Part I

Foundations

Cambridge University Press
978-1-107-03120-3 - Social Media Intelligence
Wendy W. Moe and David A. Schweidel
Excerpt
More information

1

The Beginnings of Social Media Intelligence

Let's say you just had a great dining experience at a new restaurant that opened down the street. Or you just saw the worst movie of your life. Many people with these experiences turn to social media to talk about their experiences and share their opinions (we discuss why people do this in Chapters 2–4). Some may write a lengthy review on their blog. Others may write shorter reviews and post to a review site (like Yelp or Rotten Tomatoes). And still others may choose to engage in a lengthy back-and-forth discussion about the merits and pitfalls of the experience in an online discussion forum.

Expressing opinions in this way is not new behavior. In the past, we referred to this as *word-of-mouth* behavior. Neighbors talking to neighbors about their new cars. Co-workers having conversations around the water cooler about a new computer or the latest events unfolding in their favorite television programs. But there are two fundamental differences between offline word-of-mouth activity and online conversations occurring in social media.

First, the online conversations taking place in social media are expressed in public forums. Opinions that were once expressed to friends, family, and colleagues privately are now broadcast publicly, sometimes for a wide and unrestricted audience, including the companies who create, manufacture, and sell the products being discussed as well as the competitors of these companies.

Second, online word-of-mouth is "on the record" and available for an extended period of time. In our backyard or water cooler conversations, the only people who knew what was said were the individuals present for the conversation. There was no written record, so the content of those conversations faded quickly along with people's precise memories of the event. In contrast, social media conversations can be revisited months or years later by anyone interested in the topic. The Library of Congress has archived conversations taking place on the micro-blogging platform Twitter since the service's founding in 2006.[1] Essentially, we are always "on the record" when we engage in social media conversations, and our opinions are preserved like historical documents for all to see.

As individuals flock to social media to discuss everything from the weather and politics to entertainment and other topics of interest, a wealth of data is created. However, many organizations struggle to interpret this data. Our goal for this book is to provide organizations with the knowledge needed *to use and interpret social media data in their strategic decision-making process*. We refer to this knowledge as social media intelligence (SMI).

We start by providing in this chapter an overview of how organizations currently collect and use social media data. By now, it's common practice for most organizations to engage in social media monitoring (or social media "listening") in one form or another. They scan posts, track metrics, and draw inferences all in an effort to keep their fingers on the pulse of their constituents and guide their decisions. However, there's a long way to go from this kind of social media monitoring to social media intelligence. Currently, organizations that engage in social media monitoring track a laundry list of easy-to-collect measures without an understanding of how these disparate metrics come together to inform an overall strategy. As a result, these organizations feel as if they are drowning in data and metrics.

Rather than collecting and monitoring what's easy to measure, organizations need to first put the dozens of social media reports aside and understand what's important to measure. The remainder of this book focuses on what we know about how people behave on social media and how their behaviors affect the organization. Such knowledge is critical in our ability to build social media intelligence that can leverage social media data in an effort to guide our organization's strategy.

What Is Social Media Monitoring?

People use social media for a variety of purposes: to stay connected with friends, share pictures with family, or share opinions with others. Social media monitoring is primarily concerned with this last application.

When an organization examines our behaviors on social media (for example, what comments we post to an online opinion forum or whether we "like" or "follow" a brand), it is engaging in *social media monitoring*. For the organization, the goal is to learn about its customers or other stakeholders and gauge their opinions in an effort to guide strategy.

Suppose you are a business whose customers are posting opinions online about your products. Based on these comments, you could redesign a product to correct flaws that early buyers have identified or make improvements based on specific requests. You might also alter your advertising to address misperceptions that you notice your customers hold. You could even use social media to monitor the health of your brand by watching for sudden shifts in opinions. The raw data that provide these insights are widely, and generally freely, available, allowing small and medium-sized businesses, as well as multinational corporations, to take advantage of them.

Social media monitoring is not restricted to just consumer products and services. Politicians who want to see how voters are responding to their platform, celebrities who are trying to manage their image, or government agencies that are attempting to respond to their constituents have all turned to social media monitoring.

Organizations may pursue a range of activities when they engage in social media monitoring. Some simply read a few posted comments here and there, looking for anecdotes that help them make a decision. Others use a number of data mining tools and quantitative metrics to systematically sift through the massive volumes of comments posted about them online.

Anecdotal Referencing of Social Media Comments

An easy-to-implement type of social monitoring is *anecdotal referencing*, the practice of reading a collection of comments and selecting a smaller set for further scrutiny.

The Land of Nod, a children's furniture brand owned by Crate and Barrel, sells their products exclusively through their own branded online, offline, and catalog stores. Their online store allows customers to post ratings and provide reviews for specific products. Like many brands, Land of Nod was initially apprehensive about allowing customer comments for fear that dissatisfied customers would post negative reviews that would in turn adversely affect sales. However, over time, they learned to monitor and appropriately respond to them. In fact, the company actively sought out negative reviews as a source of information. They treated these negative reviews like customer complaints and learned from the comments, redesigning products to correct flaws and improving customer service to avoid glitches. However, Land of Nod did not pursue a product modification for *every* online complaint. They simply read the comments posted and responded to the ones that in some way stood out. Many organizations pursue similar tactics. Often product teams read online comments and identify a few anecdotes to reference in order to justify a proposed product modification, hardly a systematic approach.

So why do so many organizations engage in anecdotal referencing? Because it is easy and inexpensive to implement. No data warehouse is necessary, and computer scientists and/or data mining experts are also not necessary. The associated implementation costs are limited to an employee's time spent reading and perhaps responding to comments. Many small and medium-sized businesses and other organizations with limited resources can engage in this form of social media monitoring easily.

This should be done with some caution, however. There are a number of significant drawbacks to relying on anecdotal references:

1. **Posting comments is a voluntary act and therefore posted opinions may not necessarily represent the opinions of the majority**. As we discuss in Chapter 3, extreme opinions are more likely to be expressed than more moderate ones. Take the case of a new restaurant. Suppose that those customers with negative experiences are a small minority of all customers but are more likely to express themselves on social media than other customers. If this were to happen, the business owner would find an abundance of criticism rather than praise posted online – not because all of his customers were unhappy but because those who chose to post comments online were systematically negative and not necessarily representative of his typical customer. In such a situation, businesses may create more problems than they solve if they address the concerns of these critics, especially if the critics represent a vocal minority who may have tastes and preferences that radically differ from those of the silent majority.[2] In such cases, actions taken to appease the vocal minority may actually alienate the majority of customers.
2. **Extreme opinions are oversampled**. Not only are extreme comments more likely to be contributed[3] but they're also more likely to receive the attention of the social media listener,[4] as in the Land of Nod example. When organizations scan and read select comments, they often do so without any systematic or scientific approach. This stands in sharp contrast to the way in which feedback is solicited through marketing research

surveys, where organizations carefully sample from the audience in order to ensure that the responses are representative of the population of interest. In many cases, the use of anecdotal referencing in the social media monitoring process involves the identification of extreme opinions. Comments that are either extremely positive or extremely negative tend to be selected and scrutinized. Any changes the organization makes to its product, service, or marketing strategy will then tend to over-focus on the slice of the minority that managed to attract the attention of monitors with their extreme opinions, further alienating the moderate majority.

3. **Anecdotal referencing is subject to analyst bias.** Social media researchers who monitor posted comments are also subject to researcher bias. In many cases, the analyst has a working hypothesis about the performance of the product, service, or brand that will taint how they interpret posted comments and which comments attract their attention.[5] For example, a brand manager who has a hunch that a product is overpriced may tend to focus on the comments that talk about price or interpret a comment (that may or may not express concern over price) as a complaint about price. The conclusions drawn from the anecdotal referencing of these comments may reflect the analyst's prior beliefs rather than providing an unbiased and representative view of customer response.
4. **Anecdotal referencing is not scalable.** While the practice of anecdotal referencing is relatively easy and inexpensive to implement on a small scale, it is not practical for large organizations or for topics that attract a high volume of comments. When the volume of comments is high, the drawbacks associated with anecdotal referencing are amplified as analysts cannot read all of the comments. Instead, they will have to pick and choose a few. But this sampling is not likely to be random; instead it tends to favor the negative, the extreme, and the opinions that support the analyst's prior hypotheses.

Text Mining

As mentioned, anecdotal referencing is not scalable. For a brand manager of a popular mass market product, reading every social media comment posted about the brand is not feasible. To handle large volumes of comments, many organizations have turned to text mining tools. A variety of text mining approaches exist, of varying levels of sophistication.

Word Counts and Word Clouds. Probably the most straightforward text mining approach is to simply count the number of times different words appear in a given set of documents. The appearance of words specific to a particular topic provides decision makers the ability to gauge the popularity of that topic.

While we do not intend to dwell on the underlying mechanics with which text mining is conducted, we would be remiss if we left readers with the impression that this is a trivial task. The raw data often need to be processed before they can be subjected to any quantitative analysis. For example, the text may need to be "stemmed," a process by which the suffix of a root word is removed so that words such as "buy," "buyer," and "buyers" would all be reduced down to "buy." This has the effect of significantly reducing the number of words for which word counts are being tabulated while still allowing analysts to extract the core meaning of the remaining words.[6] Most readers have experienced first-hand the practical benefits of stemming; it is incorporated into Google's search engine.[7]

Once word counts have been constructed, the conversation occurring around a particular topic can be represented using a word cloud – a visual representation of commonly used words found in a collection of comments. In a word cloud, the font size of the word represents the frequency with which that word was used. For example, when Steve Jobs (the founder and former CEO of Apple Computers) passed away in 2011, Apple set up a tribute page for Steve Jobs where fans could leave condolence messages. Selecting a sample of these messages, we can visualize them as a word cloud, as shown in Figure 1.1. In this case, *apple*, *world*, *thank*, and *family* were the words that appeared most often. Words such as *great*, *genius*, and *technology* were also common. Word clouds are straightforward to construct using word counts and provide an easy-to-understand graphic that highlights words that are often associated with the topic at issue. This can be extremely useful for marketing managers who need to understand how their products are perceived. The same line of thinking can be applied to celebrity agents or political campaign managers who are tasked with managing personal brands. Whether a company's product or an individual, brands want to be associated with positive words, such as *great*, *reliable*, or *quality*, and not negative words, such as *terrible*, *unreliable*, or *overpriced*.

Co-occurrence of Words. The next step in text analysis is to identify a few common themes in the comments. These themes emerge from the words that jointly appear in the same social media post. For example, words like *comfort*

Figure 1.1. Example word cloud

and *posh* may both appear in a word cloud for a five-star hotel, but both tap into a single underlying theme: *luxury*. Several statistical methods exist to identify underlying themes in data. These methods have historically been used to process survey data where multiple survey questions ask respondents to characterize a common aspect of the product. For example, questions about the hotel's housekeeping staff, front desk personnel, and room service all essentially represent service quality; statistical methods bring out underlying patterns in the respondents' answers to the questions about these different types of service.

These methods focus on identifying words that appear together in the same social media posting, or *co-occurrence*. After deconstructing a comment into individual word stems, researchers then look for patterns of co-occurrences.[8] For example, in reviewing a laptop computer, a consumer may comment, "The screen is beautiful and crystal clear!" The automated program would note that "screen," "beautiful," "crystal" and "clear" appear in the same comment. If other social media comments containing "screen" also contain the word "clear," then these two words are said to *co-occur*. If these words were to co-occur with a high frequency, the statistical analysis would reveal this and suggest that an analyst group the words together into a common theme.

Researchers have applied this approach to assess market structure.[9] For example, car shoppers are typically aware of and compare multiple brands before making a purchase decision. As a result, social media comments about vehicles often include both the vehicle the commenter ultimately chose to purchase and other vehicles considered. A consumer researching a car purchase may post the following question to an online community, "Should I purchase a BMW 335i or an Infiniti G37?" BMW and Infiniti both compete in the luxury car segment, but an automated computer program does not know this. All the program sees is that, in this comment, "BMW" and "Infiniti" appear together. If these words co-occur frequently throughout social media comments, a reasonable conclusion for a marketing researcher to draw is that these vehicles are considered by the same consumers and may therefore compete with each other. Historically, marketing researchers have generated such insights by administering surveys that ask shoppers to list the products they considered. Using text analysis allows for similar findings to be derived without the cost and time required to design and administer a survey. The comments available on blogs, forums, and other social media venues provide researchers with the raw data necessary to perform comparable analyses that provide market structure.

Sentiment Analysis. While the text mining approaches just described are effective at identifying words or topics commonly associated with a brand, it may not provide much insight on the sentiment associated with the comment. For example, the word *safety* may appear either for a car brand that is known for its outstanding safety record or for a car brand that has a poor safety record. The objective of sentiment analysis is to extract the sentiment (positive, negative, or neutral) expressed in social media comments. With the high volume of social media comments that need to be processed, researchers have tried to develop computer-assisted methods to automatically identify the sentiment in a given comment. However, these methods often encounter difficulty assessing sarcasm, slang, and other uses of language that do not perfectly conform to traditional grammar rules. Human analysts still provide a more accurate means of conducting sentiment analysis of social media posts.

When the topic of interest is very popular and is the subject of many comments, a combination of manual coding and automated content analysis offers a potential solution. Human analysts manually assess the sentiment in a *training set* of documents which are then compared by computer to a *test set* that contains a much larger set of documents. Documents in the test set that share similar characteristics of documents in the training set are said to express the same sentiment. While this semi-automated approach is not guaranteed