crimson hexagon

Nine Essentials of Online Opinion Monitoring

Measuring Brand Equity and Impact in the Age of Social Media

Introduction

In days gone by, marketing was a polished monologue. You orchestrated the release of your brand value proposition to the world and hoped for the best. You were able to measure the relative success of your efforts only later in the cycle—through survey results, focus groups, a spike in market share for you—or a competitor.

With the rise of self-publishing on the internet, that monologue has become more of a multilogue: a many-to-many conversation. Your customers, supporters, and detractors talk amongst themselves and declare their views on you to anyone interested. The question is: are you listening to the right opinions? Are you responding?

You may feel that this idle chatter originates from squeaky wheels with too much time on their hands. But whether the opinions are expressed by perpetual malcontents or considered influencers does not change the fact that your public is *reading* these views written by the people who experience your brand. A full 41 percent of bloggers frequently post about brands they love or hate. Their views are shaping the views of their readers. [1]

With more and more words out there, following online opinion has become almost an expectation—for large and small companies, local and national public figures alike. According to one *Wall Street Journal* writer, "(s)ince people can easily share their opinions, they expect their voices will be heard."[2]

While listening to what people are saying about your products, services or policy stance is both a nice-to-know and an expectation, it should in fact be seen as much more. Online opinions are a deeply valuable resource that organizations can extract in order to:

- receive immediate feedback
- spot opportunities
- identify fatal flaws
- measure marketing effectiveness
- shape a promising direction for your corporate or political vision, product features, or service strategy that takes root in what real people in the real world are saying.

Your brand, as defined by marketing or campaign managers, can help shape how your customers or audience feel about you. But your brand as the public perceives it is a reflection of everything you do—your products, your customer service, your public events. Social media is now uniting all of these slivers of your publicly perceived brand to form a *de facto* brand. Even when you are scrupulously on-brand in all departments, your de facto brand will be defined "out there."

In this paper, we introduce the essential components of online opinion monitoring and illustrate the benefits of each aspect. As we hear from analysts and customers, understanding the conversation helps organizations gain more insight into, and perhaps some measure of control over, how the public perceives their brands. While you may not be driving the bus, at least you can get on and listen in on the conversations.

1 "State of the Blogosphere 2008," technorati.com

2 "So Many, Many Words," *Wall Street Journal*, 28 January 2008

1. Go big—go online

> *"The blogosphere is now so large it is an accurate barometer of consumer opinion (…) All brands should be using it as a means to measure consumer opinion, track response to marketing initiatives and as a forum for research."* [3]

It is the age of big data. As the amount of information available to us spills over into petabytes, some claim that even the discipline of science is changing. The hypothesis is being replaced by the open-ended question, and surprising answers are popping out of the matrix of near-infinite test data.

Forests of servers and clouds of processors store and circulate photos, video, scientific data, songs, images and, of course, words: words written by anyone who has access to a computer and the internet.

Although the numbers are often disputed, one estimate puts the number of bloggers worldwide at 184 million and the number of readers at 346 million.[4] Of these bloggers, 41 percent are frequently moved to share their opinion on products and brands.[5] Because of blogs and other user-generated social media such as consumer communities, forums, discussion groups, and product reviews, larger and larger quantities of text are now available for analysis: a sprawling but fathomable archive.

The pioneers are already out there, gauging public opinion on political topics or figures and driving and refining marketing and product development. "Now Google and like-minded companies are sifting through the most measured age in history, treating this massive corpus as a laboratory of the human condition." [6]

All of this freely available data is sounding a warning for traditional research methods, including surveys. A spokesperson at P&G ventured that without adopting "approaches that are more in touch with the lifestyles of the consumers we seek to understand, the consumer-research industry as we know it today will be on life support by 2012."[7]

For the time being, many organizations are complementing online opinion monitoring with surveys, a database of customer feedback or possibly even an online community. But accurate insight can only come when you combine your organizational data with what is being written outside of your walls. To get all of this information to tell a story, especially if you are a well-known brand or public figure, you will need a filter—a smart one.

3 "Wave 3," Universal McCann, March 2008

4 "State of the Blogosphere 2008," technorati.com

5 "Wave 3," Universal McCann, March 2008

6 "The End of Theory," The Petabyte Age, *Wired*, 23 June 2008

7 "The End of Consumer Surveys?" *Advertising Age*, 15 September 2008

2. Filter out the noise

Scientists maintain that useful conclusions can only be reached after throwing away large amounts of thoroughly irrelevant or uninteresting data.

Take as an example the Large Hadron Collider, the world's biggest particle accelerator located on the Swiss-French border. Once particles reach top speed on the 17-mile track, cameras will record their collisions. But the resulting 10 petabytes per second is clearly too much data to store and process, so "filters will reduce the take to roughly 100 events per second that seem most promising for analysis."[8]

"Most promising for analysis." There's the rub. How to decide what is interesting and what isn't? In opinion monitoring, the data is words—unstructured data par excellence. To tag and sort a mountain of words, devoid of context, and prepare it for query costs processor time and energy. And it can be difficult to get meaningful results. For example tracking the term "bailout" might provide illustrative opinion trends about certain Wall Street events, but you would need to ensure you were capturing opinions about the right event, firms, etc.

When the mountain is actually just a gentle rise—when you are a small start-up or a local agency getting ten daily occurrences of your name in social media—no jettison is required. You can easily read the context around every mention over your morning coffee.

But for bigger names, your online sentiment corpus will likely be unwieldy. Winnowing out the irrelevant references to your brand and summarizing the true expressions of opinion is not something you'll be able to do by hand.

So how is this done? First of all, in some instances up to 60 percent of documents meeting search term criteria can be non-relevant to the question. These can include technical specifications about your product, job postings for your company, or spam. Opinion monitoring approaches built strictly on counting mentions will fail to exclude such documents. But including a preponderance of irrelevant data in the overall results can minimize and obscure important trends, warnings or opportunities. More advanced approaches automatically filter out these great swaths of irrelevant data and prevent it from clouding your insight.

Irrelevant data areas blown out, your opinion monitoring software can go to work mining the remaining (probably still ample) data and distilling the important insights. This is the second level of filtering that must sort out positive, negative and neutral brand references, and this is where it gets tricky. Opinion monitoring approaches that do not fall prey to positive words used in a negative context (or vice versa—see section 3) help you confidently gauge opinion. Allowing human input in order to "teach" the software what to look for (see section 8) refines the sort and allows interpretation of many levels of language.

Opinion monitoring with effective filtering gives you more precise readings. Only then can you see whether a negative opinion on your topics of interest is widespread, an anomaly, or a canary in a coalmine.

8 "Chasing the Quark," The Petabyte Age, *Wired*, 23 June 2008

3. Stop counting, start learning

> *"The entire literary canon may be smaller than what comes out of particle accelerators or models of the human brain, but the meaning coded into words can't be measured in bytes. It's deeply compressed. Twelve words from Voltaire can hold a lifetime of experience."*[9]

Speaking on the frontiers of big data for textual analysis, IBM mathematician and computer scientist Martin Wattenberg distils the holy grail of natural language science: Meaning. While numbers – and certainly statistics – can be misleading, there's at least the assumption that numerical values are universally understood. Human language is rife with meaning, but also plagued by ambiguity – and ambiguity can lead to misunderstanding.

Many opinion monitoring approaches simply count the raw mentions of a word in the hope that the massive sample size will drown any anomalies arising from the knottiness of language. The BBC television series "White" illustrates the difficulties involved in word counting. To determine general attitudes on the white working class in Britain, research software sorted commenters' words into categories such as "anger," "confusion" or "happiness." Words such as "good" and "dream" fell into the "happiness" bucket, even when they resided in phrases such as "good-for-nothing" and "things I can only dream about." Similarly, counting keyword mentions of brands that may resonate with a similar audience can prove challenging: the singer Pink and the Victoria's Secret lingerie line Pink and the breast cancer Pink might all occupy the same online community discussion.

Mere counting without an analysis of context does not give an accurate or comprehensive sense of opinion. At worst, the insight can be outright wrong.

4. Ask the right questions

How can an organization *act* on the knowledge of how the public feels about them in general? Most brand analyses start here, but what can you really do with this information? If you adjust your position on mass transit spending, will your polls improve? Would they buy more from you if your service was faster? Or is it a product problem? People might be excited about iPhone's internet capabilities but may hate its contact directory. Mixing together opinion on the two categories clouds your sight.

Finding out positive or negative brand opinion is simply not detailed enough. You gain a more sophisticated analysis of public sentiment through opinion monitoring that lets you define searches on specific categories, such as:

- Your brand attributes that are most important to the public
  (do they like that you're reliable, innovative, established, flexible or easy-to-use?)
- What motivates customers to buy
- Features most and least liked in your product
- Problems people are having with your product or service
- The most pressing customer service issues in an online community devoted to your industry

More specific analysis mines deeper insight into public opinion on your chosen topics. It lets you measure against how you are marketing and shape your direction or policy upon it.

9 "Visualizing Big Data," The Petabyte Age, *Wired*, 23 June 2008

5. Grow more specific over time

So you're Apple, and you're asking specific questions. For example "What features do my customers like best about the iPhone?" You're dividing their opinions into seven buckets corresponding to seven features, including email, camera and browser features.

But as the analysis proceeds, you notice that the eighth category—"other"—is growing. What other features have you not thought to ask about? Obviously these features are important to iPhone users, because they are bothering to mention them. Delving in to find the main topics within the "other" category, you may unearth new categories, possibly of even greater importance than your existing categories. You may find that many customers care deeply about the phone's thin profile, cool image or easy-to-use design.

Opinion monitoring helps you zoom in on the specifics that are important to your audience while sifting out unimportant detail.

6. Listen to the people who matter

Somewhere along the line in statistical work, the question of size and quality of sample arises. To gauge opinion, many feel that a broad, man-on-the-street sample is required, so that the response spectrum is wider, in this case towards the pole of the completely uninitiated. It is often very likely this extremely broad audience has never heard of the product, topic or person in question, and that is important information.

But the people who choose to express their brand opinions online are all among the initiated. They are a subset of the whole of possible opinion: not only are they aware of the topic, they are motivated enough to share their opinion in the public sphere. So online opinion monitoring necessarily omits the uninitiated.

Some would argue that gauging opinion on the basis of this sample subset provides only limited insight. Others feel that including the opinions of people who know nothing about the subject actually skews the results.

This question can be answered by the tectonic shift in sample size taking place online. As the barriers to online publishing fall by the wayside, the man is getting off the street and onto the internet to express himself on a range of topics. Today, eMarketer estimates that in the US alone, 25.2 million people publish a blog, and that 104.1 million people are regular blog readers.[10] The rise of social networking sites with a lower "cost of entry" than maintaining a blog – from niche topics to mainstream adoption – also feeds into the number of online opinion. [11]

To help you sort through the range of sources, there is also the option of filtering opinion by authority. In some instances, you may want to hear from the largest possible online sample to make sure you're uncovering the broad opinion trends – letting the data tell the story. In others, you may want to "weight" results as needed, and to discern whether certain sentiments are emerging only from the influencer community or more broadly.

[10] "The Blogosphere: A Mass Movement from Grassroots," eMarketer, May 2008

11 Even newspapers are getting in on the game: the *New York Times* launched a social network in the fall of 2008.

7. Watch the trend, spot the surprise

> *"(O)nce you know how people feel about something, you're a step closer to being able to guess what they'll do next."*[12]

Filtering on large amounts of data for targeted topics can present the current state of discussions, and over time show you a trend. Watching the direction of the trend and extrapolating can help you take a stab at predicting where things are going.

As we've learned from books such as *Crossing the Chasm* and *The Tipping Point*, opinions move in predictable patterns or waves. While we may be pleased to see positive opinion growing and gaining ground, we also need to keep an eye on the outliers. Are they voices in the dark, alone and without support? Or are they precursors, gathering troops outside the kingdom walls: the shape of things to come?

Following trends over time for a large sample can help you spot these surprises before they surprise you.

8. Let people do the thinking, let machines do the work

> *"(W)ithin the petabytes of information it collects may lurk things nobody has even imagined—assuming astronomers can figure out how to teach their computers to look for objects no one has ever seen."*[13]
> -- On the world's largest telescope on Cerro Pachón in Chile

Can machines get a true reading of crowd sentiment? How can an algorithm plumb the depth and range of all possible human expression? Like machine translation, can sentiment mining be accurate without access to true intelligence and world knowledge?

In fact, machine translation is improving because of the availability of vast searchable corpora in many languages and the help of human "teachers" who correct mistakes and help machines "learn."

More advanced online opinion monitoring borrows from these principles: machines and humans helping each other. Human readers prime the software by choosing an indicative range of text as examples of negative and positive opinions within each category of interest. This indicative text can include slang and grammatical errors, thereby teaching the software to identify opinion expressed through many language variations.

The algorithm then extrapolates this base knowledge to thousands or millions of documents that would be impossible for even an army of human readers to scan.

Advanced online opinion monitoring proposes a modern division of labor: humans provide the up-front intelligence, while machines take over for the high-volume execution.

12 "Tracking the News," The Petabyte Age, *Wired*, 23 June 2008

13 "Watching the Skies," The Petabyte Age, *Wired*, 23 June 2008

9. Skip the snapshot—get the motion picture

Some organizations are satisfied with a one-time static snapshot of online opinion. “Brand testing? Yes, we did that last March.”

But often these point-in-time readings are out of date by the time they hit your desktop, not to speak of six months out. With social networks changing the refresh rate of the current online corpus to hourly, and individuals publishing whenever they have something new to say, keeping that thermometer in the water is more important than ever.

Ongoing sentiment testing provides ongoing insight through live monitors. This kind of currency can keep you informed enough to be able to minimize potential brand damage or take advantage of opportunities you hadn’t spotted.

Conclusion

Gauging and tracking human opinion in context can only be done through textual analysis—the careful examination of words. As we have seen, human expression is deeply compressed. We can learn much from small amounts. But this compression does not make analysis any easier; quite the opposite. Distilling accurate meaning from natural language is a difficult thing.

Happily, the sheer size of the online corpus and the new approach to science based on mining big data is providing a solution to this puzzle. The amount of writing on any topic imaginable, whether from an influential opinion-maker or new content contributor, dropped at our feet: this is the gift of social media. And the area of study is only growing. As it does, the accuracy of online opinion monitoring will increase and the feasibility of humans performing such analysis will decrease in proportion.

For those embarking on opinion analysis, we hope that our observations have given you what you need to gain a comprehensive, accurate, ongoing view of your brand’s public currency beyond counting mentions or identifying simple positive and negative views. It is our hope that you go forward and use online opinion monitoring to navigate with confidence, to intervene early, and to shape your strategy and policy based on real public opinion.

About Crimson Hexagon:

Crimson Hexagon, founded in 2007, is the leading provider of real-time social media monitoring and analysis to brands, agencies, media firms and their partners. Powered by patent-pending technology developed at Harvard University's Institute for Quantitative Social Science, the Crimson Hexagon Opinion Analysis Platform overcomes the limits of traditional market research by delivering a real-time view of how engaged online consumers truly think and feel about a brand or issue. For more information go to: http://www.crimsonhexagon.com. Read the Crimson Hexagon Social Media Monitoring and Analysis blog: http://www.crimsonhexagon.com/blog/.