

Our Sentiments, Exactly

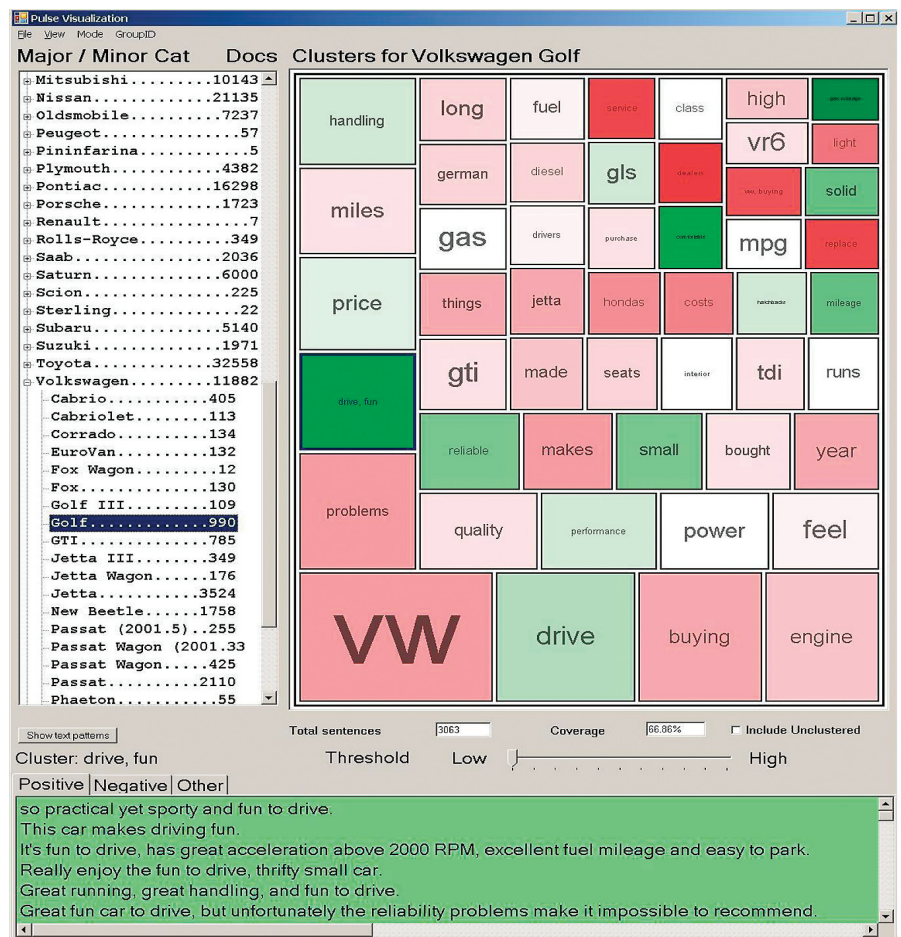
With sentiment analysis algorithms, companies can identify and assess the wide variety of opinions found online and create computational models of human opinion.

FACTS MAY BE stubborn things, as John Adams once put it, but at least they're easy to compute. From a data-processing perspective, opinions are much more stubborn.

In recent years, the Web has created a bull market in human opinion: movie reviews, product ratings, restaurant recommendations, and all kinds of other viewpoints expressed in articles, blogs, discussion groups, and elsewhere. As the Web accumulates more and more data, many of us rely on each other's opinions as a filter to help us make informed decisions. For many businesses, customer opinions have become a type of virtual currency that can make or break their products. As opinion data plays an increasingly important role on the Web, however, computer scientists are discovering the limitations of traditional text analytics algorithms for sorting opinions from raw facts.

The distinction between facts and opinions might seem clear enough on the surface, but in practice teasing them apart involves parsing many linguistic shades of gray. This is where the emerging field known as sentiment analysis comes in. Sometimes called opinion mining or subjectivity analysis, sentiment analysis is a new term that broadly refers to the identification and assessment of opinions, which for the purposes of computation might be defined as written expressions of subjective mental states.

Traditional text analytics algorithms work by scanning a body of text to extract and analyze keywords. That approach works well for identifying simple factual statements, but assessing opinions requires delving much deeper into the subtleties of human language. "Sentiments are very different from conventional facts," says analytics consultant Seth Grimes. While direct expressions of opinion are fairly easy to spot—for example, "I hated *Revenge of*



An example of sentiment analysis from Michael Gamon in which topics from reviews of the Volkswagen Golf are depicted. The size of each topic box indicates the number of mentions of the topic, and the shading of each topic box indicates the average sentiment, ranging from negative (red) to neutral/none (white) to positive (green).

the Sith”—most human sentiments fall somewhere along a continuum from objective fact to subjective experience. For example, “It’s fifteen degrees outside” is an objective statement; “It’s cold” reveals a somewhat more subjective point of view; while “I’m putting on two pairs of socks” constitutes a completely indirect expression of opinion disguised as a statement of fact.

“We are dealing with sentiment that can be expressed in subtle ways,” says Yahoo! researcher Bo Pang, co-author of the book *Opinion Mining and Sentiment Analysis*. To penetrate those sub-

tleties, sentiment analysis algorithms assess written statements through a series of overlapping filters. They usually begin by attempting to determine the polarity of a particular sentiment—i.e., Is it positive or negative? Once that’s established, they may try to determine the intensity of sentiment being expressed—i.e., How positive or negative is this statement? Next, an even more subtle layer of analysis might attempt to determine the degree of subjectivity—i.e., How partial or impartial is the point of view being expressed here? (This is often determined by looking at

the number of adjectives in a sentence.) Using these and other criteria, sentiment analysis algorithms can then begin to create computational models of human opinion.

Complicating matters even further are questions of context (who's speaking, and to whom?) and linguistic nuances like slang and ambiguity. A "bad motorcycle" might actually be a good one; whereas a "bad movie" is probably just plain bad. Sentiment analysis algorithms sometimes have to go beyond literal interpretations of a text to discern an author's original intent. Given the wide varieties of idiomatic writing on the Web, this is no small task. As Grimes notes, "You don't see 'Genistein inhibits protein histidine kinase...Not!' in a scientific paper."

Mining Collective Opinions

"With opinions, so much depends on the point of view of the user," says David Pierce, chief technology officer of Jodange, whose sentiment analysis software grew out of a research project by Claire Cardie at Cornell University and Jan Wiebe at the University of Pittsburgh. Drawing on a body of theory in linguistics, philosophy, and computational linguistics, their team developed an algorithm that tries to determine the context of any particular statement by isolating three key data points: the topic, the opinion holder, and the opinion itself. First, the algorithm employs an entity extraction routine that locates keywords to identify particular topics and opinion holders. Next, it layers that data onto a linguistic analysis of the opinion being expressed. The resulting unit of data is a triple consisting of opinion, opinion holder, and topic. These triples are then stored in a relational database, where they can be cross-referenced across multiple documents to create what Jodange vice president of product management and marketing Pia Chong calls a "walled garden of opinion."

By connecting opinions from multiple sources about a particular topic, the application can provide users with a bird's-eye view of a particular topic presented in a variety of different formats: straightforward lists, heat maps that show the concentration of opinions on particular topics, an opinion index that calculates positive or negative trends, or a so-called Doppler view that shows a

graphical summary of opinion data. The company is currently working on a new predictive model that could use opinion data to predict future developments, such as the impact of written opinion on trends in a company's stock price.

A number of other companies are now developing their own variations of sentiment analysis software. Companies like Attensity, Clarabridge, Lexalytics Limited, SPSS, and TEMIS are developing their own proprietary versions of sentiment analysis software. All of these products employ some combination of keyword extraction and linguistic analysis to provide their customers with a particular understanding of collective opinion. Some of these products are targeted toward business applications, others toward consumer-facing Web applications.

For consumers, the most obvious applications for sentiment analysis involve enhancing search engines with more opinion data. "Sentiment analysis software could enable a much smoother user experience" for consumer research, says Pang. Microsoft's Product Search, which is part of Live Search, and Yelp's review highlights, which include phrases automatically extracted from user reviews, already rely on basic sentiment analysis to enhance their search results. Such interactions could eventually find their way into the general Web search experience. Pang suggests that such interactions could be fine-tuned for users at different stages of the research process, allowing them to narrow down from reviews of a product category to comparisons between products, then finally to in-depth product reviews.

Beyond the realm of consumer prod-

For many businesses, online customer opinions have become a type of virtual currency that can make or break their products.

ucts, Pang also sees opportunities for sentiment analysis to shape the way people consume news. "When the media is having a field day, [users] might want to get a digest of different perspectives on the breaking news with analysis." Similar applications might eventually lead to new types of interfaces where readers could track the movement of opinion about particular stories over time.

That kind of opinion-trending insight is particularly valuable to users working in business or government. These potential users might include business intelligence professionals, market researchers, or public relations specialists. Today, sentiment analysis vendors are already marketing their products to companies in the form of hosted services that provide opinion dashboards and other management tools. At this stage, sentiment analysis software is too new to have penetrated most IT firewalls. Eventually, however, companies may start exploring how to integrate sentiment analysis data with their core management systems. "Unified analysis is coming," says Grimes, "but it's not here yet."

Attensity is taking a step in that direction by marketing a suite of tools designed to help companies integrate sentiment analysis data with their internal business operations. In addition to providing sentiment analysis data, Attensity provides mechanisms for funneling that data into operational "queues" like marketing campaigns or call center scripts. "For example, if a valuable customer is upset they can route them to a special marketing campaign that compensates them through points or other things of value," explains Michelle de Haaff, Attensity's vice president of marketing and products.

As sentiment analysis finds its way into the business mainstream, vendors will likely continue to develop similar services that bring sentiment analysis into the IT mainstream. Once that integration starts to happen, companies will be able to feed opinion data into core business processes that can help them strengthen their customer relationships—and, ultimately, boost profits: a decidedly unsentimental goal. ■

Alex Wright is a writer and information architect who lives and works in New York City.

© 2009 ACM 0001-0782/09/0400 \$5.00