Available online at www.sciencedirect.com
SCIENCE DIRECT
Library Collections, Acquisitions, & Technical Services 29 (2005) 180–191
Library Collections, Acquisitions, & Technical Services

Use of negation phrases in automatic sentiment classification of product reviews

Jin-Cheon Na*, Christopher Khoo, Paul Horng Jyh Wu

Division of Information Studies, School of Communication and Information, Nanyang Technological University, 31 Nanyang Link, Singapore 637718, Singapore

Available online 9 June 2005

Abstract

This paper reports a study in automatic sentiment classification, i.e., automatically classifying documents as expressing positive or negative sentiments. The study investigates the effectiveness of using a machine-learning algorithm, support vector machine (SVM), on various text features to classify on-line product reviews into recommended (positive sentiment) and not recommended (negative sentiment). In the first part of this study, several approaches, unigrams (individual words), selected words (such as verb, adjective, and adverb), and words labeled with part-of-speech tags were investigated. Using SVM, the unigram approach obtained an accuracy rate of around 76%. Error analysis suggests various approaches for improving classification accuracy: handling of negation phrases, inferencing from superficial words, and handling the problem of comments on parts of the product. The second part of the study investigated the use of negation phrase n-grams to improve classification accuracy. This approach increased the accuracy rate to 79.33%. Compared with traditional subject classification which mainly uses unigrams, syntactic and semantic processing of text appear more important for sentiment classification. We expect that deeper linguistic processing will help increase accuracy for sentiment classification.

* Corresponding author. Fax: +65 6792 7526.
E-mail addresses: tjcna@ntu.edu.sg (J.-C. Na), assgkhoo@ntu.edu.sg (C. Khoo), hjwu@main.ntu.edu.sg (P.H.J. Wu).

1464-9055/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.lcats.2005.04.007

1. Introduction

Research in automatic text classification seeks to develop models for assigning category labels to new documents or document segments based on a training set of documents that have been pre-classified by domain experts. Most studies of automatic text classification have focused on "topical classification," i.e., classifying documents by subject or topic (e.g., education vs. entertainment). However, researchers are increasingly turning their attention to non-topical classification. Examples of non-topical classification include genre classification (e.g., Refs. [1,2]) and sentiment classification.

The study reported in this paper is in the area of automatic sentiment classification. Sentiment is a broad concept covering the following types of mental phenomena [3,4]:

1. human judgment (opinion or view), including evaluation of human behavior with respect to social norms, and evaluation of objects and products with reference to aesthetic principles;
2. affective response (emotion or feeling) towards things, actions, events and processes; and
3. attitude (cast of mind or general mental disposition), which reflects the values by which a person passes judgment and associate affective responses to things.

Sentiment-related concepts can be characterized by at least three dimensions: type of sentiment/emotion/attitude, sentiment orientation (positive versus negative), and intensity of the sentiment. Our focus is on sentiment orientation—identifying a piece of text as having an overall positive or negative sentiment. Sentiment orientation has been referred to in the literature by various terms, including semantic orientation, affective orientation, semantic valence, and polarity. This study investigated the application of a machine-learning method called support vector machines (SVM) to develop a model for automatically classifying product reviews into two categories: recommended (positive sentiment) and not recommended (negative sentiment).

Automatic sentiment classification is useful in many areas. It can be used to classify product reviews into positive and negative reviews, so that potential customers can have an overall idea of how a product is perceived by other users [5–7]. A sentiment timeline system can track on-line discussion about movies and display a plot of the number of positive and negative sentiment messages over time [8]. A search engine can allow the user to specify the topic and the polarity of the desired reviews, such as recommended or not recommended [9]. It can help the user to focus on Web articles containing positive or negative comments, enabling the user to browse Web pages more efficiently. The technique can be used for filtering out email messages with impolite or abusive words [10]. In the area of social science research, it can be used to categorize news articles into those conveying positive and negative views, for various research purposes [11,12].

Though machine-learning techniques have long been used in topical text classification with good results, they are less effective when applied to sentiment classification [13]. Sentiment classification is a more difficult task compared to traditional topical classification, which classifies articles by comparing individual words (unigrams) in various subject areas.

In sentiment classification, unigrams may not be enough for accurate classification. For instance, the two phrases "you will be disappointed" and "it is not satisfactory" do not share the same words, but both express negative sentiments. Thus, this study investigates the effectiveness of using a machine-learning algorithm, support vector machine (SVM), on various text features including unigrams to classify on-line product reviews into recommended and not recommended. Especially we investigate the use of n-gram negation phrases identified through shallow linguistic processing to improve sentiment classification accuracy.

In the remainder of this paper, Section 2 discusses related works on sentiment classification. Sections 3 and 4 describe our experiment settings and investigation of several approaches to perform the classification using different text features: unigrams (individual words), selected words (such as verb, adjective, and adverb), and words labeled with part-of-speech tags. In Section 5, we analyze the product reviews that are wrongly classified by the SVM model to identify the sources of error and directions for improving the automatic classification. Section 6 describes the second part of the study that investigates the use of negation phrases identified through shallow linguistic processing, and finally Section 7 discusses future work.

2. Related work

Other researchers have carried out studies of automatic sentiment classification. Pang, Lee, and Vaithyanathan [14] examined the effectiveness of three machine-learning methods (Naïve Bayes, Maximum Entropy, and Support Vector Machines) for the sentiment classification of movie reviews (http://www.reviews.imdb.com/Reviews). They used mainly features based on unigrams (with negation tagging) and bigrams. A simple approach was used to handle negation phrases: the tag "NOT_" was added to every word between a negation word (e.g., "not") and the first punctuation mark following the negation word. This negation handling provided a negligible effect on performance since they used too simplistic approach to handle the negation problem. SVM provided the best accuracy, 82.9%, using unigrams with the presence or absence of a feature (unigrams with feature frequency decreased the accuracy to 72.8%).

Turney [15] used an unsupervised machine-learning technique to estimate the semantic orientation of a word based on its association with the words "excellent" and "poor," i.e., the extent to which the word co-occurs with "excellent" and "poor" in a text collection. Mutual information was used as the association measure and was computed using statistics (i.e., the number of hits returned) gathered by the Alta Vista search engine. The document phrases are bigrams, where one member of the pair is an adjective or an adverb and the second provides context. An average accuracy of 74% was achieved when the algorithm was evaluated on reviews from Epinions (http://www.epinions.com), sampled from four different domains (automobiles, banks, movies, and travel destination). The main limitation of this algorithm is the time required to calculate the semantic orientation of document phrases by sending queries to a search engine.

Dave, Lawrence, and Pennock [16] developed a method for automatic semantic classification of product reviews, collected from C|net (http://www.cnet.com) and Amazon.com. They used information retrieval techniques for feature extraction and classification rather than traditional machine-learning ones. Firstly a score for each term is calculated using the following equation:

$$\text{score}(f_i) = \frac{p(f_i|C) - p(f_i|C')}{p(f_i|C) + p(f_i|C')}$$

The normalized term frequency, $p(f_i|C)$, is determined by taking the number of times a feature f_i occurs in the class C and dividing it by the total number of tokens in C. Once each term has a score, the scores of the words in an unknown document are totaled, and the sign of the total is used to determine a class, i.e., negative or positive review. In their experiments, n-grams and arbitrary length of substring features increased accuracy of the classifier (this is contrary to the results obtained by other researchers using machine-learning techniques). They also tried to identify negation phrases and mark all words following the phrase as negated, such as converting "not good or useful" into "NOTgood NOTor NOTuseful." However, this simple approach did not improve accuracy.

Some researchers have attempted to develop methods for identifying the semantic orientation of adjectives using corpus statistics. Hatzivassiloglou and McKeown [17] developed an automatic method to find the semantic orientation (positive or negative) of adjectives from a large corpus. Their method is based on information extracted from conjunctions between adjectives in a large corpus. Conjoined adjectives usually are of the same orientation (e.g., "simple and well received"), whereas conjoined adjectives with "but" usually connect two adjectives of different orientations (e.g., "simplistic but well received"). They obtained more than 90% classification precision for adjectives that occur in a modest number of conjunctions in the corpus.

Wiebe [18] and Hatzivassiloglou and Wiebe [19] worked on automatic learning of subjective adjectives from corpora. The learned subjective adjectives can be used for subjectivity tagging, i.e., distinguishing sentences used to present opinions from sentences used to objectively present factual information. This subjectivity tagging can help a sentiment classifier to focus only on subjective sentences rather than objective sentences.

As related work, some researchers have worked on genre classification. Kessler, Nunberg, and Schutze [20] studied automatic detection of text genre using logistic regression and neural networks techniques. The genre they investigated were reportage, editorial, scientific/technical, legal, non-fiction, and fiction. In addition to structural cues (e.g., counts of a specific part of speech), surface cues including text features were explored: lexical cues (e.g., terms of address, such as Mr. and Mrs., which predominate in papers), character-level cues (e.g., counts of question marks), and derivative cues (e.g., average sentence length). They argued that genre classification based on surface cues was as successful as classification based on structural cues. Finn, Kushmerick, and Smyth [21] also investigated automatic genre detection which decides whether a document presents the opinion of its author or reports facts (i.e., genre of subjectivity). C4.5, a decision tree induction program, [22] was used with various text features: bag of words, parts of speech, and hand-crafted shallow linguistic features (e.g., average sentence length). For the part-of-speech approach, a document is represented as a vector of 36 part-of-speech features expressed as percentages of the total number of words for the document. They argued that the part-of-speech approach provided the best accuracy when the learned classifiers were generalized from the training corpus to a new domain corpus. Genre detection will be useful for extracting review (opinion) documents from various kinds of Web documents as a pre-processing task for sentiment classification.

3. Research method

3.1. Sampling

Using a spider program, product reviews were automatically downloaded from Review Centre (www.reviewcentre.com, 2003), which hosts millions of product reviews by consumers. After filtering out blank Web pages, a sample of 1800 product reviews was systematically selected, comprising 900 positive reviews and 900 negative reviews. The sample was divided into a training set of 1200 reviews (600 positive and 600 negative) for developing the classification model, and a test set of 600 reviews (300 positive and 300 negative) for evaluating the accuracy of the model. The majority of reviews are of mobile phones and electronic equipment.

Review Centre rates product reviews using a 10-star rating system. In this study, reviews with 7 stars or above are coded as recommended (positive), while reviews with 4 stars or below are non-recommended (negative). The ratings are generally consistent with the reviewer's comments, though there are some instances of inconsistency between the ratings and the reviewers' comments (see Section 5). The aim of the classification model is to predict from the natural language text of the review whether the review is coded as recommended or non-recommended.

3.2. Pre-processing

The texts of the reviews were tokenized and the words extracted were stemmed using the Conexor parser [23]. Each review was converted into a vector of terms (i.e., words) with term weights, indicating the importance of each term in the review. Three weighting schemes were investigated: term presence (binary weighting), term frequency (TF), and term frequency inverse document frequency (TFIDF). Term presence (binary weighting) has the value 1 if the term exists in the review, 0 otherwise. Term frequency (TF) uses the frequency of the term in the review as the weight. The TFIDF weight is defined by the formula:

$$\mathrm{TF} \times \log\left(\frac{N}{\mathrm{DF}}\right)$$

where TF is the number of times the term occurs in the current review document, N the number of reviews in the training set, and DF the document frequency—the number of reviews in the training set containing the term. The TFIDF weight has been used in many studies on topical text classification.

3.3. Machine-learning methods

A machine-learning method, support vector machine (SVM), was used in this study. SVM has been applied to text classification in Joachims' study [24], and later used in many studies [25,26]. The core idea is to find a hyperspace surface H, which separates positive and negative examples with the maximum distance. In our study, SVM^{light} (www.svmlight.joachims.org), a publicly available SVM program, was used for automatic review classification.

Many studies have found SVM to perform better than other machine-learning algorithms. Joachims [27] and Yang and Liu [28] found that SVM worked better than decision tree induction on text classification. However, a decision tree model is easy to interpret and can be converted to IF-THEN rules. Yang [29] claimed that SVM and k-NN methods were significantly better than other classifiers. Sebastiani reported that SVM delivered very good performance in some experiments [30].

3.4. Approaches investigated

Different kinds of linguistic features were investigated in developing the classification models:

- Unigram (baseline)—Using all the individual stemmed words (unigrams) that appeared in product reviews.
- Selected words (such as verbs, adjectives, and adverbs)—Conexor parser was used to tag individual words with part-of-speech tags. Only words with specific parts of speech, such as verb, adjective, and adverb, were used in developing the classifier.
- Words labeled with part-of-speech tags—the individual words were combined with their part-of-speech tags. For instance, the words "better:adjective" and "better:verb" were considered different terms.

4. Results

Table 1 lists the results for the various approaches attempted in this study. The use of unigrams, the simplest approach, obtained 75.39% accuracy (average of ID1, ID2, and ID3). The TFIDF weighting that is effective in traditional topical text classification performed a little better than TF and Presence when applied to sentiment classification in this study (ID3). Document frequency (DF) was used to retain (or consider) only words that occur in at least the specified number of documents in the training set. In general, the value 3 for DF performed a little better than the value 1 (ID3 and ID4).

The use of additional part-of-speech information did not improve results, possibly because it increased the number of dimensions (each word is subdivided into different part of speech)

Table 1
Various approaches and results

ID	Approach	Selected terms	Term weighting	DF	Terms labeled with POS tags	Negation	Accuracy (%)
1	Unigram with TF	All	TF	3	No	No	74.17
2	Unigram with Presence	All	Presence	3	No	No	75.50
3	Unigram with TFIDF	All	TFIDF	3	No	No	76.50
4	Unigram with TFIDF and DF = 1	All	TFIDF	1	No	No	74.17
5	Unigram labeled with POS	All	TFIDF	3	Yes	No	75.83
6	Unigram with selected words (V, A, Adverb)	Verb, adjective, adverb	TFIDF	3	No	No	77.33
7	Unigram with selected words (N, V, A, Adverb)	Noun, verb, adjective, adverb	TFIDF	3	No	No	75.50

and reduced the term weight (such as TF) for each term (ID5). Limiting the terms to just verbs, adjectives and adverbs improved the accuracy rate: 77.33% (ID6). This indicates that positive and negative sentiments are expressed mostly through verbs, adjectives, and adverbs. But when we included nouns in addition to verbs, adjectives, and adverbs, the accuracy rate degraded a little bit (ID7).

5. Error analysis

Generally, when applied to topical text classification, the accuracy of SVM is above 85% [31,32]. Thus, we reapplied the learned SVM model to the training set (using unigram, DF = 3 and TF options) to identify the sources of error and directions for improving the automatic classification. Out of 1200 training set documents, the total number of wrongly classified documents was 87.

The possible reasons for failure in automatic classification are summarized in Table 2 and explained as follows (note that some documents are counted multiple times since they have more than one reason for misclassification):

1. Negation phrase. Negation phrases in the reviews affected the effectiveness of the simple unigram-based classifier. For instance, the sentence “I’d never regretted purchasing it” is

Table 2
Error analysis

Reasons	Number of documents
Negation phrase	34
Comments on parts of the product	25
Need inferencing	17
Inconsistency between rating and comments	13
Comments on other products	10

actually a positive comment. However, the unigram approach treats "never" and "regretted" as separate negative words. This seems to be one of the most common problems in sentiment classification.
2. Comments on parts of the product. Sometimes, though the reviewer comments negatively on parts of the product, he is actually satisfied with the product as a whole, e.g., "The best phone I've had yet. The ONLY bad point is that. . .".
3. Need inferencing. Some comments are complex and need inferencing to identify the sentiment category. The sentence "if the price dropped, the company would be surprised how it would sell" contains no apparent positive or negative words.
4. Inconsistency between rating and comments. In 13 cases, there is no obvious relation between the reviewer's comments and the number of stars given. For instance, the comment "Good if you constantly listen to music on the move. This phone is still the best looking phone on the market" is apparently positive; however, the reviewer gave it 3 stars.
5. Comments on other products. The reviewer uses indicative words to comment on or make comparisons with other related products. For example, "8210 is better. More valuable".

In addition, some reviews are too short to be classified accurately (23 documents are no longer than two lines). For example, the comment "This is an OK phone but slow" is difficult to classify without more context.

6. The negation phrase approach

As shown in Table 2, the errors attributed to negation phrase, comments on parts, and need inferencing account for a large portion of the wrong classifications. Thus, we investigated the use of simple linguistic processing to address the problems of negation phrase. Each negation (not, never, or no) and its adjacent words are combined to generate a new composite term (i.e., negation phrase). To extract negation phrases, we constructed syntactic patterns for negation phrases, such as "<Verb>-Not-<Verb>" and "<Verb>-Not-<Adverb>-<Adjective>" after carefully reviewing various negation expressions in the product reviews.

The complete list of "not" negation rules that are used in the study is shown in Fig. 1a. "Not" negation rules form negation phrases like "not good", "not very beautiful", "do not mind", etc. Fig. 1b displays "never" negation rules. "Never" negation rules generate patterns like "never say goodbye", "never grow old", "never adequately explained", etc. A complete "no" negation rules can be seen in Fig. 1c. "No" negation rules generate negation phrases like "no access", "no complaints", "no doubt", etc. The meanings of the negation rule symbols in Fig. 1 are described in Table 3.

Negation phrases are extracted by using a simple algorithm. The algorithm first arranges the tokens in the data set into a list of sentences. Adjacent tokens in each sentence are concatenated to form a token string. The token string is checked against negation rules and if it matches with one of the rules, a negation phase is formed and treated as a single term. If the

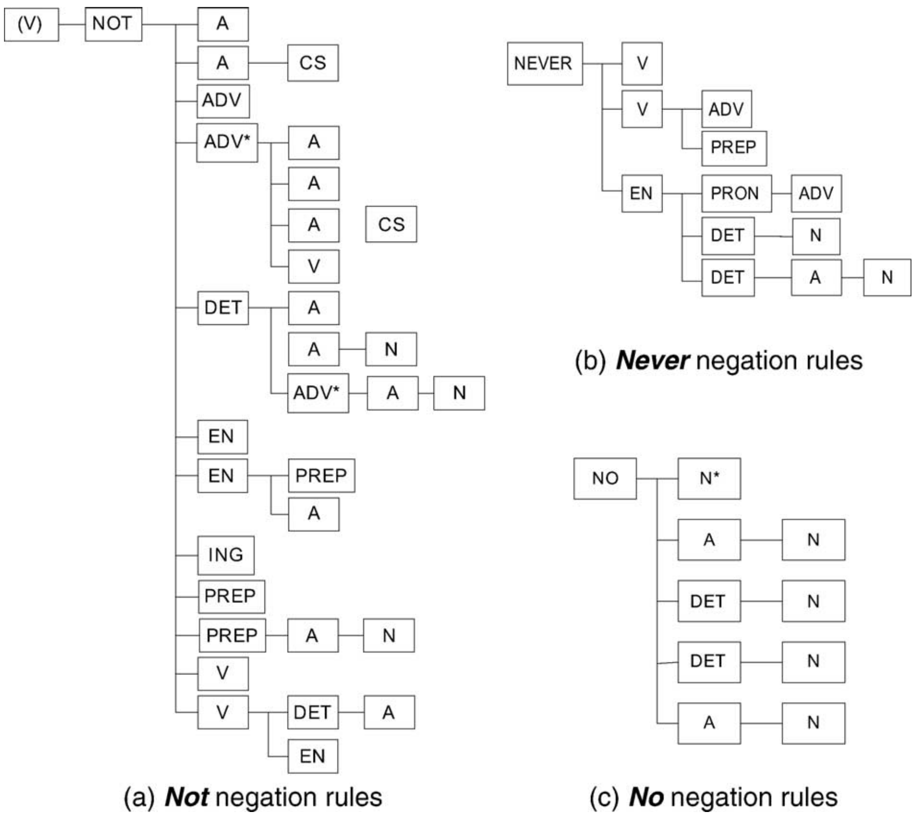

Fig. 1. Negation rules.

token string does not match any of the negation rules, the original tokens in the token string are returned. Table 4 lists some samples of negation phrases extracted automatically. For instance, "do not buy" occurs in 34 reviews out of 1200 training reviews.

Table 3
Meanings of negation rule symbols

Symbol	Meaning
A	Adjective
ADV	Adverb
CS	Subordinating conjunction
DET	Determiner
EN	Past participle
ING	Present participle
N	Noun
PRON	Pronoun
PREP	Preposition
V	Verb
( )	Optional
*	Multiple

Table 4
Negation phrases

Negation phrases	DF
Do not buy	34
Do not work	24
Would not recommend	14
Do not want	14
Do not like	9
Be not worth	6
Not bad	5
Not the good	5
Have not regret	4
Will not work	4
Do not recommend	3
Will not regret	3
Be not as good as	3
Would not buy	2
Be not very impressed	2
Be not happy	2
Not so bad	2
Do not purchase	1
Do not dislike	1
Not so good	1
Not too bad	1
Be not a good choice	1

Table 5 lists the results for the negation phrase approaches where negation phrases were treated as unique terms. This approach improved the accuracy rate to 79.33%. From the results and error analysis, linguistic processing of text appears to be useful for sentiment classification. We expect that deeper linguistic processing will further increase the accuracy of sentiment classification.

7. Conclusion

From our experiments, compared with the "unigram" approach, the use of "negation phrase" through simple linguistic processing improved classification accuracy. This result and

Table 5
Negation phrase approaches and results

ID	Approach	Selected terms	Term weighting	DF	Terms labeled with POS tags	Negation	Accuracy (%)
1	Unigram with negation phrase and DF = 3	All	TFIDF	3	No	Yes	78.33
2	Unigram with negation phrase and DF = 1	All	TFIDF	1	No	Yes	79.33

the error analysis of wrongly classified documents suggest that the simple unigram approach for sentiment classification is not good enough. Some form of deeper linguistic processing is required for sentiment classification, such as further syntactic and semantic processing and discourse analysis.

As future work, we plan to improve the current negation phrase processing. Since there are many variations of negation phrases, we need to generalize various negation phrases. For instance, currently we consider "do not buy", "not to buy", and "would not buy" as different negation phrases but plan to generalize them into one negation phrase, "not buy". This will reduce the number of variations and increase frequencies of negation phrases. Also we plan to filter unimportant negation phrases using various feature selection techniques, such as information gain [33]. Further work will include development of an opinion sentence tagging tool that distinguishes sentences that present opinions from sentences that objectively present factual information. This will help the sentiment classifier to focus more on opinion sentences. Finally, ordinal classification (e.g., very positive, positive, neutral, negative, and very negative) in addition to binary classification will be investigated to support more fine-grained sentiment classification.

References

[1] Kessler, B., Nunberg, G., & Schutze, H. (1997). Automatic detection of text genre. *Proceedings of the Eighth Conference On European Chapter of the ACL (Association for Computational Linguistics)* (pp. 32–38).

[2] Lee, Y. -B., & Myaeng, S. H. (2002). Text genre classification with genre-revealing and subject-revealing features. *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* Tampere, Finland (pp. 145–150).

[3] Downes, W. (2000). The language of felt experience: Emotional, evaluative and intuitive. *Language and Literature*, *9*(2), 99–121.

[4] *The American Heritage Dictionary* (4th ed.). (2001). New York: Dell Publishing. *An introductory tour through appraisal theory* (n.d.). Retrieved December 1, 2004, from http://www.grammatics.com/appraisal/AppraisalOutline/UnFramed/AppraisalOutline.htm

[5] Turney, P. D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *Proceedings of the 40th Annual Meeting of the ACL (Association for Computational Linguistics)* Philadelphia, PA, July 7–12 (pp. 417–424).

[6] Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment classification using machine-learning techniques. *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing,* Philadelphia, PA, July 6–7 (pp. 79–86).

[7] Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the 12th International World Wide Web Conference,* Budapest, Hungary, May 20–24 (pp. 519–528).

[8] Tong, R. M. (2001). An operational system for detecting and tracking opinions in on-line discussions. *Working Notes of the ACM SIGIR 2001 Workshop on Operational Text Classification* (pp. 1–6).

[9] Hearst, M. A. (1992). Direction-based text interpretation as an information access refinement. In P. Jacobs (Ed.), *Text-Based Intelligent Systems: Current Research and Practice in Information Extraction and Retrieval*. Mahwah, NJ: Erlbaum.

[10] Spertus, E. (1997). Somkey: Automatic recognition of hostile messages. *Proceedings of the 5th International Conference on Intelligent User Interfaces,* Providence, RI, July 27–31 (pp. 1058–1065).

[11] Semetko, H. A., & Valkenburg, P. M. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, *50*(2), 93–109.

[12] Lind, R. A., & Salo, C. (2002). The framing of feminists and feminism in news and public affairs programs in U.S. Electronic Media. *Journal of Communication*, *52*(1), 211–228.
[13] Pang, et. al.
[14] Ibid.
[15] Turney.
[16] Dave.
[17] Hatzivassiloglou, V., & McKeown, K. R. (1997). Predicting the semantic orientation of adjectives. *Proceedings of the 8th Conference of the European Chapter of the ACL (Association for Computational Linguistics),* Madrid, Spain, July (pp. 174–181).
[18] Wiebe, J. M. (2000). Learning subjective adjectives from corpora. *Proceedings of the 17th National Conference on Artificial Intelligence and 12th Conference on Innovative Applications of Artificial Intelligence* (pp. 735–740).
[19] Hatzivassiloglou, V., & Wiebe, J. M. (2000). Effects of adjective orientation and gradability on sentence subjectivity. *Proceedings of 17th International Conference on Computational Linguistics, Saarbrücken, Germany* (pp. 299–305).
[20] Kessler.
[21] Finn, A., Kushmerick, N., & Smyth, B. (2002). Genre classification and domain transfer for information filtering. *Proceedings of ECIR-02, 24th European Colloquium on Information Retrieval Research* (pp. 353–362).
[22] Quinlan, R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufman.
[23] Tapanainen, P., & Järvinen, T. (1997). A non-projective dependency parser. *Proceedings of the 5th Conference on Applied Natural Language Processing,* Washington, WA, March 3 (pp. 64–71).
[24] Joachims, T. (1998). Text categorization with support vector machines: Learning with many relevant features. *Proceedings of 10th European Conference on Machine-learning,* Chemnitz, Germany, April 21–24 (pp. 137–142).
[25] Joachims, T. (1999). Transductive inference for text classification using support vector machines. *Proceedings of 16th International Conference on Machine-learning,* Bled, Slovenia, June, 27–30 (pp. 200–209).
[26] Schohn, G., & Cohn, D. (2000). Less is more: Active learning with support vector machines. *Proceedings of the 17th International Conference on Machine-Learning,* Stanford, CA, June 2 (pp. 839–846).
[27] Joachims (1998).
[28] Yang, Y., & Liu, X. (1999). A re-examination of text categorization methods. *Proceedings of 22nd ACM International Conference on Research and Development in Information Retrieval,* Berkeley, CA, August 16–19 (pp. 42–49).
[29] Yang, Y. (1999). An evaluation of statistical approaches to text categorization. *Information Retrieval*, *1*(1–2), 69–90.
[30] Sebastiani, F. (2002). Machine-learning in automated text categorization. *ACM Computing Surveys*, *34*(1), 1–47.
[31] Joachims (1998).
[32] Yang and Liu.
[33] Yang, Y., & Pedersen, J. (1997). A comparative study on feature selection in text categorization. *Proceedings of ICML-97, 14th International Conference on Machine Learning* (pp. 412–420).