Editor: **Hsinchun Chen,** University of Arizona, hchen@eller.arizona.edu

AI and Opinion Mining

Hsinchun Chen and David Zimbra, *University of Arizona*

The advent of Web 2.0 and social media content has stirred much excitement and created abundant opportunities for understanding the opinions of the general public and consumers toward social events, political movements, company strategies, marketing campaigns, and product preferences. Many new and exciting social, geopolitical, and business-related research questions can be answered by analyzing the thousands, even millions, of comments and responses expressed in various blogs (such as the blogosphere), forums (such as Yahoo Forums), social media and social network sites (including YouTube, Facebook, and Flikr), virtual worlds (such as Second Life), and tweets (Twitter). *Opinion mining*, a subdiscipline within data mining and computational linguistics, refers to the computational techniques for extracting, classifying, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated content. Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and other emotional states in online text. For example, we might seek to answer these such questions:

- What were the opinions of young US voters toward the Democratic and Republican presidential candidates during the most recent election?
- Since September 11, how do the international Jihadi forums introduce radical ideology and incite young members?
- What are the opinions and comments of investors, employees, and activists toward Wal-Mart in light of its cost-reduction efforts and global business practices?
- What was the most successful McDonald's promotional campaign conducted recently in China, and why did it succeed? Which McDonald's product is most preferred by young students in China and why?

Much advanced research in this area has recently focused on several critical areas.[1,2] In this installment of Trends & Controversies and the next, we review several contributions to this emerging field. The topics covered include how to extract opinion, sentiment, affect, and subjectivity expressed in text. For example, resources online might include opinions about a product or the violent and racist statements expressed in political forums. Researchers have also been able to classify text segments based on sentiment, affect, and subjectivity by analyzing positive or negative sentiment expressed in sentences, the degree of violence expressed in forum messages, and so on.

Future Challenges

In spite of recent advances, there are still several promising new directions for developing and advancing new opinion mining research. For example, much past and current opinion mining research has focused on English, Chinese, Arabic, and several European languages. Advanced and mature techniques have been developed especially for English. However, in light of the large amount of public opinions expressed by citizens in different parts of the world, new, scalable opinion mining and sentiment analysis resources and techniques need to be developed for various languages. Future work in multilingual opinion mining will require bootstrapping techniques for analyzing obscure and lesser-known languages for quick situation assessment.

Frameworks and methods for integrating sentiments and opinions expressed with other

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

computational representations—such as interesting topics or product features extracted from user-generated text, participant reply networks, spikes and outbreaks of ideas or events—are also critically needed. Sentiment and opinion alone will not allow researchers to sufficiently understand the complex dynamics in opinions expressed by a large group of participants.

Much of the current opinion-mining research has focused on business and e-commerce applications, such as product reviews and movie ratings. Little research has tried to understand opinions in the social and geopolitical context. For example, what are the opinions of the different Muslim (Shia and Sunni) populations toward the US rebuilding effort in Iraq? What are the potential risks associated with the recent US policy toward Somalia and other Maghreb countries in North Africa? Many researchers believe that the *geopolitical Web*—that is, the geopolitical content and opinions expressed in the Web-based social media contributed by the local population or constituents in a given country or region—can help shed light on public opinions and concerns. Such information might help advance "soft power" or "smart power" in an at-risk region.[3,4] There are significant interdisciplinary research opportunities between computer scientists and social and political researchers in this area.

As we look forward, many emerging social media or Web 2.0 applications create new challenges and opportunities for opinion mining. For example, how do we properly capture the burst and fading of interests expressed in the short and often cryptic 140-character-long tweets? How do we capture the opinion-related body language expressed by Second Life participants (such as thumbs-up, thumbs-down, and applause) for opinion mining?

Lastly, in addition to extracting and quantifying online opinions, we need new research to identify the causal and association relationships between online opinions and real-world events or performances. Will the online opinions expressed in the activist forums help change and shape future political events or policies? How effective are the international Jihadi forums in radicalizing members and increasing new terrorist recruits? Are the volume and sentiment of online movie word-of-mouth comments good predictors of eventual movie sales? Will online employee or investor comments and sentiment help predict a corporation's future stock performance?

Opinion Mining for Wal-Mart

As an example of how online opinion-mining works, we present our research that sought to understand the stock performance of a large US corporation, Wal-Mart. This research, which is based on a Market Intelligence 2.0 (MI2) analysis framework, applies automatic topic and sentiment extraction methods to various online discussions to assess the opinions of various business constituents toward a given company.[5]

Our work builds on previous studies focusing on the relationship between the discussions held in firm-specific finance Web forums and public stock behavior. However, instead of assuming a shareholder view of participants in a finance Web forum as in previous research, and considering them to be uniformly representative of investors, we adopted a stakeholder perspective. This perspective more accurately represents the diversity of the constituency groups participating in the Web forum and closely aligns the analysis with the corporation's stakeholder theory.

To address the broad questions posed in this research, and guided by the literature reviewed, we developed a framework for analysis with four major stages: stakeholder analysis, topical analysis, sentiment analysis, and stock modeling. During the stakeholder analysis stage, we identified the stakeholder groups participating in Web forum discussions. In the topical analysis stage, the major topics of discussion driving communication in the Web forum are determined. The sentiment analysis stage consists of assessing the opinions expressed by the Web forum participants in their discussions. Finally, in the stock modeling stage, we examine the relationships between various attributes of Web forum discussions and the firm's stock behavior.

Our study collected messages from the Yahoo Finance Wal-Mart forum from 4 January 1999 to 10 July 2008, resulting in 433,325 messages. We also used Yahoo Finance to collect stock information to calculate the stock behavior variables for the 2,394 trading days covered in the analysis. News articles containing Wal-Mart related keywords were also collected from the *Wall Street Journal* over the same time period, for a total of 4,077 articles.

The stock models we developed to test the hypotheses regarding the relationship between the Web forum discussions and the Wal-Mart stock behavior directly followed those proposed in previous research.[6] The models examined the correlation and developed both contemporaneous and predictive regression models using the variables presented in the research design. In the contemporaneous regressions, the stock behavior variables are regressed on the measures of the forum discussions occurring on the same day. In the predictive regressions, measures of the forum

Table 1. Predictive regression models for the overall forum.

Overall forum$_t$	Market$_t$	Sentiment$_{t-1}$	Disagreement$_{t-1}$	Message volume$_{t-1}$	Message length$_{t-1}$	Subjectivity$_{t-1}$
Return$_t$	0.8723† (31.33)	0.0025 (0.31)	0.0000 (0.04)	−0.0007** (−2.29)	0.0002 (1.42)	0.0015 (1.46)
Volatility$_t$	−0.0010 (−0.25)	0.0074 (0.47)	−0.0023† (−4.94)	−0.0122† (−19.09)	0.0030† (7.82)	0.0149† (7.27)
Trading volume$_t$	0.7627† (15.06)	−0.4275** (−2.06)	0.0140** (2.29)	0.1957† (−13.24)	−0.0668† (−13.24)	−0.3014† (−11.11)

** $p < 0.05$, †$p < 0.01$

discussions occurring on a specific day are utilized to predict the stock behavior on the following trading day. We summarize selected results from predictive regression using overall forum data.

As Table 1 shows, message volume in the forum holds a significant negative relationship with stock return, with high volume indicating subsequent negative returns. Disagreement and subjectivity also held significant relationships with volatility, where less disagreement and high levels of subjectivity predicted periods of high stock volatility. When the forum consolidated or intensified its sentiment position and utilized highly subjective language, perhaps in efforts to influence other participants, increases in stock volatility were likely to follow in the subsequent trading day. Disagreement also predicted trading volume on the next day, characterized by a positive relationship, as buyer and seller investor positions are communicated in the discussions followed by high trading volume. High levels of subjectivity, on the other hand, indicated suppressed trading volume to follow. Efforts to influence other forum participants might cause the market to become more conservative, maintaining their stock holdings or postponing a purchase. Additionally, sentiment expressed in the Web forum holds a significant relationship with the trading volume on the following day. Positive sentiment reduces trading volume, perhaps because satisfied shareholders hold their stock, while negative sentiment induces trading activity as shareholders defect.

In This Issue and the Next

In this Trends & Controversies department and the next, we include three articles on opinion mining from distinguished experts in computer science and information systems. Each article presents a unique innovative research framework, computational methods, and selected results and examples. In this first issue, Bing Liu's article "Sentiment Analysis: A Multifaceted Problem" argues that sentiment analysis is not a single problem, but a combination of many facets or subproblems. Liu introduces some of these problems and suggests several technical challenges, including object identification, feature extraction and synonym grouping, opinion-orientation classification, and integration.

The July/August 2010 issue of *IEEE Intelligent Systems* will follow up with two additional articles: "Sentiment Quantification," by Andrea Esuli and Fabrizio Sebastiani, and "Intelligent Feature Selection for Opinion Classification," by Ahmed Abbasi.

Acknowledgments

The research is supported in part by the following grants: US Department of Defense HDTRA-09-0058, US National Science Foundation CNS-070933, NSF CBET-0730908, and NSF IIS-0428241, awarded to the University of Arizona Artificial Intelligence Lab (PI, Hsinchun Chen).

References

1. A. Abbasi et al., "Affect Analysis of Web Forums and Blogs using Correlation Ensembles," *IEEE Trans. Knowledge and Data Eng.*, vol. 20, no. 9, 2008, pp. 1168–1180.
2. A. Esuli and F. Sebastiani, "SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining," *Proc. 5th Conf. Language Resources and Evaluation* (LREC), 2006, pp. 417–422.
3. H. Chen, *Intelligence and Security Informatics for International Security*, Springer, 2006.
4. J. Nye, *Soft Power: The Means to Success in World Politics*, PublicAffairs, 2004.
5. H. Chen, "Business and Market Intelligence 2.0," *IEEE Intelligent Systems*, vol. 25, no. 1, 2010, pp. 68–71.
6. W. Antweiler and M. Frank, "Is All that Talk Just Noise? The Information Content of Internet Stock Message Boards," *J. Finance*, vol. 59, no. 3, 2004, pp. 1259–1295.

Hsinchun Chen is the McClelland Professor of Management Information Systems at the University of Arizona and director of the Artificial Intelligence Lab. Contact him at hchen@eller.arizona.edu.

David Zimbra is research associate in the Management Information Systems Department at the University of Arizona. Contact him at zimbra@email.arizona.edu.

Sentiment Analysis: A Multifaceted Problem

Bing Liu, *University of Illinois at Chicago*

Sentiment analysis is the computational study of people's opinions, appraisals, and emotions toward entities, events and their attributes. In the past few years, this field has attracted a great deal of attention from both the academia and industry due to many challenging research problems and a range of applications.

Opinions are important because whenever people need to make a decision, they want to hear others' opinions. The same is true for organizations. However, few computational studies on opinions existed before the Web because there was little opinionated text (text with opinions or sentiments) available. In the past, when making a decision, individuals typically asked for opinions from friends and families. When an organization wanted to find opinions of the general public about its products and services, it conducted surveys and focus groups. However, with the explosive growth of the social media content on the Web in the past few years, the world has been transformed. People can now post reviews of products at merchant sites and express their views on almost anything in discussion forums and blogs, and at social network sites. Hence, individuals are no longer limited to asking friends and families because of the plethora of user-generated product reviews and opinions available on the Web. In turn, companies might no longer need to conduct surveys or focus groups to gather consumer opinions about its products and those of its competitors because there's plenty of such information publicly available.

However, finding opinion sites and monitoring them on the Web is a formidable task because there are numerous, diverse sources, each of which might also have a huge volume of opinionated text. In many cases, opinions are hidden in long forum posts and blogs, so it is difficult for a human reader to find relevant sites, extract related sentences with opinions, read them, summarize them, and organize them into usable formats. Automated opinion discovery and summarization systems can address this need.

In this article, I first give a brief introduction to this problem and present some technical challenges. As we will see, sentiment analysis is not a single problem, but a combination of many facets or subproblems. This article introduces and explains some of these problems. I will then share some of my thoughts on the past and future of sentiment analysis based on my research in the past few years and my experience in the industry for a short while.

Sentiment-Analysis Problem

The research in the field started with sentiment and subjectivity classification, which treated the problem as a text classification problem. *Sentiment classification* classifies whether an opinionated document (such as product reviews) or sentence expresses a positive or negative opinion.[2] *Subjectivity classification* determines whether a sentence is subjective or objective.[3] Many real-life applications, however, require more detailed analysis because users often want to know the subject of opinions.[1,4] For example, from a product review, users want to know which product features consumers have praised and criticized.

To explore this generic problem, let's use the following review segment on iPhone as an example:

> (1) I bought an iPhone two days ago. (2) It was such a nice phone. (3) The touch screen was really cool. (4) The voice quality was clear too. (5) However, my mother was mad with me as I did not tell her before I bought it. (6) She also thought the phone was too expensive, and wanted me to return it to the shop....

(I numbered each sentence for easy reference.[1]) The question is, what do we want to extract from this review? The first thing that we might notice is that there are several opinions in this review. Sentences 2, 3, and 4 express three positive opinions, while sentences 5 and 6 express negative opinions or emotions. We can also see that all the opinions are expressed about some targets or objects. For example, the opinion in sentence 2 is on the iPhone as a whole, and the opinions in sentences 3 and 4 are on the iPhone's touch screen and voice quality, respectively.

Importantly, the opinion in sentence 6 is on the iPhone's price, but the opinion/emotion in sentence 5 is about "me," not the phone. In an application, the user might be interested in opinions on certain targets but not necessarily on user-specific information. Finally, we can also see the sources or holders of opinions. The source or holder of the opinions in sentences 2, 3, and 4 is the author of the review, but in sentences 5 and 6, it is "my mother."

Definitions

With this example in mind, we can now define the sentiment analysis or opinion mining problem. We start with the *opinion target*. In general, people can express opinions on any *target entity*—products, services, individuals, organizations, or events. In this context, the term object is used to denote the target entity that has been commented on. An object can have a set of components (or parts) and a set of attributes (or properties),[1,4] which we collectively call the features of the object.

For example, a particular brand of cellular phone is an object. It has a set of components (such as battery and screen) and a set of attributes (such as voice quality and size), which are all called *features* (or *aspects*). An opinion can be expressed on any feature of the object and also on the object itself.

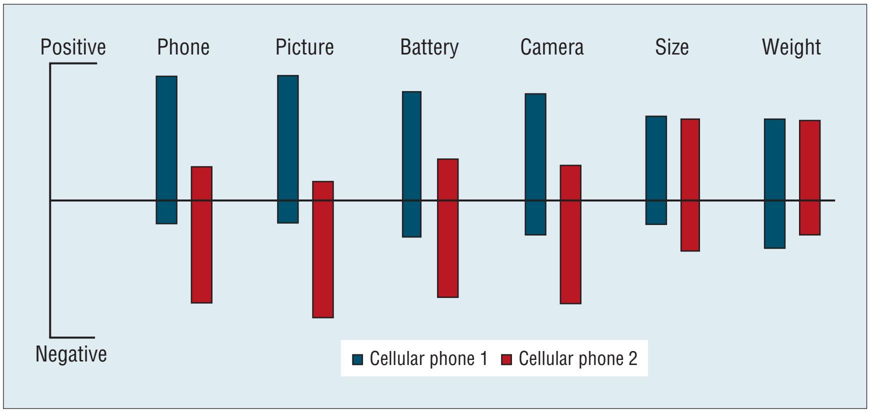

Figure 1. Visual comparison of feature-based opinion summaries of two cellular phones. Each bar above the x-axis shows the number of positive opinions on a feature (given at the top), and the bar below shows the number of negative opinions on the same feature.

For example, in "I like the iPhone. It has a great touch screen," the first sentence expresses a positive opinion on the iPhone itself, and the second sentence expresses a positive opinion on its touch screen feature.

The *opinion holder* is the person or organization that expresses the opinion. In the case of product reviews and blogs, opinion holders are usually the authors of the posts. Opinion holders are more important in news articles because they often explicitly state the person or organization that holds a particular opinion.

An *opinion* on a feature f (or object o) is a positive or negative view or appraisal on f (or o) from an opinion holder. Positive and negative are called opinion orientations.

With these concepts in mind, we can define an object model, a model of an opinionated text, and the mining objective, which are collectively called the *feature-based sentiment analysis model*.[1,4] In the *object model*, an object o is represented with a finite set of features, $F = \{f_1, f_2, \ldots, f_n\}$, which includes the object itself as a special feature. Each feature $f_i \in F$ can be expressed with any one of a finite set of words or phrases $W_i = \{w_{i1}, w_{i2}, \ldots, w_{im}\}$, which are the feature's *synonyms*.

In the *opinionated document model*, an opinionated document d contains opinions on a set of objects $\{o_1, o_2, \ldots, o_r\}$ from a set of opinion holders $\{h_1, h_2, \ldots, h_p\}$. The opinions on each object o_j are expressed on a subset F_j of features of o_j. An opinion can be one of the following two types:

- A *direct opinion* is a quintuple $(o_j, f_{jk}, oo_{ijkl}, h_i, t_l)$, where o_j is an object, f_{jk} is a feature of the object o_j, oo_{ijkl} is the orientation of the opinion on feature f_{jk} of object o_j, h_i is the opinion holder, and t_l is the time when the opinion is expressed by h_i. The opinion orientation oo_{ijkl} can be positive, negative, or neutral.
- A *comparative opinion* expresses a preference relation of two or more objects based their shared features. A comparative opinion is usually conveyed using the comparative or superlative form of an adjective or adverb, such as "Coke tastes better than Pepsi." (Due to space limitations, this article does not discuss such opinions; see previous work for more details.[1])

Therefore, given an opinionated document d, the objective of sentiment analysis (or opinion mining) is twofold:

- discover all opinion quintuples $(o_j, f_{jk}, oo_{ijkl}, h_i, t_l)$ in d and
- identify all synonyms (W_{jk}) of each feature f_{jk} in d.

In practice, not all five pieces of information in the quintuple need to be discovered for every application because some of them might be known or not needed. For example, in the context of online forums, the site typically displays the time when a post is submitted and identifies the opinion holder. (We will not discuss them further in this article.)

Applications

A simple way to use the results is to produce a feature-based summary of opinions on an object or multiple competing objects.[1,4] Figure 1 shows the summary of opinions on two competing cellular phones along different product features dimensions. We can clearly see how consumers view the different features of each product. Phone 1 is clearly a better product.

Technical Challenges

The objective of sentiment analysis gives us a good clue of the main tasks involved and technical challenges. The following blog excerpt gives a more complex example:

> (1) Yesterday, I bought a Nokia phone and my girlfriend bought a moto phone. (2) We called each other when we got home. (3) The voice on my phone was not clear. (4) The camera was good. (5) My girlfriend said the sound of her phone was clear. (6) I wanted a phone with good voice quality. (7) So I was not satisfied and returned the phone to Best Buy.

Object Identification

The objects discussed in the blog are "moto" (Motorola) and Nokia. This object identification is important because without knowing the object on which an opinion has been expressed, the opinion is of little use. The problem is similar to the classic named-entity recognition problem, but there is a difference. In a typical opinion-mining application, the user wants to find opinions on some competing objects, such as competing products or services. Thus, the system needs to separate relevant and irrelevant objects. For example, Best Buy is not a competing product name, but a store name.

Feature Extraction and Synonym Grouping

In this example, the phone features are voice, sound, and camera. Although there have been attempts to solve this problem, it remains to be a major challenge. Current research mainly finds nouns and noun phrases. The recall might be good, but the precision can be low. Furthermore, verb features are common as well but are harder to identify.

To produce a summary similar to the one in Figure 1, we also need to group synonym features because people often use different words or phrases to describe the same feature. For example, voice and sound refer to the same feature in this excerpt.

Opinion-Orientation Determination

The next task is to determine whether a sentence contains an opinion on a feature and, if so, whether it is positive or negative. Existing approaches are based on different supervised and unsupervised methods using opinion words and phrases and the grammar information. One key issue is to identify opinion words and phrases (such as good, bad, poor, or great), which are instrumental to sentiment analysis. However, there are seemly an unlimited number of expressions that people use to express opinions, and in different domains, they can be significantly different. Even in the same domain, the same word might indicate different opinions in different contexts.[1]

For example, in the sentence "The battery life is long," long indicates a positive opinion about the battery life feature. However, in the sentence "This camera takes a long time to focus," long indicates a negative opinion. Also, sentence 6 in the blog excerpt seemingly expresses a positive opinion, but of course it does not. There are still many problems that need to be solved.[1]

Integration

Integrating these tasks is also complicated because we need to match the five pieces of information in the quintuple. That is, the opinion oo_{ijkl} must be given by opinion holder h_i on feature f_{jk} of object o_j at time t_l. To make matters worse, a sentence might not explicitly mention some pieces of information, but they are implied using pronouns, language conventions, and context.

To deal with these problems, we need to apply natural-language processing techniques in the opinion mining context, such as parsing, word-sense disambiguation, and coreference resolution.

As an example, we can use coreference resolution to give a glimpse of the issues. For the blog example, figuring out what is "my phone" and what is "her phone" in sentences 3 and 5 is not a simple task. Sentence 4 does not mention any phone and does not have a pronoun; the question is which phone "the camera" belongs to. Coreference resolution is a classic problem in natural-language processing, and the research community has not yet found an accurate solution.

My Perspective

I would now like to share some of my thoughts on the past and future of the field based on my research and practical application experiences.

The Past

The research community has studied almost all main aspects of this problem. The most well-studied subproblem is opinion orientation classification—that is, at the document, sentence, and feature levels. The existing reported solutions are still far from perfect because current studies are still coarse and not much has been done on finer details. For example, on opinion classification, there are many conceptual rules that govern opinions,[1] and there are even more expressions (possibly unlimited) that can convey these concepts. However, little in-depth study has been done on many of them. On feature extraction and synonym grouping, they remain challenging. Object extraction is probably the easiest because we can apply many existing information extraction algorithms. Integration and matching of all five pieces of information in the quintuple is still lacking, which is probably not surprising since the research community likes to focus on individual subproblems.

This leads us to the question of sentiment analysis accuracy of the current algorithms. The question is, however, very difficult to answer because there are so many subproblems. Although for some individual problems, researchers have annotated data for testing, a comprehensive corpus still does not exist to help test the accuracy on all tasks in a unified way. For accurate evaluation, the benchmark data needs to cover a large number of domains because a system that does well in one domain might not do well in another, as opinions in different domains can be expressed so differently. Precision and recall are commonly used as evaluation measures. In most applications, high precision is critical but high recall might not be necessary as long as the system can extract enough opinions to be statistically significant. A crucial issue, however, is ensuring the correct proportions of positive and negative opinions on each feature. Hence, the system errors should be balanced so that they do not destroy the natural distribution of positive and negative opinions.

The Future

Building on what has been done so far, I believe that we just need to conduct more refined and in-depth investigations as well as build integrated systems that try to deal with all the problems together because their interactions can help solving each individual problem. I am optimistic that the problems will be solved satisfactorily in the next few years for widespread applications. In fact, we might already begin to see the light at the end of the tunnel. For instance, based on our tests using 10 diverse data sets, the system that my group is building (called Opinion Parser) can achieve 80 to 90 percent of recall and precision on feature-based opinion orientation determination. It can also perform integration to a good extent based on several automated discovery functions.

For real-life applications, a completely automated solution is nowhere in sight. However, it is possible to devise effective semiautomated solutions. The key is to fully understand the whole range of issues and pitfalls, cleverly manage them, and determine what portions can be done automatically and what portions need human assistance. In the continuum between the fully manual solution and fully automated solution, we can push more and more toward automation.

Beyond what have been discussed so far, we also need to deal with the issue of opinion spam, which refers to writing fake or bogus reviews that try to deliberately mislead readers or automated systems by giving untruthful positive or negative opinions to promote a target object or damage the reputation of another object.[5] Detecting such spam is vital as we go forward because spam can make sentiment analysis useless.

Finally, despite these difficulties and challenges, the field has made significant progress over the past few years. This is evident from the large number of start-up companies that provide sentiment-analysis and opinion-mining services. A real, substantial need exists in industry for such services. This practical need and the technical challenges will keep the field vibrant and lively for years to come.

References

1. B. Liu, "Sentiment Analysis and Subjectivity," *Handbook of Natural Language Processing*, 2nd ed., N. Indurkhya and F.J. Damerau, eds., Chapman & Hall, 2010, pp. 627–666.
2. B. Pang and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, nos. 1–2, 2008, pp. 1–135.
3. J. Wiebe et al., "Learning Subjective Language," *Computational Linguistics*, vol. 30, Sept. 2004, pp. 277–308.
4. M. Hu and B. Liu, "Mining and Summarizing Customer Reviews," *Proc. ACM SIGKDD Conf. Knowledge Discovery and Data Mining* (KDD), ACM Press, 2004, pp. 168–177.
5. N. Jindal and B. Liu, "Opinion Spam and Analysis," *Proc. Conf. Web Search and Web Data Mining* (WSDM), ACM Press, 2008, pp. 219–230.

Bing Liu is a professor of computer science at the University of Illinois at Chicago. He has a PhD in artificial intelligence from the University of Edinburgh. Contact him at liub@cs.uic.edu.