A Lexicon-Enhanced Method for Sentiment Classification: An Experiment on Online Product Reviews

Yan Dang, Yulei Zhang, and Hsinchun Chen, *University of Arizona*

A proposed lexicon-enhanced method for sentiment classification combines machine-learning and semantic-orientation approaches into one framework that significantly improves sentiment-classification performance.

As an emerging communication platform, Web 2.0 has led the Internet to become increasingly user-centric. People are participating in and exchanging opinions through online community-based social media, such as discussion boards, Web forums, and blogs. Along with such trends, an increasing amount of user-generated content containing rich opinion and sentiment information has appeared on the Internet. Understanding such opinion and sentiment information has become increasingly important for both service and product providers and users because it plays an important role in influencing consumer purchasing decisions.[1]

Sentiment-classification techniques can help researchers study such information on the Internet by identifying and analyzing texts containing opinions and emotions (also referred to as direction-based text). They can help determine whether a text is objective or subjective and whether a subjective text contains positive or negative sentiments.[2] We can classify the approaches adopted by previous sentiment-classification studies into two categories.[3] In the *machine-learning approach*,[4] each classifier is trained using a collection of representative data. In contrast, the *semantic-orientation approach* does not require prior training; instead, it measures a word containing positive or negative sentiment.[5,6] Each approach has its own benefits and drawbacks. For example, the machine-learning approach tends to be more accurate, but the semantic-orientation approach has better generality.[7,8]

Few studies have combined these approaches to improve sentiment-classification performance. This study leverages both approaches by combining them into one framework. To do this, we developed a lexicon-enhanced method to generate a set of sentiment words based on a sentiment lexicon as a new feature dimension. We combined these sentiment features with content-free and content-specific features used in the existing machine-learning approach. To demonstrate our proposed method's performance, we conducted experimental studies using five sets of online product reviews. Our experimental results on different data

1541-1672/10/$26.00 © 2010 IEEE
Published by the IEEE Computer Society

sets consistently show that adding the new set of sentiment features can increase sentiment-classification performance. We also found that conducting feature selection can further improve the performance, especially for large data sets.

Sentiment-Classification Approaches and Applications

In general, sentiment analysis is concerned with analyzing direction-based text.[2] Determining whether a text is objective or subjective and whether a subjective text contains positive or negative sentiments is a common two-class problem that involves classifying sentiments as positive or negative.[4,5] Additional variations include classifying sentiments as opinionated/subjective or factual/objective.[9] Some studies have attempted to classify emotions (such as happiness, sadness, anger, or horror) instead of sentiments.[10]

The machine-learning approach treats the sentiment-classification problem as a topic-based text-classification problem.[3] Any text-classification algorithm can be employed, such as Naïve Bayes or support vector machines (SVMs). Bo Pang and her colleagues experimented with this approach to classify movie reviews into two classes: positive and negative.[4] They compared Naïve Bayes, Maximum Entropy, and SVM and achieved the highest classification accuracy (82.9 percent) using SVM.

The semantic-orientation approach, on the other hand, performs classification based on positive and negative sentiment words and phrases contained in each evaluation text and mining the data requires no prior training.[3,8] Two types of techniques have been used in previous semantic-orientation-approach-based sentiment-classification research: corpus-based and dictionary-based.[3] The corpus-based techniques aim to find co-occurrence patterns of words to determine their sentiments. Researchers have proposed different strategies to determine sentiments. For example, Peter Turney calculated a phrase's semantic orientation to be the mutual information between the phrase and the word "excellent" (as the positive polarity) minus the mutual information between the phrase and the word "poor" (as the negative polarity).[5] Ellen Riloff and Janyce Wiebe used a bootstrapping process to learn linguistically rich patterns of subjective expressions to distinguish subjective expressions from objective expressions.[6] Dictionary-based techniques

> A text's overall sentiment is determined by the sentiments of a group of words or phrases appearing in the text.

use synonyms, antonyms, and hierarchies in WordNet (or other lexicons with sentiment information) to determine word sentiments.[11] Building upon WordNet, SentiWordNet is a lexical resource for sentiment analysis that has more sentiment-related features.[12] It assigns to each synset of WordNet three sentiment scores regarding positivity, negativity, and objectivity, respectively. SentiWordNet has been used as the lexicon in recent sentiment classification studies.[13–15]

The corpus-based techniques, however, often rely on a large corpus to calculate the statistical information needed to decide the sentiment orientation for each word or phrase. Therefore, they might not be as efficient as the dictionary-based techniques. Still, a good lexicon is critical for the dictionary-based techniques.

Sentiment classification has been applied to online reviews, Web discourse,[16] and online news.[9] For example, one study used sentiment classification to distinguish objective from subjective news articles in a *Wall Street Journal* collection. For online reviews, previous research has looked at movie, product, and music reviews.[4,5] This work is challenging because movie reviews often contain long summaries of plot details and a single product review might express both positive and negative sentiments about particular aspects of the product.

Sentiment-Classification Features

Previous machine-learning-approach-based sentiment-classification studies have mainly adopted content-free or content-specific features.

Content-free features (used as the baseline features in this study) include lexical, syntactic, and structural features.[17] Lexical features are character- or word-based statistical measures of lexical variation. They mainly include character-based lexical features, vocabulary richness measures, and word-based lexical features. Syntactic features indicate the patterns used to form sentences, such as function words, punctuation, and parts of speech. Structural features show the text organization and layout, such as greetings, signatures, the number of paragraphs, and the average paragraph length. Other structural features include the use of various file extensions, fonts, sizes, colors, and so forth. Content-specific features consist of important keywords and phrases on certain topics, such as word n-grams. Previous studies have shown that these features help improve text-classification performance.[17]

In previous semantic-orientation-approach-based sentiment-classification studies, a text's overall sentiment is determined by the sentiments of a group of words or phrases appearing in the text. Different categories of words or phrases have been used to determine this. For example, some approaches use adjectives appearing in a text,[18,19] while others use all the two-word phrases containing adjectives or adverbs in a given text[5] or using the combination of adjectives, verbs, and nouns.[14] Our work refers to such sentiment words and phrases as *sentiment features* and incorporates them into the machine-learning approach as an additional feature dimension.

In text-classification studies, the full text is often transformed to a vector of features describing the text's content to reduce its complexity. However, not all features are necessary or sufficient to learn the concept of interest and many are noisy or redundant. Feature selection, which aims to identify a minimal-sized subset of features relevant to the target concept, can be applied.[20]

Design, Implementation, and Evaluation

Our proposed lexicon-enhanced method for sentiment classification consists of three main steps: data acquisition, feature generation, and classification and evaluation.

Data Acquisition

For this study, we used online product reviews as the application domain because of their increasing importance in influencing individuals' purchasing decisions. As shown in recent research compiled by PriceRunner.co.uk, the UK's most comprehensive and impartial price-comparison service, people rate online reviews of products and companies (98 percent) to be a more important factor than recommendations from friends (88 percent) during their buying process.[21] The results of a large-scale survey with 2,100 participants showed that online shopping consumers are almost four times more likely to rely on online reviews from strangers (20 percent believed this to be the most important factor) than advice from a friend (6 percent).[21] However, the sheer volume of opinion and sentiment information posted online (such as online product reviews) is overwhelming and grows exponentially each day, making it difficult for individuals to locate and digest the information they need.

> Online shopping consumers are almost four times more likely to rely on online reviews from strangers than advice from a friend.

We used spidering programs to collect textual data from online sources as HTML pages. Parsing programs then parsed out the review data and stored it in a relational database. The first testbed we collected using this approach consists of digital camera reviews from Epinions (epinions.com), a well-known public product review data source that has served as the testbed for many previous sentiment-classification studies.[5,15] Each Epinions review has a polarity rating from one to five, with one star being the most negative and five stars being the most positive. The middle point, three, is the cutoff. In total, we collected 307 negative reviews and 1,499 positive reviews. To keep the training set relatively balanced, we used all 307 negative reviews and randomly selected 307 positive reviews.

To make our experiment scientifically more rigorous, we included another four testbeds from the well-known, publicly available Blitzer's multidomain sentiment data set.[22] Each of the four testbeds we used in this study contains reviews on books, DVDs, electronics, and kitchen appliances, respectively, and each has 1,000 positive and 1,000 negative reviews.

Feature Generation

As we mentioned earlier, our proposed sentiment classification method uses three types of features: content-free (F1), content-specific (F2), and sentiment (F3) features. Among them, F1 and F2 come from the machine-learning approach, and F3 is from the semantic-oriented approach.

For each testbed, according to previous sentiment-analysis research, we utilized 87 lexical features, 158 syntactic features (150 function words and eight punctuation marks), and five structured features, for a total of 250 F1 features. We used unigrams and bigrams as F2 features. After removing the semantically empty stop words, we kept the unigrams and bigrams that appeared more than five times. The number of F2 features varies for each testbed and is much larger (with thousands or tens of thousands of features) than the number of F1 features.

To extract F3 features, we first conducted part-of-speech (POS) tagging on each data collection. We then calculated each word's sentiment score by

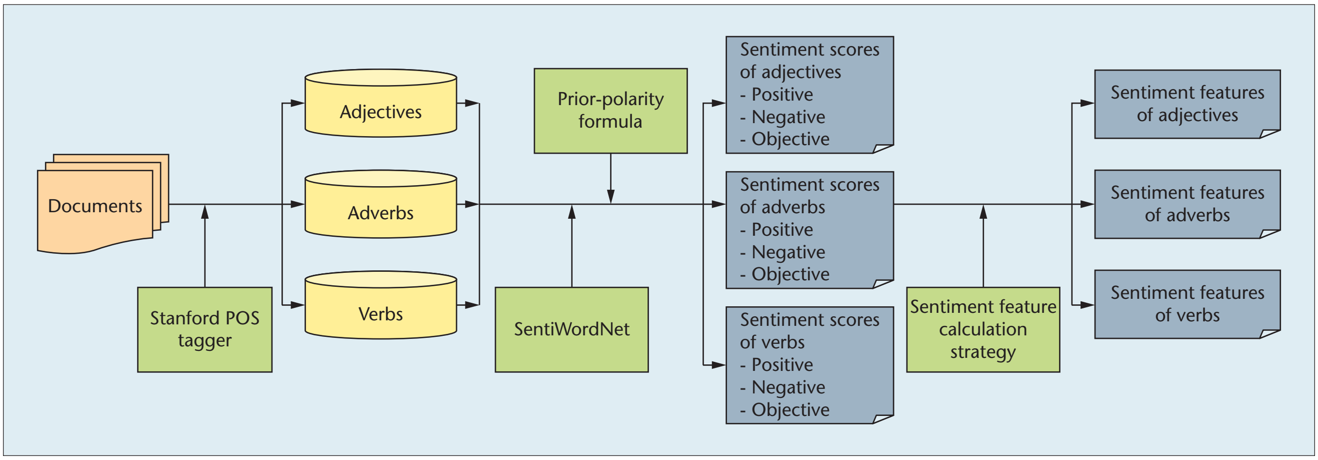

Figure 1. Extracting sentiment features. We used the Stanford POS tagger to perform the tagging and used all the adjectives, adverbs, and verbs as the sentiment features.

looking up a sentiment-based lexicon. We used the dictionary-based technique to generate sentiment features instead of the corpus-based technique[5] because the latter relies on a large corpus to calculate statistical information and thus might not be as efficient as the dictionary-based technique in general.[19] Previous studies also emphasized the need and importance of sentiment-based lexicons,[19] which can be effectively and efficiently used to decide each word's semantic orientation.

Figure 1 shows the detailed process for extracting sentiment features. First, we used the Stanford POS tagger (http://nlp.stanford.edu/software/tagger.shtml) to perform the tagging. As previous literature suggests, we used all the adjectives, adverbs, and verbs as the sentiment features.[18,19,23] We did not include nouns since they are more context dependent.

Second, we used SentiWordNet to determine the sentiment scores of the extracted adjectives, adverbs, and nouns. Because each word in SentiWordNet has multiple senses, we calculated the average of the three polarity scores (positive, negative, and objective) for its adjective, adverb, and verb senses separately using the prior-polarity formula adopted from previous literature.[14,15] That is,

If $\text{Score}(\text{word} = \text{POS})_{\text{objective}} > 0.5$
    We consider the word in the given POS sense to be objective and exclude it from our sentiment feature set.
Else
    If $\text{Score}(\text{word} = \text{POS})_{\text{positive}} > \text{Score}(\text{word} = \text{POS})_{\text{negative}}$
        We add $(\text{word} = \text{POS}, |\text{Score}(\text{word} = \text{POS})_{\text{positive}}|)$ to our sentiment feature set.
    If $\text{Score}(\text{word} = \text{POS})_{\text{positive}} < \text{Score}(\text{word} = \text{POS})_{\text{negative}}$
        We add $(\text{word} = \text{POS}, -|\text{Score}(\text{word} = \text{POS})_{\text{negative}}|)$ to our sentiment feature set.
    If $\text{Score}(\text{word} = \text{POS})_{\text{positive}} = \text{Score}(\text{word} = \text{POS})_{\text{negative}}$
        We exclude it from our sentiment feature set.

Figure 2. The sentiment feature-calculation strategy. We used 0.5 (the midpoint of the 0 to 1 scale) to differentiate subjective words from objective words in a given POS sense.

$$\text{Score}(\text{word} = \text{POS})_i = \left(\Sigma_{k \in \text{SentiWordNet}(\text{word=POS}\ \&\ \text{polarity}=i)} \text{SentiWordNet\_Score}(k)_i\right) / |\text{synsets}(\text{word} = \text{POS})|$$

where $\text{POS} \in$ {adjective, adverb, verb}, $i \in$ {positive, negative, objective}, and k denotes the synsets of a given word in a particular sense. For example, the word "bad" has 14 synsets with an adjective POS sense, two with an adverb POS sense, and no synsets with a verb POS sense. Therefore, we calculated the following scores based on the prior-polarity formula:

$\text{Score}(\text{“bad”} = \text{adjective})_{\text{positive}} = 0.11$,
$\text{Score}(\text{“bad”} = \text{adjective})_{\text{negative}} = 0.64$,
$\text{Score}(\text{“bad”} = \text{adjective})_{\text{objective}} = 0.25$,
$\text{Score}(\text{“bad”} = \text{adverb})_{\text{positive}} = 0.06$,
$\text{Score}(\text{“bad”} = \text{adverb})_{\text{negative}} = 0.56$, and
$\text{Score}(\text{“bad”} = \text{adverb})_{\text{objective}} = 0.38$.

For each word in a given POS sense (adjective, adverb, or verb), we obtained three scores regarding positivity, negativity, and objectivity, respectively. We developed a feature-calculation strategy to determine the word's final score as a sentiment feature. As Figure 2 shows, we used 0.5 (the midpoint of the 0 to 1 scale) to differentiate the subjective words from the objective words in a given POS sense. For a word with objective scores smaller than or equal to 0.5, we then compared its positive and negative scores. If the positive score was greater than the negative score, we treated the word (in the given POS sense) as a positive sentiment feature. Otherwise, we treated it as a negative sentiment feature. In addition, we excluded words that had equal positive and negative scores.

Table 1. The number of features in each feature set.

Feature set	Number of features				
	Digital cameras	Books	DVDs	Electronics	Kitchen appliances
F1	250	250	250	250	250
F1 + F2	17,619	8,138	7,992	5,282	4,670
F1 + F3	887	1,141	1,090	641	582
F1 + F2 + F3	18,256	9,029	8,832	5,673	5,002
Selected F1 + F2 *	3,012–3,978	524–555	597–668	612–670	545–592
Selected F1 + F2 + F3 *	3,085–4,055	567–606	647–724	658–715	583–627

*Because the number of selected features varies in each fold when conducting the 10-fold cross validation, we list the number range instead of 10 individual numbers for each selected feature set on each data set.

Figure 2 shows detailed calculations. For the previous example of the word "bad," using this strategy resulted in two sentiment features: ("bad" = adjective, −0.64) and ("bad" = adverb, −0.56).

An accurate sentiment-based lexicon is important to ensure a good performance of F3 features because any error in this step will propagate to the final classification results. We manually checked the synsets in SentiWordNet related to the words having apparent semantic orientations. We found a few words that are always used to express positive feelings (such as love, enjoy, favorite, perfect, and great) or negative feelings (such as horrible, weak, useless, stupid, and silly), but their objective scores were assigned greater than 0.5 in SentiWordNet. We then manually added these words to our sentiment feature collection.

Among the three types of features, F1 and F3 features are domain independent and F2 features are domain dependent. To test the performance of different types of features, especially the newly introduced F3 features, we created four feature sets. We first built three feature sets incrementally:

1. Feature set F1 includes content-free features.
2. Feature set (F1 + F2) consists of content-free and content-specific features.
3. Feature set (F1 + F2 + F3) consists of content-free, content-specific, and sentiment features.

This incremental order represents the evolutionary sequence of features used for online text classification.[24] In addition, because both F1 and F3 features are domain independent, we composed another feature set (F1 + F3). Among the four feature sets, F1 and (F1 + F3) are domain independent, and (F1 + F2) and (F1 + F2 + F3) are domain dependent (because F2 features are domain dependent).

When the number of features is large, feature selection might improve classification performance by selecting an optimal subset of features.[25] Previous classification studies using n-gram features usually included some form of feature selection to extract the most important words or phrases.[26] We used the Information Gain (IG) heuristic to conduct feature selection due to its reported effectiveness in previous online text-classification research.[2,26] All the features with an information gain greater than 0.0025 were selected.[2] Thus, we built two selected feature sets (F1 + F2) and (F1 + F2 + F3).

Table 1 summarizes the number of features in each feature set for all five testbeds. Feature sets (F1 + F2) and (F1 + F2 + F3) are much larger than feature sets F1 and (F1 + F3) due to the large number of F2 features. After feature selection, the sizes of both feature sets (F1 + F2) and (F1 + F2 + F3) were reduced significantly.

Classification and Evaluation

To examine our proposed method, we compared the performances of different feature sets using a SVM as the classifier because of its reported performance in previous sentiment-analysis studies.[2] For each testbed, we randomly chose 90 percent of the reviews as training data and the remaining 10 percent as testing data. We used 10-fold cross validation to conduct the evaluation. Table 2 summarizes the performance measures in terms of overall accuracy, average precision, average recall, and average F-measure for all five testbeds.

Our method achieved classification accuracy comparable to previous sentiment-classification research that also used Blitzer's multidomain sentiment data set and adopted unigrams and bigrams as features.[22,27] The accuracy differences between Blitzer's study and ours can be attributed to two causes. First, the evaluation procedures of the two studies differed. For each data set, Blitzer used 80 percent of the reviews as the training data and the remaining 20 percent as the testing data. Also, Blitzer's study did not use cross validation. Second, some parameters used in the two studies also differed, such as the use of stop words and the feature-filtering criteria.

Table 2. Experimental results for different feature sets.*

Measure	Feature set	Data set				
		Digital cameras	Books	DVDs	Electronics	Kitchen appliances
Overall accuracy (%)	F1	74.92	70.55	65.45	69.85	69.95
	F1 + F2	80.94	75.30	78.00	80.65	82.40
	F1 + F3	81.92	75.55	72.00	76.35	77.85
	F1 + F2 + F3	**83.22**	76.95	78.40	80.90	83.30
	Selected F1 + F2	78.50	77.45	80.45	82.80	83.40
	Selected F1 + F2 + F3	79.32	**78.85**	**80.75**	**83.75**	**84.15**
Average precision (%)	F1	75.20	70.60	65.50	69.90	70.27
	F1 + F2	81.10	75.30	78.00	80.70	82.40
	F1 + F3	82.00	75.60	72.00	76.40	78.07
	F1 + F2 + F3	**83.30**	77.00	78.40	80.90	83.30
	Selected F1 + F2	79.40	77.74	80.63	82.97	83.48
	Selected F1 + F2 + F3	80.10	**79.15**	**80.95**	**83.85**	**84.28**
Average recall (%)	F1	74.90	70.60	65.50	69.90	69.95
	F1 + F2	80.90	75.30	78.00	80.70	82.40
	F1 + F3	81.90	75.60	72.00	76.40	77.85
	F1 + F2 + F3	**83.20**	77.00	78.40	80.90	83.30
	Selected F1 + F2	78.49	77.45	80.45	82.80	83.40
	Selected F1 + F2 + F3	79.31	**78.40**	**80.75**	**83.75**	**84.15**
Average F-measure (%)	F1	75.05	70.60	65.50	69.90	70.11
	F1 + F2	81.00	75.30	78.00	80.70	82.40
	F1 + F3	81.95	75.60	72.00	76.40	77.96
	F1 + F2 + F3	**83.25**	77.00	78.40	80.90	83.30
	Selected F1 + F2	78.94	77.59	80.54	82.89	83.44
	Selected F1 + F2 + F3	79.70	**78.99**	**80.85**	**83.80**	**84.21**

*The bold-faced values indicate the best performances.

The values of all four measures increased as additional types of features were added. The feature set (F1 + F2 + F3) outperformed both feature sets (F1 + F2) and (F1 + F3), each of which in turn outperformed the feature set F1. The increases from the feature set F1 to (F1 + F3) and from the feature set (F1 + F2) to (F1 + F2 + F3) indicated the considerable contribution of the newly introduced F3 features. Although in most cases the increases from the feature set F1 to (F1 + F2) and from the feature set (F1 + F3) to (F1 + F2 + F3) were relatively larger, they were caused by a much larger number of F2 features, which are domain dependent. Therefore, we believe the F3 features introduced in this study can play an important role in improving sentiment-classification performance.

Except for the data set on digital cameras, all the other testbeds showed that feature selection increased the classification performance on all four measurement dimensions. As Table 2 indicates, the best performances for all five testbeds was achieved by combining all three types of features and conducting feature selection for the four testbeds from Blitzer's multidomain sentiment data set. The fact that feature selection did not improve the performance on the digital camera testbed is not surprising. This testbed has fewer reviews than the other four, and generalizing the feature-selection model achieved from the training data to the testing data often requires numerous training data points.

These results show that adding the newly introduced sentiment features, which are often used in the existing semantic-orientation approach, and the content-free and content-specific features that come from the existing machine-learning approach can improve sentiment-classification performance significantly. The better

THE AUTHORS

Yan Dang is a doctoral candidate in the Department of Management Information Systems and a research associate in the Artificial Intelligence Lab at the University of Arizona. Her research interests include text mining and human–computer interaction. Dang has an MS in computer science from Shanghai Jiao Tong University. Contact her at ydang@email.arizona.edu.

Yulei Zhang is a doctoral candidate in the Department of Management Information Systems and a research associate in the Artificial Intelligence Lab at the University of Arizona. His research interests include text mining, data mining, and bioinformatics. Zhang has an MS in bioinformatics from Shanghai Jiao Tong University. Contact him at ylzhang@email.arizona.edu.

Hsinchun Chen is the McClelland Professor of Management Information Systems and the director of the Artificial Intelligence Lab at the University of Arizona. His research interests include knowledge management, business intelligence, and data and Web mining. Chen has a PhD in information systems from New York University. He is a Fellow of IEEE and the American Association for the Advancement of Science (AAAS). Contact him at hchen@eller.arizona.edu.

performance can be attributed to the rich polarity information introduced by the sentiment features. Conducting feature selection was also helpful.

One direction for further study in this area is to refine the lexicon and extend the sentiment feature-extraction procedure. Further research can also explore other sentiment feature-generation methods, such as corpus-based techniques, and compare their performance. In addition, feature selection on large feature sets has shown to improve the classification performance on relatively large data sets. Comparing different feature-selection algorithms to find the best one for our proposed sentiment-classification method could be an additional future research direction. Moreover, although we used English language review data in this study, the proposed method can also be applied to other languages, and a multilingual sentiment-based lexicon needs to be developed in the future.

Acknowledgments

This work was supported by the US National Science Foundation's Computer and Network Systems (CNS) Program, under grant number CNS-0709338, between September 2007 and August 2010.

References

1. Rubicon Consulting, "Online Communities and Their Impact on Business: Ignore at Your Peril," 25 Mar. 2009; http://rubiconconsulting.com/downloads/whitepapers/Rubicon-web-community.pdf.
2. A. Abbasi et al., "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," *ACM Trans. Information Systems*, vol. 26, no. 3, 2008, pp. 1–34.
3. B. Liu, *Web Data Mining*, Springer, 2007.
4. B. Pang et al., "Thumbs up? Sentiment Classification Using Machine Learning Techniques," *Proc. Conf. Empirical Methods in Natural Language Processing* (EMNLP), ACL Press, 2002, pp. 79–86.
5. P.D. Turney, "Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews," *Proc. 40th Ann. Meetings of the Assoc. Computational Linguistics*, ACL Press, 2002, pp. 417–424.
6. E. Riloff and J. Wiebe, "Learning Extraction Patterns for Subjective Expressions," *Proc. Conf. Empirical Methods in Natural Language Processing*, ACL Press, 2003, pp. 105–112.
7. P.D. Turney and M.L. Littman, "Measuring Praise and Criticism: Inference of Semantic Orientation from Association," *ACM Trans. Information Systems*, vol. 21, no. 4, 2003, pp. 315–346.
8. L. Zhou and P. Chaovalit, "Ontology-Supported Polarity Mining," *J. Am. Soc. Information Science and Technology* (JASIST), vol. 59, no. 1, 2008, pp. 98–110.
9. J. Wiebe et al., "Learning Subjective Language," *Computational Linguistics*, vol. 30, no. 3, 2004, pp. 277–308.
10. P. Subasic and A. Huettner, "Affect Analysis of Text Using Fuzzy Semantic Typing," *IEEE Trans. Fuzzy Systems*, vol. 9, no. 4, 2001, pp. 483–496.
11. A. Andreevskaia and S. Bergler, "Mining WordNet for Fuzzy Sentiment: Sentiment Tag Extraction from WordNet Glosses," *Proc. 11th Conf. European Chapter of the Assoc. Computational Linguistics*, ACL Press, 2006, pp. 209–216.
12. A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining," *Proc. 5th Conf. Language Resources and Evaluation* (LREC-06), Evaluation and Language Resource Agency, 2006, pp. 417–422.
13. A. Devitt and K. Ahmad, "Sentiment Polarity Identification in Financial News: A Cohesion-based Approach," *Proc. 45th Ann. Meeting Assoc. Computational Linguistics*, ACL Press, 2007, pp. 984–991.
14. K. Denecke, "Using SentiWordNet for Multilingual Sentiment Analysis," *Proc. IEEE 24th Int'l Conf. Data Eng. Workshop* (ICDEW 2008), IEEE Press, 2008, pp. 507–512.
15. A. Fahrni and M. Klenner, "Old Wine or Warm Beer: Target-Specific Sentiment Analysis of Adjectives," *Proc. AISB Symp. Affective Language in Human and Machine*, Soc. Study of Artificial Intelligence and Simulation of Behavior Press, 2008, pp. 60–63.
16. M. Efron, "Cultural Orientations: Classifying Subjective Documents by Cocitation Analysis," *Proc. AAAI Fall Symp. Series on Style and Meaning in*

Language, Art, Music, and Design*, AAAI Press, 2004, pp. 41–48.
17. A. Abbasi and H. Chen, “Applying Authorship Analysis to Extremist-Group Web Forum Messages,” *IEEE Intelligent Systems*, vol. 20, no. 5, 2005, pp. 67–75.
18. V. Hatzivassiloglou and J. Wiebe, “Effects of Adjective Orientation and Gradability on Sentence Subjectivity,” *Proc. 18th Int’l Conf. Computational Linguistics*, ACL Press, 2000, pp. 299–305.
19. M. Hu and B. Liu, “Mining and Summarizing Customer Reviews,” *Proc. ACM SIGKDD Int’l Conf.*, ACM Press, 2004, pp. 168–177.
20. M. Dash and H. Liu, “Feature Selection for Classification,” *Intelligent Data Analysis: An Int’l J.*, vol. 1, no. 3, 1997, pp. 131–156.
21. R. Maynard, “Online Reviews More Important than a Recommendation from a Friend,” 2 July 2008; http://newsroom.pricerunner.co.uk/72/online-reviews-more-important-than-a-recommendation-from-a-friend.
22. J. Blitzer et al., “Biographies, Bollywood, Boom-boxes and Blenders: Domain Adaptation for Sentiment Classification,” *Proc. Assoc. Computational Linguistics*, ACL Press, 2007, pp. 440–447.
23. F. Benamara et al., “Sentiment Analysis: Adjectives and Adverbs are Better Than Adjectives Alone,” *Proc. Int’l Conf. Weblogs and Social Media* (ICWSM), AAAI Press, 2007, pp. 203–206.
24. R. Zheng et al., “A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques,” *J. Am. Soc. Information Science and Technology* (JASIST), vol. 57, no. 3, 2006, pp. 378–393.
25. B. Guo and M.S. Nixon, “Gait Feature Subset Selection by Mutual Information,” *IEEE Trans. Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 39, no. 1, 2009, pp. 36–46.
26. M. Koppel and J. Schler, “Exploiting Stylistic Idiosyncrasies for Authorship Attribution,” *Proc. IJCAI Workshop Computational Approaches to Style Analysis and Synthesis*, IJCAI, 2003, pp. 69–72.
27. S. Li and C. Zong, “Multi-domain Sentiment Classification,” *Proc. Assoc. Computational Linguistics*, ACL Press, 2008, pp. 257–260.

cn *Selected CS articles and columns are also available for free at http://ComputingNow.computer.org.*