arg3 == advmodANDarg2 \in belowsynset$)	then label arg1 as blw
else if($arg3 == prepANDarg1 \in abovesynset$)	then label arg2 as abv
else if($arg3 == prepANDarg1 \in belowsynset$)	then label arg2 as blw
else if($arg3 == objANDarg1 \in increasesynset$)	then label arg1 as inc
else if($arg3 == objANDarg1 \in decreasesynset$)	then label arg1 as dec

Table 2: Definition of Detect(arg1, arg2, arg3)

3.3 Generalized Sentiment-bearing Expression Features

In order to make a further generalization, we conduct POS and grammatical relation clustering. The POS tags and grammatical relations are categorized as shown in Table 3

For negation indicator terms, we add the tag “not-” to all the dependency relations where it occurred. For the range and trend pattern indicator terms, a status tag based on its semantic meaning will used in the corresponding relations. Table 4 present some examples for these type of specific generalizations.

POS-cluster	Contained POS tags
J	JJ, JJS, JJR
R	RB, RBS, RBR
V	VB,VBZ, VBD, VBN, VBG, VBP
N	NN, NNS, NNP, NNPS, PRP
O	The other POS tags
Relation-cluster	**Contained grammatical relations**
mod	amod, advmod, partmod, rcmod, acomp
subj	nsubj, nsubjpass, xsubj, agent
obj	dobj, iobj, xcomp
prep	prep, prepc

Table 3: POS clustering (the Penn Corpus Style) and grammatical relation clustering.

Dep	Indicator	G-Feature
amod(camera, great)	not-great	not-mod-N-great
amod(interest, high)	high	mod-abv-interest
prep(level, below)	below	prep-blw-level
dobj(abate, damage)	abate	obj-dec-damage
dobj(improve, quality)	improve	obj-inc-quality

Table 4: Different types of generalized sentiment-bearing expression feature.

4 Experiments

Details of our experimental evaluation and results follow.

4.1 Experimental Setup

Datasets: Three datasets are used in our sentiment polarity classification experiments:

1. NPS survey dataset v1.0 to which we refer to as "surveys" (3000 promoter and 3000 detractor survey entries, with avg. 10 words)

2. sentences/snippets polarity dataset v1.0 (Pang and Lee, 2005) to which we refer to as "short reviews" (5331 positive and 5331 negative reviews, with avg. 21 words)[2].

3. polarity dataset v2.0 (Pang and Lee, 2004) to which we refer to as "long reviews" (1000 positive and 1000 negative reviews, with avg. 780 words)[3].

[2]http://www.cs.cornell.edu/people/pabo/movie-review-data/rt-polaritydata.tar.gz
[3]http://www.cs.cornell.edu/people/pabo/movie-review-data/review-polarity.tar.gz

	Negation	Range	Trend
review	261	142	8
survey	308	162	47

Table 5: Negation, Range, Trend Pattern Occurrence Information

The three datasets are of different lengths. The polarity dataset is composed of relatively long movie reviews. The sentence/snippets polarity dataset v1.0 is composed of formal written sentence level examples and text in survey sentences are usually short and incomplete. We conduct polarity classification experiments over these three datasets to evaluate the proposed method and investigate the effect of text length on classification performance.

Classifier: We performed n-fold cross-validation experiments on the above datasets, using Joachims' SVM-light (Joachims, 1999) [4] package to train an SVM polarity classifier. All learning parameters were left at their default values. Following (Pang et al., 2002), we use frequency to determine word presence. Each document is first tokenized and downcased, and then represented as a vector of features with 2-norm. A χ^2 feature selection strategy (Yang and Pedersen, 1997) is applied to back-off sentiment-bearing expression features, where we reject features if their χ^2 score is not significant at the 0.2 level.

4.2 Pattern Detection Evaluation

Considering the critical role of sentiment-bearing expression detection in the proposed approach, we evaluated the accuracy of this step separately. For this purpose, we annotated several subsets of the datasets. Specifically, we created a subset which consists of 200 positive and 200 negative sentences from the sentences/snippets polarity dataset v1.0 and a subset which consists of 200 positive and 200 negative sentences from the NPS survey dataset v1.0. Table 5 presents information of negation, range, and trend patterns in the labeled subsets. We see that negation patterns are in general the most frequent, and occur in a majority of the documents whereas trend patterns are in general less frequent.

[4] http://svmlight.joachims.org

Patterns	Precision/Recall(%)	
	review	survey
Negation	74.9/70.9	88.3/75.6
Range	75.8/79.6	82.4/86.4
Trend	88.9/100.0	100.0/97.9
Average	75.6/74.5	87.3/81.0

Table 6: Negation, Range, Trend Pattern Detection Accuracy

The Stanford parser[5] was used to extract dependency relations in our experiments. Table 6 shows the performance of the sentiment-bearing expression detection component. As can be observed, our detection component performs better on sentences of the survey dataset compared with the review dataset. This is related to fact that sentences from the review set are longer and more complex compared with sentences from the survey set, which indicates increase in complexity for the review set.

4.3 Results and Discussion

Finally, the accuracy of an SVM classifier using different sets of features is shown in Table 7. We used the the SVM-light classifier over unigram (uni), unigram with bigram (uni+bi), unigram with all dependencies (uni+dep), and an ensemble with the proposed sentiment-bearing expression features (uni+gdep) using 10-fold cross-valuation. As can be observed the proposed feature set yields the best results when compared with several baseline techniques. Compared with the baseline of bag-of-words expression, the proposed feature set yields a significant performance improvement with the sentence and review datasets. And a minor improvement is achieved by both bigram and generalized sentiment-bearing features for the survey dataset. The minor improvement in the survey dataset is due to the fact that sentences in this dataset are simpler and shorter, and some of the negation, range, and trend pattern have already been captured by bigrams.

A comparison between our results and results reported in the literature for the movie review polarity dataset v2.0 (Pang and Lee, 2004) (Ng et al., 2006) (Matsumoto et al., 2005) indicate that our results surpass the known state-of-the-art regarding this dataset.

[5] http://nlp.stanford.edu/software/lex-parser.shtml

Features	Accuracy(%)		
	survey	sentence	review
uni	90.4	76.6	87.1
uni+bi	91.5	77.8	88.0
uni+dep	91.1	77.4	87.7
uni+gdep	**91.7**	**84.4**	**93.3**

Table 7: Sentiment Classification Accuracy using 10-fold Cross-evaluation

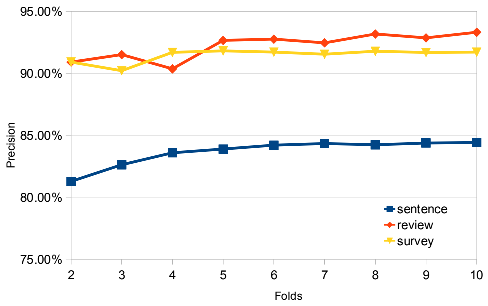

Figure 1: Sentiment classification accuracy using variable cross-validation

To evaluate the influence of the training set size on performance, we performed evaluation using from 2 to 10-fold cross-validation using three datasets. The results shown in Figure 1 indicate that the accuracy of the proposed approach improves with increase in the training set size. As can be observed, the precision fluctuates under 4 folds and stays steady above the 5 folds.

5 Conclusions

The focus of this paper is the construction of more accurate composite sentiment-bearing expression features for sentiment classification. Three patterns are defined to cover more sentiment-bearing expressions and we investigate how to construct more sentiment feature by considering both explicit and implicit negations in the sentence. We propose a set of heuristic rules to detect negations and sentiment-bearing expressions and a dataset is manually annotated for the evaluation of the pattern detection component. Results show that the performance of the pattern detection components can meet the practical applications' requirement and the proposed methods can improve the accuracy significantly.

References

Aho, A. V. and Ullman, J. (1972). *The Theory of Parsing, Translation and Compiling*, volume 1. Prentice-Hall, EngleWood Cliffs, NJ.

Arora, S., Mayfield, E., Penstein-Rose, C., and Nyberg, E. (2010). Sentiment classification using automatically extracted subgraph features. In *Proc. NAACL HLT Workshop on Computational Approaches to Analysis and Generation of Emotion in Text*, pages 131–139.

Ashok K. Chandra, D. C. K. and J.Stockmeyer, L. (1981). Alternation. *Journal of the Association for Computing Machinery*, 28(1):114–133.

Benarmara, F., Hatout, N., Muller, P., and Ozdowska, S., editors (2007). *Actes de TALN 2007 (Traitement automatique des langues naturelles)*, Toulouse. ATALA, IRIT.

Choi, Y. and Cardie, C. (2008). Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proc. EMNLP*, pages 793–801.

Choi, Y., Cardie, C., Riloff, E., and Patwardhan, S. (2005). Identifying sources of opinions with conditional random fields and extraction patterns. In *Proc. conf. Human Language Technology and Empirical Methods in Natural Language Processing*, pages 355–362.

Choi, Y., Kim, Y., and Myaeng, S.-H. (2009). Domain-specific sentiment analysis using contextual feature generation. In *Proc. 1st intl. CIKM workshop on Topic-sentiment analysis for mass opinion*, pages 37–44.

Councill, I. G., McDonald, R., and Velikovich, L. (2010). What's great and what's not: learning to classify the scope of negation for improved sentiment analysis. In *Proc. Workshop on Negation and Speculation in Natural Language Processing*, pages 51–59.

Danescu-Niculescu-Mizil, C., Lee, L., and Ducott, R. (2009). Without a doubt?: unsupervised discovery of downward-entailing operators. In *Proc. Human Language Technologies: NAACL*, pages 137–145.

Das, S. and Chen, M. (2001). Yahoo! for amazon: Extracting market sentiment from stock message boards. In *Proc. Asia Pacific Finance Association Annual Conf. (APFA)*.

Dave, K., Lawrence, S., and Pennock, D. M. (2003). Mining the peanut gallery: opinion extraction and semantic classification of product reviews. In *Proc. 12th intl. conf. on World Wide Web*, pages 519–528.

Gamon, M. (2004). Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In *Proc. 20th intl. conf. on Computational Linguistics (COLING)*.

Joachims, T. (1999). Making large-scale support vector machine learning practical. In Scholkopf, B., Burges, C. J. C., and Smola, A. J., editors, *Advances in kernel methods*, pages 169–184. MIT Press.

Joshi, M. and Penstein-Rose, C. (2009). Generalizing dependency features for opinion mining. In *Proc. ACL-IJCNLP Conf. Short Papers*, pages 313–316.

Kim, J., Li, J.-J., and Lee, J.-H. (2009). Discovering the discriminative views: measuring term weights for sentiment analysis. In *Proc. Joint Conf. Annual Meeting ACL and Intl. Joint Conf. Natural Language Processing of the AFNLP*, pages 253–261.

Laignelet, M. and Rioult, F. (2009). Repérer automatiquement les segments obsolescents à l'aide d'indices sémantiques et discursifs. In *Actes de TALN 2009 (Traitement automatique des langues naturelles)*, Senlis. ATALA, LIPN.

Langlais, P. and Patry, A. (2007). Enrichissement d'un lexique bilingue par analogie. In (Benarmara et al., 2007), pages 101–110.

Matsumoto, S., Takamura, H., and Okumura, M. (2005). Sentiment classification using word sub-sequences and dependency sub-trees. In *Proc. 9th Pacific-Asia conference on Advances in Knowledge Discovery and Data Mining*, pages 301–311.

Moilanen, K. and Pulman, S. (2007). Sentiment composition. In *Proc. of RANLP*.

Mukherjee, S. and Bhattacharyya, P. (2012). Wikisent : Weakly supervised sentiment analysis through extractive summarization with wikipedia. In *European Conference on Machine Learning (ECML-PKDD 2012)*, pages 250–260, Bristol, U.K. Springer.

Ng, V., Dasgupta, S., and Arifin, S. M. N. (2006). Examining the role of linguistic knowledge sources in the automatic identification and classification of reviews. In *Proc. Intl. Conf. on Computational Linguistics (COLING): Posters*, pages 611–618.

Pang, B. and Lee, L. (2004). A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Proc. ACL.

Pang, B. and Lee, L. (2005). Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *Proc. ACL*, pages 115–124.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proc. EMNLP*, pages 79–86.

Wilson, T., Wiebe, J., and Hoffmann, P. (2005). Recognizing contextual polarity in phrase-level sentiment analysis. In *Proc. conf. on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354.

Wilson, T., Wiebe, J., and Hwa, R. (2004). Just how mad are you? finding strong and weak opinion clauses. In *Proc. 19th national conference on Artifical intelligence (AAAI)*, pages 761–767.

Xia, R. and Zong, C. (2010). Exploring the use of word relation features for sentiment classification. In *Proc. Intl. Conf. on Computational Linguistics (COLING): Posters*, pages 1336–1344.

Yang, Y. and Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In *Proc. 14th Intl. Conf. on Machine Learning (ICML)*, pages 412–420.