More than words: Social networks’ text mining for consumer brand sentiments

Mohamed M. Mostafa

Instituto Universitário de Lisboa, Business Research Unit, Avenida das Forças Armadas, Lisbon, Portugal

ARTICLE INFO

Keywords:
Consumer behavior
Global brands
Sentiment analysis
Text mining
Twitter

ABSTRACT

Blogs and social networks have recently become a valuable resource for mining sentiments in fields as diverse as customer relationship management, public opinion tracking and text filtering. In fact knowledge obtained from social networks such as Twitter and Facebook has been shown to be extremely valuable to marketing research companies, public opinion organizations and other text mining entities. However, Web texts have been classified as noisy as they represent considerable problems both at the lexical and the syntactic levels. This research used a random sample of 3516 tweets to evaluate consumers’ sentiment towards well-known brands such as Nokia, T-Mobile, IBM, KLM and DHL. We used an expert-predefined lexicon including around 6800 seed adjectives with known orientation to conduct the analysis. Our results indicate a generally positive consumer sentiment towards several famous brands. By using both a qualitative and quantitative methodology to analyze brands’ tweets, this study adds breadth and depth to the debate over attitudes towards cosmopolitan brands.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Opinions expressed in social networks play a major role in influencing public opinion’s behavior across areas as diverse as buying products, capturing the “pulse” of stock markets and voting for the president (Bai, 2011; Eirinaki, Pisal, & Singh, 2012). An opinion may be regarded as a statement in which the opinion holder makes a specific claim about a topic using a certain sentiment (Kim & Hovy, 2004). Web-generated opinions in blogs and social networks have recently become a valuable resource for mining user sentiments for the purpose of customer relationship management, public opinion tracking and text filtering (Zhang, Zeng, Li, Wang, & Zuo, 2009). Online opinions have been recently analyzed using sentiment analysis (SA). This is basically a natural language processing (NLP) application that uses computational linguistics and text mining to identify text sentiment, typically as positive, neutral or negative. This technique is also known in the text mining literature as emotional polarity analysis (EPA), opinion mining, review mining, or appraisal extraction (Zagal, Tomuro, & Shepitsen, 2012). Thus, SA can be regarded as an automated knowledge discovery technique that aims at finding hidden patterns in a large number of reviews, blogs or tweets. To calculate a sentiment score, the sentiment obtained from the text is compared to a lexicon or a dictionary to determine the strength of the sentiment. For example, the lexical resource SentiWord, which includes around 200,000 entries, uses a semi-supervised method to assign each word with positive, negative and objective scores. For instance, as Fig. 1 illustrates, a negative word might have in one of its senses a sentiment score of negative 0.375, positive 0.125 and objective 0.5.

Knowledge obtained from social networks are extremely valuable because millions of opinions expressed about a certain topic are highly unlikely to be biased. The affective nature of such opinions makes them easily understandable by the majority of readers, which increasingly make them the basis for making decisions regarding marketing research, business intelligence, stock market prediction and image monitoring (Montoyo, Martíniz-Barco, & Balahur, Forthcoming). However, almost all online text-based communications ignore the rules of spelling and grammar. In fact, Web texts have been classified as noisy as they still pose considerable problems both at the lexical and the syntactic levels (Boiy & Moens, 2009). At the lexical level, jargon, contractions of existing words/abbreviations, the use of emoticons and the creation of new words are the norm. At the syntactic level, we can hardly speak of real sentences. This writing style is evident in most forms of computer-mediated communication forums such as social network sites, bulletin boards and chat rooms (e.g., Derks, Fischer, & Bos, 2008). Although language purists might argue that such tendency represents poor language use, Thelwall, Buckley, Paltoglou, Cai, and Kappas (2010) claim that such use is prompted by technological advancements along with social factors. This complicating factor pertaining to informal Web texts’ sentiment detection has been dealt with through several techniques, including word sense disambiguation (Pederson, 2001), accurate detection of negation (Dave, Lawrence, & Pennock, 2003), and inferring semantic orientation from association (Turney & Littman, 2003). Dealing successfully with this problem has led to a plethora of online sentiment analyses in texts written in languages as diverse as English...
(e.g., Jansen, Zhang, Sobel, & Chowdury, 2009), Chinese (e.g., Xu, Liao, & Li, 2008) Arabic (Ahmed & Almas, 2005), and multi-languages (Abhasi, Chen, & Salem, 2008).

Although several studies have recently investigated SA (e.g., Cai, Spangler, Chen, & Zhang, 2010; Leong, Lee, & Mak, 2012), no previous studies have focused solely on investigating consumers’ sentiments towards major worldwide brands such as IBM, Nokia and DHL. In this study we aim to fill this void. We believe that by investigating brands’ polarity the study adds depth to the knowledge base on text mining. By using both a qualitative and quantitative methodology to analyze brand comments, this study also adds breadth to the debate over brand quality as perceived by consumers. Finally, by focusing solely on online texts, rather than on traditional offline data, this study enriches the knowledge base of this under-represented area. More specifically, this research aims attempts to answer the following research questions:

RQ1. Can social networks’ opinion mining techniques be used successfully to detect hidden patterns in consumers’ sentiments towards global brands?; and

RQ2. Can companies effectively use the blogosphere to redesign their marketing and advertising campaigns?

This paper is organized as follows. Next section provides a brief literature review on the major areas of SA applications. Section 3 deals with the method used to conduct the analysis. In this section issues related to research design and, sampling and data analysis techniques are presented. In Section 4, the results of sentiment analysis are presented. Finally, Section 5 presents research implications and limitations. This section also explores avenues for future research.

2. Literature review

SA techniques have been recently utilized in applications such as extracting suggestions from consumers’ product reviews (e.g., Vishwanath & Aishwarya, 2011), classifying consumers’ positive and negative product reviews (e.g., Turney, 2002), tracking sentiment trends in online discussion boards (e.g., Tong, 2001), detecting Internet hot spots (e.g., Li & Wu, 2010), tracking political opinions (e.g., Thomas, Pang, & Lee, 2006), determining consumers’ dissatisfaction with online advertising campaigns (e.g., Qiu et al., 2010b), tracking emotions in emails (Mohammad, 2012), predicting stock market movements (e.g., Wong, Xia, Xu, Wu, & Li, 2008) and differentiating between informative and emotional social media content (e.g., Denecke & Nejdi, 2009). An extensive literature review suggests that most SA applications might be classified into four distinct categories: product reviews, movie reviews, political orientation extraction and stock market predictions.

2.1. Product reviews

Blair-Goldensohn et al. (2008) used Google Maps data as input in order to analyze consumer sentiments towards hotels, department stores and restaurants. Using polarity values (positive/negative), the system developed was able to summarize sentiment regarding different aspects of the service provided such as value for money and ambience. In the same vein, Yi, Nasukawa, Bunesuc, and Niblack (2003) developed a sentiment analyzer to evaluate consumers’ opinions regarding digital camera features. The system used online text reviews to extract consumers’ sentiments regarding important features of digital cameras such as resolution and picture quality. Liu, Huang, An, and Yu (2007) used probabilistic Latent Sentiment Analysis (PLSA) to predict future product sales by examining bloggers’ sentiment.

Feldman, Fresko, Netzer, and Ungar (2007) developed a polarity system to analyze consumers’ comparison comments posted on discussion boards. The system used online information such as “300° C Touring looks so much better than the Magnum” to analyze consumers’ sentiments regarding several product aspects such as style, noise, quality and price. Hu and Liu (2004) used machine learning methods to extract and summarize consumers’ sentiments related to several electronic products, including mp3 players, digital cameras and mobile/ cellular phones. The system developed classified each review into positive or negative opinion and predicted future buying behavior. In a similar study, Miyoshi and Nakagami (2007) analyzed consumer sentiments regarding electronic products using adjective-noun pairs in a sentence.

In a recent study, Zhang, Xu, and Wan (2012) developed an expert system, called Weakness Finder, in order to analyze consumers’ sentiments in Chinese language online texts. The system extracts attitudes towards product features such as quality and price based on a morpheme-based analysis. The system was trained to utilize explicit and implicit sentiments to determine each sentiment’s polarity regarding products’ weaknesses. This study extended previous work by Ding, Liu, and Yu (2008) and by Liu (2010) because it took into consideration several linguistic aspects such as the adverbs of degree and the negation. Similarly, Abrahams, Jiao, Wang, and Fan (Forthcoming) employed text mining techniques to detect online consumer complaints regarding several automobile models. The authors found that consumer sentiments may be used to categorize and prioritize vehicle defects.

Pekar and Ou (2008) used sentiment analysis technique to evaluate 268 reviews of major hotels based on customers’ reviews posted on the website “epinions.com”. The authors used attributes such as food, room service, facilities and price to automatically analyze sentiments expressed towards those features. Finally, Na, Khoo, and Wu (2005) used support vector machines to classify 1800 product reviews into either recommended/positive sentiment or not recommended/negative sentiment. The authors used error analysis to improve initially obtained classification accuracy. Major sources of error in classification were due to negation, superficial words and comments on parts of the products.

2.2. Movie reviews

Na, Thet, and Khoo (2010) used a sample of 520 online movie reviews to conduct sentiment analysis. The authors compared textual characteristics of consumers’ reviews across four different genres to investigate sentiments expressed towards movies such as “Slumdog Millionaire”, “American Gangster” and “Burn after Reading”. Genres analyzed included discussion board threads, user reviews, critic reviews and bloggers’ postings. This study focused
on linguistic aspects of comments such as vocabulary, sentence length and part-of-speech distribution. The authors found that comments on discussion boards and user reviews contain more verbs and adverbs compared to the heavy usage of nouns and prepositions found in bloggers and critic postings. The study also identified the most frequent positive and negative terms used across different genres along with the distribution patterns of such terms.

Zhuang, Jing, and Zhu (2006) used machine learning methods to summarize online texts movie reviews sentiments. The authors aimed to find feature opinion pairs in consumers’ reviews by detecting feature classes such as “sound effects” and the stated opinion such as “excellent.” In a similar research design, Pang, Lee, and Vaithyanathan (2002) used support vector machines (SVM) to classify online sentiment classification of movie reviews. The authors used both single words (unigrams) and pairs of adjacent words (bigrams) to conduct the analysis. Compared with other machine learning classification methods, the SVM technique achieved the highest accuracy (83% correct classification). Based on the Page Rank algorithm, Wijaya and Bressan (2008) used online user reviews to evaluate movies. The authors’ reported results compared favorably with the rankings reported by the box office.

Thet, Na, and Khoo (2008a) used machine learning and information extraction techniques such as pronoun resolution and co-referencing to analyze sentiment orientation of movie review online texts. The authors correctly segmented customers’ reviews into relevant sections pertaining to different aspects of the movie such as the cast, the director and the overall rating. In a second study, Thet, Na, and Khoo (2008b) proposed an automatic method for determining movie reviews’ sentiment orientation and strength. The authors used a computational linguistics approach taking into consideration the grammatical dependency structure of each clause analyzed.

2.3. Political orientation

Larsson and Moe (2011) investigated Twitter’s emotional polarity during the 2010 Swedish election using around 100,000 tweets dealing with the election. The authors suggested a novel approach to classify high-end tweets messages among microbloggers into several categories such as senders, receivers and sender-receivers. Similarly, Tumasjan, Sprenger, Sandner, and Welpe (2011) investigated 100,000 tweets message referring either to a politician or to a political party in Germany to predict election outcome. Williams and Gulati (2008) also found that electoral success may be predicted accurately by the total number of Facebook supporters. The authors also found that tweets sentiment analysis can be used to accurately predict election outcome.

---

Table 1
Brands included in the study and their average sentiment scores.

<table>
<thead>
<tr>
<th>ID</th>
<th>Airline name</th>
<th># of tweets</th>
<th>Mean sentiment score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IBM</td>
<td>825</td>
<td>0.2508</td>
</tr>
<tr>
<td>2</td>
<td>Nokia</td>
<td>600</td>
<td>0.5351</td>
</tr>
<tr>
<td>3</td>
<td>Pfizer</td>
<td>373</td>
<td>−0.4021</td>
</tr>
<tr>
<td>4</td>
<td>Lufthansa</td>
<td>239</td>
<td>0.2343</td>
</tr>
<tr>
<td>5</td>
<td>KLM</td>
<td>222</td>
<td>0.0717</td>
</tr>
<tr>
<td>6</td>
<td>DHL</td>
<td>203</td>
<td>0.2562</td>
</tr>
<tr>
<td>7</td>
<td>Comcast</td>
<td>200</td>
<td>−0.3701</td>
</tr>
<tr>
<td>8</td>
<td>Mobinil</td>
<td>123</td>
<td>0.7951</td>
</tr>
<tr>
<td>9</td>
<td>Citi group bank</td>
<td>115</td>
<td>−0.0608</td>
</tr>
<tr>
<td>10</td>
<td>Air India</td>
<td>108</td>
<td>−0.0128</td>
</tr>
<tr>
<td>11</td>
<td>Novartis</td>
<td>101</td>
<td>0.1584</td>
</tr>
<tr>
<td>12</td>
<td>T-Mobil</td>
<td>100</td>
<td>−0.3312</td>
</tr>
<tr>
<td>13</td>
<td>US Bank</td>
<td>98</td>
<td>−0.8617</td>
</tr>
<tr>
<td>14</td>
<td>Samsung</td>
<td>86</td>
<td>0.0697</td>
</tr>
<tr>
<td>15</td>
<td>Al-Jazeera English</td>
<td>66</td>
<td>0.0735</td>
</tr>
<tr>
<td>16</td>
<td>Egypt air</td>
<td>57</td>
<td>−0.1853</td>
</tr>
<tr>
<td></td>
<td>Total and grand mean</td>
<td>3516</td>
<td>0.0270</td>
</tr>
</tbody>
</table>
In their seminal study on informal political texts, Malouf and Mullen (2008) argued that SA is a useful technique that might be used in order to analyze possible ideological biases, political opinions and political judgments' favorability. The authors used this technique to investigate political orientation among the users of a specific US web site dedicated to political discussions. This study is important because it is probably the first to investigate sentiment analysis in informal online political discussions—an area that is fast “becoming an important feature of the intellectual landscape of the Internet” (p. 175). Golbeck, Grimes, and Rogers (2010) also used SA to classify 6000 political tweets by US Congress’ members. The authors found that the major reason behind tweeting was to disseminate useful information, followed by tweets related to personal daily activities. The authors labeled the latter as a “vehicle for self-promotion” (p. 1620). Similarly, Ekdale, Namkoong, and Perlmutter (2010) empirically analyzed political bloggers’ behavior in the US. The authors found that the main reason for blogging was prompted by extrinsic motivations such as influencing public opinion or using the blogosphere as a plausible alternative to traditional media. In a study investigating microblogging participation within political environment, Gil De Zuniga, Puig-I-Abreu, and Rojas (2009) found that the major reasons for blogging are basically extrinsic motivations.

Heavy tweets used by Egyptian protesters from January 25 to February 11, 2011, which led ultimately to the forced resignation of ex-dictator Hosni Mubarak were extensively investigated (e.g., Lim, 2012; Papacharissi & Oliveira, 2012). These studies found that Twitter was used by protesters as an alternative to the blocked access to the Internet. The continuous stream of events provided by Twitter users was also found to be an accurate predictor of the crisis outcome. Park, Lim, Sams, Nam, and Park (2011) analyzed 2000 comments posted on ten Korean politicians’ visitor boards. The authors started by classifying the comments as positive, negative, or irrelevant. The authors found that positive comments represent the majority of all comments with a 51.3%. Negative comments represented 20.8% while irrelevant comments represented 27.9%. In terms of gender, the authors found that female comments were associated with more positive comments compared to male users (75.5% vs. 67.3%). Zappavigna (2011) used a large corpus of tweets (45,000) posted immediately after Obama’s US presidential elections victory in 2008. Using computational linguistic techniques, the author showed that the hashtag has “extended its meaning potential to operate as a linguistic marker referencing the target of evaluation in a tweet” (p. 788). Other studies investigating political sentiment analysis include studies by Efron (2004), Thomas et al. (2006), Park, Kim, and Barnett (2004) and Sobkowicz, Kaschesky, and Bouchard (Forthcoming).

2.4. Stock market prediction

Using automated natural language processing and machine learning techniques, Das and Chen (2001) classified sentiments expressed on Yahoo! Finance’s discussion board. The authors reported 82% accuracy in classifying posts into positive sentiment,
negative sentiment or neutral/irrelevant sentiment. In a similar study, Gu, Konana, Liu, Rajagopalan, and Ghosh (2006) used comments posted on Yahoo! Finance’s discussion board to predict different stocks’ future returns. Each post was classified into five possible categories: (2) for “strong buy”; (1) for “buy”; (0) for “hold”; (−1) for “sell”; and (−2) for “strong sell.” In this study the authors also used a weighting scheme to assign weights for each sentiment obtained based on the reputation and previous accuracy of the poster. Using a simulated environment to mimic real trading, the authors reported around four percent increase in returns over one month based on sentiment analysis.

Bollen, Mao, and Zeng (2011) found that the aggregation of millions of tweets posted daily on Twitter can be used to predict stock market over time. The authors used measures such as daily Twitter posts over around ten months to predict the Dow Jones Industrial average closing values. To cross-validate the results, the authors also used the resulting time series of Twitter moods to detect the general public’s response towards the outcome of the US presidential campaign. Other studies investigated the relationship between investors' sentiments and other factors such as stock returns following a major earthquake (Shan & Gong, 2012), air disasters involving US vs. foreign airlines (Kaplanski & Levy, 2010) and local sports events (Chang, Chen, Chou, & Lin, 2012).

3. Method

3.1. Twitter sampling

Twitter is a microblogging service that was launched formally on July 13, 2006. Unlike other social media, Twitter is considered a microblog because its central activity revolves around posting short updates or tweets using the Web or mobile/cell phones. The maximum size of the blog is 140 characters—roughly the size of a newspaper headline. According to Semiocast.com (2012), a marketing research company, there are now around 500 million active Twitterers. Fig. 2 shows top ranked countries according to active tweets in 2012 (Semiocast.com, 2012). Tweets are available publicly as a default, and are also directly broadcasted to the user’s followers (Bliss, Klouman, Harris, Danforth, & Dodds, 2012).

A recent analysis of Twitter activities found that more than 80% of the users either update their followers on what they actually doing or disseminate information regarding their daily experiences (Thelwall, Buckley, & Paltoglou, 2011). Since Twitter is the most large, popular and well-known microblog Web site, it was selected to conduct the analysis reported in this study. The data used represent a random set of Twitter posts from July 18, 2012, to August 17, 2012. The data comprised 3516 tweets for sixteen brands. To guarantee representativeness, sample selection has been varied by day of the week and hours in the day. Our sample size is comparable in size to Qiu, He, Zhang, Shi, Bu, and Chen's (2010) sample, which included 3783 opinion sentences. Table 1 shows the random sample of tweets for each brand included in the study. Following Thelwall et al. (2011), only tweets in English was chosen in order to remove complications that might arise with analyzing multilingual tweets.

Table 2 shows a sample of tweets for Air India with a manual classification of customers’ sentiments. As can be seen from the table, tweets represent a very noisy environment in which messages posted to virtual audience includes abbreviated words, the @ and the hashtag (#) characters, and heteroglossia-referring to other voices in the tweets in order to convey interpersonal and ideational
meanings (Bakhtin, 1981). Huang, Thornton, and Efthimiadis (2010) found that the hashtag was invented by Twitter users early in 2008 to help followers find a specific tweet or post. As opposed to the hashtag, the @ character has been introduced to address a tweet to another follower, which allows Twitter to function effectively as a collaboration and conversation system (Honeycutt & Herring, 2009).

3.2. Lexicon

Categorizing words for SA is a major step in applying the technique. Broadly speaking, there are two widely used methods for sentiment orientation identification: the lexicon-based approach and the corpus-based method (Miao, Li, & Zeng, 2010). However, since the corpus-based method has rarely been used to analyze sentiment orientation, we will focus here on the lexicon-based method. Nevertheless, both methods require a pre-defined dictionary or corpus of subjective words. The sentiment is determined by comparing tweets against the expert-defined entry in the dictionary, which makes it easy to determine the polarity of a specific sentence. Thus, it is crucial to have an accurate classifier to be used to construct indicators of sentiment. Previous research has typically incorporated lexicons such as the manually coded General Inquirer (Stone, Dunphy, Smith, & Ogilvie, 1966), which includes over 11000 hand-coded word stems in 182 categories, the LIWC dictionary (Pennebaker, Mehl, & Niederhoffer, 2003), the SentiWordNet (Baccianella, Esuli, & Sebastiani, 2010), the Q-WordNet (Agerri & Garcia-Serrano, 2010) or the lexicon of subjectivity clues (Wiebe, Wilson, Bruce, Bell, & Martin, 2004). Automatically-coded lexicons have recently been developed, including the sentiment-based lexicon (Taboada, Brooke, Tofiloski, Voll, & Stede, 2011).

In this study we used the Hu and Liu (2004) lexicon to conduct the analysis because this dictionary has been used successfully in a similar application (Miner et al., 2012). This lexicon includes around 6800 seed adjectives with known orientation (2006 positive words and 4783 negative words). The lexicon has recently been updated by adding words based on a thorough search in the WordNet. The lexicon size is similar to the recently used Opinion Finder lexicon, which included 2718 positive words and 4912 negative words (Bollen et al., 2011). Our lexicon is also comparable to the dictionary used by Hu, Bose, Köh, and Liu (2012), which includes 1635 positive words and 2005 negative words. Modifications on this approach include the handling of negation (Das & Chen, 2001) and the weight of enhancer/intensifier words such as the use of the “really” or “absolutely” (Turney, 2002). Other attempts have been recently tried to develop lexicons capable of detecting deep sentiments expressed in discussions among several actors in informal situations (Maks & Vossen, 2012). Similarly, Reyes and Rosso (2012) developed a corpus to detect irony in consumers’ reviews.

Fig. 4. A 3-D map with link strengths and base lines shown.
4. Results

4.1. Exploratory data analysis and visualizations

To conduct the qualitative part of this study we used QDA Miner 4.0 software package (Provalis Research, 2011) for coding textual data posted on Twitter. This software was selected because of its extensive exploratory tools that can be used to identify hidden patterns in textual data. In order to analyze consumer sentiments towards brands, we started by generating relative frequency word counts. Table 3 shows the percentage of words in a random set of tweets.

From Table 3 we can see that words such as “global”, “flight” and “price” have the highest frequency for a brand such as Egypt Air. However, references were also made in the tweets to countries such as Syria, probably because of the ongoing uprising in that country. Analyzing frequency of appearance or simply the incremental count of appearance of particular words or phrases might provide insights into a particular topic. In fact O’Leary (2011) argues that despite the simplicity of such approach, it can be used to predict characteristics of the topic analyzed. Fig. 3 shows a proximity plot constructed based on Egypt Air tweets. This figure shows visually, on a single axis, the distance from a particular object to all other objects. This graph was constructed to extract huge amount of data based on a distance matrix. From this graph we see that most tweets were concerned with things like “price”, “tickets” and “monopoly.” However, Twitters seem to be also concerned with fighting in neighboring Syria.

Fig. 4 shows a 3-D concept map constructed based on multidimensional scaling (MDS) technique. In this graph the closer the cases, the higher the tendency of co-occurrence and vice versa. The lines on the map represent levels of association among words. From the graph we can reconstruct the most influential tweets. For example, for Egypt Air, we can see the following pattern: closeness to home as represented by words such as “Egypt”, “Dubai” and “Resident”, personification as represented by names of people tweeting, and relevance and significance as represented by words such as “Dreamliner”, “Service”, and “Flight.” This result is in line with a recent study using centering resonance analysis, a computational discourse analysis technique, on a random sample of 9000 Egyptian tweets (Papacharissi & Oliveira, 2012).

4.2. Overall sentiment scores

We used the twitteR, the plyr, stringr and the ggplot2 libraries in the R software package version 2.15 to conduct the quantitative sentiment score. Fig. 5 shows the distribution of sentiment scores obtained for Nokia and Pfizer (similar graphs were constructed for remaining brands). From the graph, we immediately recognize some asymmetry. For example, the bars at +1 are much larger for Nokia compared to the +1 bars for Pfizer. It is also evident that the bars at –1 are much larger for Pfizer compared to the –1 bars for Nokia. This makes it clear that the overall sentiment score for Nokia is generally better than the sentiment score for Pfizer.

The visualization of the sentiment distribution in Fig. 5 further underlines the fact that most tweets fall either on the neutral point (0) or within the band of circa –1/+1, which is an indication that several tweets are not very affective. Although this result is in line with Lindgren (2012), we can focus only on positive or negative sentiments. Following Miner et al. (2012), we ignored the middle and constructed sentiment scores for a random sample of only positive and negative sentiments. Fig. 6 shows results for only three brands. From Fig. 6, we clearly see that most tweet comments were positive for both Lufthansa and DHL. However, most of the tweets were negative for T-Mobile.

Finally, we used the StreamGraph software package (Clark, 2008) to visualize the trend of tweets across a period of time for all brands. Fig. 7 shows tweets trend for three brands: Nokia, Pfizer and Al-Jazeera English. This graph is useful since it represents multiple time series data stacked on top of the other (Havre, Hetzler, Whitney, & Nowell, 2002). Since the total frequency of all features represents the height of the curve, each time series data should be read off the figure not as the cumulative height but rather as starting with zero. As can be seen the graph is characterized by a number of spikes, indicating an increase in tweets’ frequency at those particular times. Interestingly, we see spikes on the top part of the figure representing Nokia’s introduction of the new smart phone Nokia Lumia brand.

5. Implications, limitations and future research

In this study we analyzed sentiment polarity of more than 3500 social media tweets expressing attitudes towards sixteen global brands. Social media users represent 67% of around a billion Internet active users (Eirinaki et al., 2012). Although a single tweet is limited to 140 characters in length, the millions of tweets posted on Twitter almost on a daily basis might provide an unbiased representation of consumers’ sentiment towards services and brands. Capturing consumers’ opinions and gaining knowledge about consumer preferences has long been a major concern for marketing researchers. However, traditional marketing methods such as focus groups and face-to-face interviewing are both costly and time consuming. In contrast, tweets and blogs are readily available for free. Such consumer generated media are also free of bias that might be introduced by the interviewer in case of...
personal interviews. Moreover, consumers opinion-based expressed qualitatively may easily be benchmarked against objective measures such as sales data, revenues, or stock price. Thus, companies can utilize such online textual content in an effort to gain insight into consumers' opinions regarding available products and services. Ignoring consumer generated sentiments might put companies in a competitive disadvantage and could also create significant brand image problems. The speed of social media might also render companies' advertising and publicity using traditional media useless.

Based on the fact that around 20% of microblogs mention a brand name (Jansen et al., 2009), we argue that managing brand perception on Twitter and other social media should form part of the company's overall proactive marketing strategy. Maintaining a constant presence on such media channels should also be an important part of the company's branding and advertising campaigns. Companies can use the blogosphere in a smart way to disseminate information needed by its customers and to monitor Twitter's and bloggers' discussions regarding its brand. By doing this, companies can track tweets and intervene immediately to communicate with dissatisfied customers. On the other hand, advertising campaigns might make use of positive tweets, which can form a part of the company's viral marketing efforts. Thus, companies may use consumers' tweets as a feedback about services and products by encouraging electronic word of mouth (e-WOM). This can be done online without investing huge amounts on traditional advertising and marketing campaigns. On the other hand, companies should not ignore negative tweets because such tweets might be used to detect what is not going right with a product or a service. Ultimately, tweets can be used effectively to identify consumers' preferences, to detect dissatisfaction related to a product defect, and to correct unintended errors. Since tweets enable companies to be more efficient and functional in dealing with customers, they should incorporate SA into their text retrieval technologies and into their search engines. This is because SA can be extremely useful in areas such as analyzing consumer trends, handling customers' feedback and targeting advertising campaigns.

However, it should be noted that while we conducted SA to objectively classify consumers' opinions, our analysis does not reveal the underlying reason behind forming such opinions. Future research using sentiment topic recognition (STR) should be conducted to determine the most representative topics discussed behind each sentiment. Through this analysis, it should be possible to gain overall knowledge regarding the underlying causes of positive or negative sentiments. It should also be noted that while the lexicon-based approach used in this study can detect basic sentiments, such approach may sometimes fall short of recognizing the subtle forms of linguistic expression used in situations such as sarcasm, irony or provocation. For instance, Boiy and Moens (2009) provide an excellent example by showing that the part of the sentence that follows “même si/even if” in a French sentence expresses the least affective feeling “[mêmen si le film a eu beaucoup de succès, je le trouvais vraiment nul!/even though the movie had a lot of success, I really found it nothing!!]”. Future research should attempt to find a way to deal with this problem. A huge corpus that includes large training data sets representing such idiomatic usage may be worth trying. Finally, opinions expressed by consumers might in fact be a manipulation of online vendors' opinions posing as real consumers. Future research should also attempt to detect genuine sentiments from opinions that merely reflect the position of vendors interested in selling more products or services.

Despite these limitations, we believe that this study contributes to the existing literature in text mining and consumer behavior. First, we used a number of well-known brands, which ensures that...
our findings are practical, influential and generalizable. Second, we approached our analysis using the most widely-used microblogging site-Twitter by employing a mixed methods approach based on both qualitative and quantitative methods. This ensures our results robustness. Finally, by focusing only on consumer tweets we contribute to the growing body of literature on e-WOM (e.g., Jansen et al., 2009). Therefore, we believe that our research is timely and impactful.

Fig. 7. Twitter stream graph (1000 tweets each) for Nokia (top), Pfizer (middle) and Al-Jazeera English brands (bottom).
References


