Industrial Management & Data Systems

Gaining competitive intelligence from social media data: Evidence from two largest retail chains in the world

Wu He, Jiancheng Shen, Xin Tian, Yaohang Li, Vasudeva Akula, Gongjun Yan, Ran Tao,

Received 26 March 2015
Revised 22 August 2015
7 September 2015
Accepted 13 September 2015

Gaining competitive intelligence from social media data

Evidence from two largest retail chains in the world

Wu He
Department of Information Technology & Decision Sciences, Old Dominion University, Norfolk, Virginia, USA
Jiancheng Shen
Department of Finance, Old Dominion University, Norfolk, Virginia, USA
Xin Tian
Department of Information Technology & Decision Sciences, Old Dominion University, Norfolk, Virginia, USA
Yaohang Li
Department of Computer Science, Old Dominion University, Norfolk, Virginia, USA
Vasudeva Akula
VOZIQ Company, Reston, Virginia, USA
Gongjun Yan
Romain College of Business, University of Southern Indiana, Evansville, Indiana, USA, and
Ran Tao
School of Computer Science & Technology, Donghua University, Shanghai, China

Abstract

Purpose – Social media analytics uses data mining platforms, tools and analytics techniques to collect, monitor and analyze massive amounts of social media data to extract useful patterns, gain insight into market requirements and enhance business intelligence. The purpose of this paper is to propose a framework for social media competitive intelligence to enhance business value and market intelligence.

Design/methodology/approach – The authors conducted a case study to collect and analyze a data set with nearly half million tweets related to two largest retail chains in the world: Walmart and Costco in the past three months during December 1, 2014-February 28, 2015.

Findings – The results of the case study revealed the value of analyzing social media mentions and conducting sentiment analysis and comparison on individual product level. In addition to analyzing the social media data-at-rest, the proposed framework and the case study results also indicate that there is a strong need for creating a social media data application that can conduct real-time social media competitive intelligence for social media data-in-motion.

Originality/value – So far there is little research to guide businesses for social media competitive intelligence. This paper proposes a novel framework for social media competitive intelligence to illustrate how organizations can leverage social media analytics to enhance business value through a case study.

Keywords Competitive intelligence, Social media analytics

Paper type Research paper

Industrial Management & Data Systems
Vol. 115 No. 9, 2015
pp. 1622-1636

0263-5577
DOI 10.1108/IMDS-03-2015-0098

Introduction

According to the IBM research, about 2.5 quintillion bytes of data are generated every day (Zikopoulos *et al.*, 2012). Many organizations are developing business analytics initiatives in order to extract new insights from the massive data sets they collected over time. These massive data sets typically contain data from many varieties of sources including internal systems and external data sources such as social media. Most early business analytics efforts are focussed on analyzing internal data, which is considered as the primary source of data within their organizations. For example, Walmart's transactional databases are estimated to contain more than 2.5 petabytes of data about their customers, products and others (Tan *et al.*, 2013).

In addition to analyzing internal data, a recent trend is to analyze massive volume of external data such as social media data related to the organizations (Zikopoulos *et al.*, 2012). In recent years we have observed the rapid development of social media which have greatly influenced the way in which people communicate with one another and obtain information (Ngai *et al.*, 2015). Nowadays social media have become ubiquitous and are playing an important role in people's daily lives. As a result, a large amount of user-generated content is available on social media sites. For example, in the business field, more and more consumers rely on user-generated reviews to evaluate products and services prior to making a purchase. Many tourists choose a restaurant based on Yelp reviews and ratings (Kang *et al.*, 2013). In the health care field, many patients use social media to discuss medical services and their doctors in order to optimize treatments; in education, many students and parents use social media platforms to exchange opinions and discuss the pros and cons of their interested colleges. User-generated social media content is offering unprecedented opportunities as well as challenges to organizations because they contain a deluge of opinions, viewpoints and conversations by millions of users. There is a need for organizations to efficiently manipulating and analyzing user-generated social media content related to their organizations (Barbier and Liu, 2011; He *et al.*, 2013). For example, CEOs may want to gain business insight or knowledge based on what customers are discussing on Facebook or Twitter about their organizations so that they can improve decision making and other aspects of business performance. A recent study by Bain & Company (Pearson and Wegener, 2013), a leading management consulting firm, found that large companies who have adopted advanced data analytics capabilities outperform their competitors by wide margin. They also found that the big data analytics adopters are five times faster in making good decisions when compared to their competitors and twice as likely to be in the top quartile of financial performance within their industries due to their insight knowledge they obtained from the data analytics.

The rest of the paper is organized as follows. The following section provides a brief literature review of social media analytics and its application. Penultimate section presents a case study that analyzes a massive data set composed of customer tweets about two large retailers: Walmart and Costco. Lastsection discusses the findings, implications and limitations. Finally, conclusions and suggestions for future studies are provided.

Literature review: social media analytics and intelligence

Zeng *et al.* (2010) define social media analytics as "using advanced informatics tools and analytics techniques to collect, monitor, and analyze social media data to extract useful patterns and intelligence." So far a number of studies including multiple special issues related to social media analytics have been published (Chen and Yang, 2011; Fan and Yan, 2015; Zeng *et al.*, 2010).

Different from traditional transaction data, social media data have several unique characteristics. First, social media cover general users' opinions about almost every aspect of our life. Second, there are always fresh content on social media and the content are updated consistently and timely by numerous online users. Third, social media contents are associated with metadata in various attributes such as user, location, likes, time, dislikes, etc. Fourth, social media data have quality issues and contain a lot of noise and spams, which need to be sifted through to figure out what data can be trusted. Real benefit can be obtained by analyzing massive social media data in real time and gaining trustworthy insights while social media data are continuously coming in high speed. For example, the US Geological Survey monitors Twitter messages containing the word "earthquake" to detect identify possible earthquakes events because people begin sending tweets with in tens of seconds after feeling shaking (Earle *et al.*, 2012). The New York Stock Exchange announced to provide investors with real-time information about trade-related public sentiment in 2013 (NYSE, 2013). Bollen *et al.* (2011) did sentiment analysis on a large corpus of tweets and found that the mood of the Twitter population can be used to predict the movement of the Dow Jones Industrial Average (DJIA) on the following day with a claimed accuracy of 87.6 percent. Risius and Beck (2015) conducted a multilevel analysis of approximately five million tweets regarding the main Twitter accounts of 28 large global companies and identified different social media activities in terms of social media management strategies, account types (broadcasting or receiving information) and communicative approaches (conversational or disseminative). Lansdall-Welfare *et al.* (2012) analyzed 484 million tweets posted by more than 9.8 million users over a period of three years and revealed changes in public mood and associated mood changes with important events. However, the rapid increase of social media data evolves into a problem because most organizations do not have the capability and skills to process such huge amounts of data in a timely manner, identify trustworthy data and gain values from these massive data to enhance their business such as marketing, products, sales and online commerce.

Furthermore, Zeng *et al.* (2010) came up with a new term called social media intelligence. By leveraging technology, solution frameworks and toolsets from social media analytics, social media intelligence aims to "derive actionable information from social media in context-rich application settings, develop corresponding decision-making or decision-aiding frameworks, and provide architectural designs and solution frameworks for existing and new applications that can benefit from the 'wisdom of crowds' through the Web." Compared to social media analytics, social media intelligence research is still limited and immature. More discussions and research on social media intelligence in various application settings with an aim to support decisions are needed.

A proposed framework for social media competitive intelligence

Business analytics techniques enable organizations to conduct deep analysis of their business data to identify potential issues, problems, opportunities and best practices. However, as RemitDATA (2013) questions:

- How do organizations know whether or not their own best practices are actually "best"?
- What if an organization's internal best practices are still poor when compared to peers?

Therefore, there is a need to benchmark comparisons against competitors. Competitive intelligence offers an approach for organizations to compare their performance against their peer organizations (Sanderson, 2013; Bose, 2008). As a result of the comparison, organizations can focus their efforts on improving the areas that are still poor when compared to peers and also develop efforts that can have the greatest impact (RemitDATA, 2013). In this paper, we propose a framework for social media data competitive intelligence (Figure 1) to help interested organizations leverage big data solutions to contextually compare their social media data against peers. Competitive intelligence was frequently described as the external sourcing of information for decision making (Ross *et al.*, 2012). Competitive intelligence offers a number of benefits such as creating new growth opportunities, minimizing the impact of surprises, enabling faster responses to changes in the market place, improving the quality of strategic planning processes, identifying potential vulnerabilities, providing early warning or alert for competitive threats (Chen and Das, 2010; Bose, 2008; Ross *et al.*, 2012).

The social comparison theory (Festinger, 1954) suggests that competitions are ubiquitous and people commonly seek to achieve a superior position to others in various contexts, from daily social situations to organizational settings and market transactions (Garcia *et al.*, 2013). From the buyer perspective, consumers today tend to compare products and read online reviews from different sources on the internet before they make the purchase decision. From the seller perspective, businesses face increasing pressure to increase income and retain clients and thus it is natural for them to compare themselves with competitors in the same market place in terms of products and services. Instead of being reactive, literature implies that all businesses should be more proactive and conduct market intelligence activities as a formal regular business process to enhance their capability and ability to respond to changing market conditions, customer requirements and competitor's actions (Ross *et al.*, 2012).

So far there is no framework to guide businesses for social media competitive intelligence. Due to privacy and data sharing concerns, it could be difficult for organizations to compare their internal data with their competitors. However, contextually comparing their external social media data against competitors is possible because social media data are publicly available. Social media data can be divided into two main categories: social media data-at-rest and social media data-in-motion (Gupta *et al.*, 2012). Social data competitive intelligence can be used to analyze both data-at-rest and data-in-motion (Bifet and Frank, 2010; Benhardus and Kalita, 2013; He *et al.*, 2015), understand what is happening in real time, and identify events, problems and issues in real time, such as determining when customer sentiment in social media is becoming more negative. In today's highly competitive business world, social media data competitive intelligence could become a key differentiator for organizations in industries such as making a difference between profit and loss in financial services markets (Zikopoulos *et al.*, 2012).

In the proposed framework, data mining technology is used as a solution to analyze the social media data including both data-in-motion and data-at-rest from the organizations and their peers, visualize and benchmark comparisons against peers across events, products, issues and any other areas that may impact the business. Existing data mining platforms from vendors like IBM, SAP, Oracle and Microsoft can be integrated to store, manage, analyze and compare data across numerous social media sources from different companies. Such comparisons are supported by advanced analytics, business intelligence and contextual data found from the data set. Social media competitive intelligence can be used to identify and assess industry

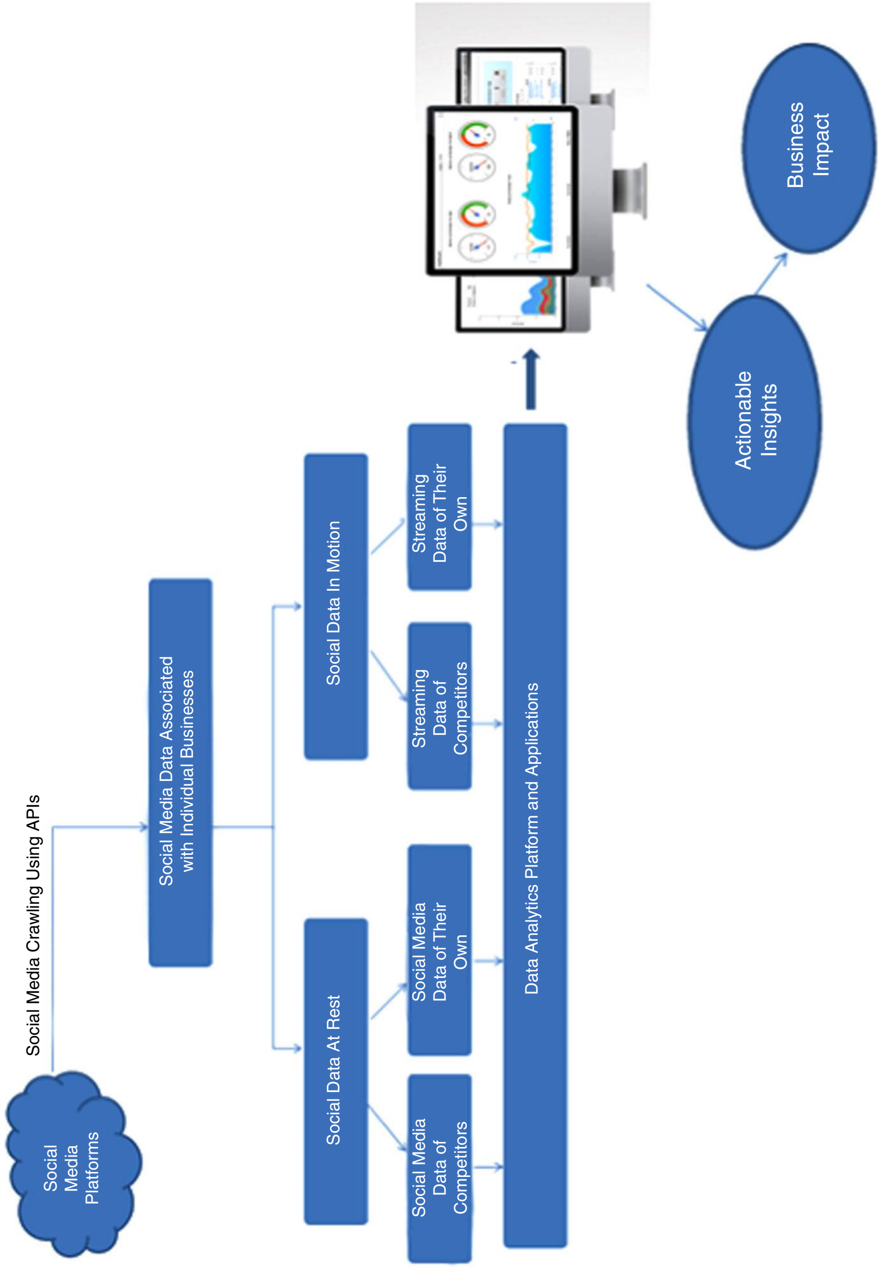

Figure 1.
A proposed framework for social media competitive intelligence

trends that shape the business environment. This proposed framework is especially applicable to consumer service-oriented sectors such as financial services, health care, retail and e-business.

Case study

Implementing the proposed social media competitive intelligence framework can result in numerous business benefits such as improving decision making and operational excellence. As an initial effort to implement the proposed framework, we conducted a pilot case study by using tweets related to two largest retail chains in the world: Walmart and Costco. In 2015, Walmart is the largest and Costco is the second largest retail chain in the world in terms of retail revenue. Both Walmart and Costco have stores in multiple countries. Tweets are short and constantly generated by online users. We collaborated with VOZIP, a social media analytics company based in USA, to gather the tweets (containing the two companies' names) that were submitted to the Twitter service during December 1, 2014 to February 28, 2015. The tweets were collected using Twitter search APIs (see Table I). VOZIQ crawls a huge amount of Twitter data that contain specific keywords utilizing Twitter search APIs on a daily basis. In particularly, VOZIQ uses Apache Solr for text searches, used Hadoop for big data analysis, used MySQL for storing processed data and used JavaServer Pages (JSP) for web-based visualization. This tool is currently available for online access at www.voziq.com. This tool is set up on Amazon AWS cloud platform to offer cloud-based services and for on-demand scaling to process large volumes of social media data. As many social media sites have millions of users and number of tweets often grows at high speed, it is not always possible to save the data in full due to resource constraint. Thus, this case study used the captured tweets (nearly half million tweets) in the past three months during December 1, 2014 to February 28, 2015 as our data set for analysis. Table I shows our data set for this case study.

A popular sentiment analysis tool called Lexalytics was used to detect sentiments of each tweet in our data set. Lexalytics offers a sentiment analysis algorithm to identify the emotive phrases within a document. After each phrase is scored (roughly −1-+1), the scores of all the emotive phrases are combined to discern the overall sentiment of the sentence. We created three pieces of information based on VOZIP raw Twitter data: the number of tweets for a day, average positive sentiment, average negative sentiment. The number of tweets for a day recorded the actual number of tweets posted on Twitter related to a specific company in a particular day. Average positive sentiment comes from adding up all the absolute value of positive sentiment, and then averaging the total out. It ranges from 0 to 1. Average negative sentiment is calculated by adding up the absolute value of negative sentiment, and then averaging the total out. It also ranges from 0 to 1. Figures 2 and 3 illustrate the overall volume and sentiment trend analysis for the two big retailers over a period of three months.

In addition to the overall volume and sentiment trend analysis, we were particularly interested in volume and sentiment trend analysis on individual product level since both Walmart and Costco are direct competitors and often sell the same type of

Companies	Number of Tweets containing the company name
Walmart	246,442
Costco	229,517

Table I. Twitter data set used in this case study products. We decided to choose four highly popular grocery products: muffin, cookie, pizza and chicken and analyzed their tweets for comparative analytics. We used a well-known text analysis tool called NVivo 10 to query the four products from the gathered tweets, and then analyzed the content and sentiment of these products. Tables II and III show the total mentions, number of positive, neutral and negative tweets related to the four products of the two big retailers. As an example, we used Figures 4 and 5 to illustrate the volume and sentiment trend analysis for both cookie and muffin of the two big retailers, respectively. Due to the page length limitation, we did not include the figures for pizza and chicken. Based on the product-level analysis, we found that although customers mentioned Walmart more than Costco on Twitter during that period, people tend to talk about Costco's muffin, cookie, pizza and chicken more than Walmart's muffin and cookie on Twitter. We saw a higher number of positive comments and negative comments on Costco's muffin, cookie, pizza and chicken than on Walmart's muffin, cookie, pizza and chicken. However, Costco received a lower percentage of positive comments and a higher percentage of negative comments than

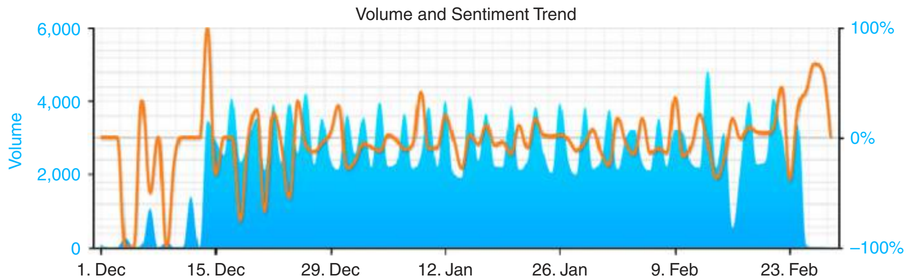

Figure 2.
Walmart volume and sentiment trend

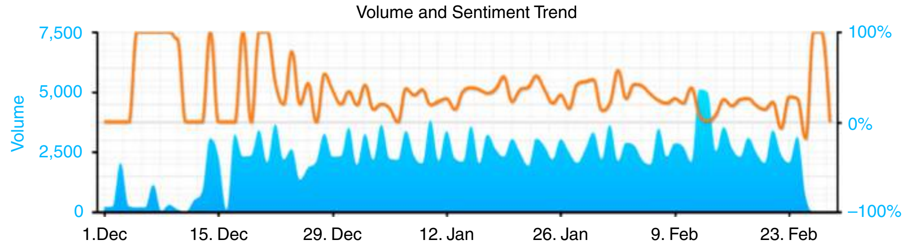

Figure 3.
Costco volume and sentiment trend

Table II.
Costco Twitter data from December 1, 2014-February 28, 2015

Product name	No. of mentions	Positive sentiment	Neutral sentiment	Negative sentiment	Overall sentiment
Muffin	1,103/229,517 (0.48%)	191/1,103 (17.32%)	850/1,103 (77.06%)	62/1,103 (5.62%)	Positive
Cookie	784/229,517 (0.34%)	113/784 (14.41%)	609/784 (77.68%)	62/784 (7.91%)	Positive
Pizza	13,242/229,517 (5.77%)	1,850/13,242 (13.97%)	10,759/13,242 (81.25%)	633/13,242 (4.78%)	Positive
Chicken	2,670/229,517 (1.16%)	424/2,670 (15.88%)	2,051/2,670 (76.82%)	195/2,670 (7.30%)	Positive

Walmart for muffin; in contrast, Costco received a lower percentage of positive comments and a lower percentage of negative comments than Walmart for cookie and pizza. As for the chicken, Costco received a similar percentage of positive comments and a lower percentage of negative comments than Walmart. These product-level comparisons reveal potential space for improvement.

Furthermore, we used a popular text mining tools called Leximancer to mine and cluster the tweets related to each of the four products in order to better understand what customers are talking about for each of the products (Liau and Tan, 2014; Wang and Wang, 2014). As an example, Figure 6 lists the generated cluster diagrams for cookie-related Costco and Walmart tweets. After examining and analyzing the

Product name	No. of mentions	Positive sentiment	Neutral sentiment	Negative sentiment	Overall sentiment
Muffin	41/246,442 (0.01%)	13/41 (31.71%)	26/41 (63.41%)	2/41 (4.88%)	Positive
Cookie	506/246,442 (0.21%)	84/506 (16.60%)	366/506 (72.33%)	56/506 (11.07%)	Neutral-positive
Pizza	410/246,442 (0.17%)	60/410 (14.63%)	310/410 (75.61%)	40/410 (9.76%)	Positive
Chicken	468/246,442 (0.19%)	72/468 (15.38%)	344/468 (73.51%)	52/468 (11.11%)	Neutral-positive

Table III. Walmart Twitter data from December 1, 2014 to February 28, 2015

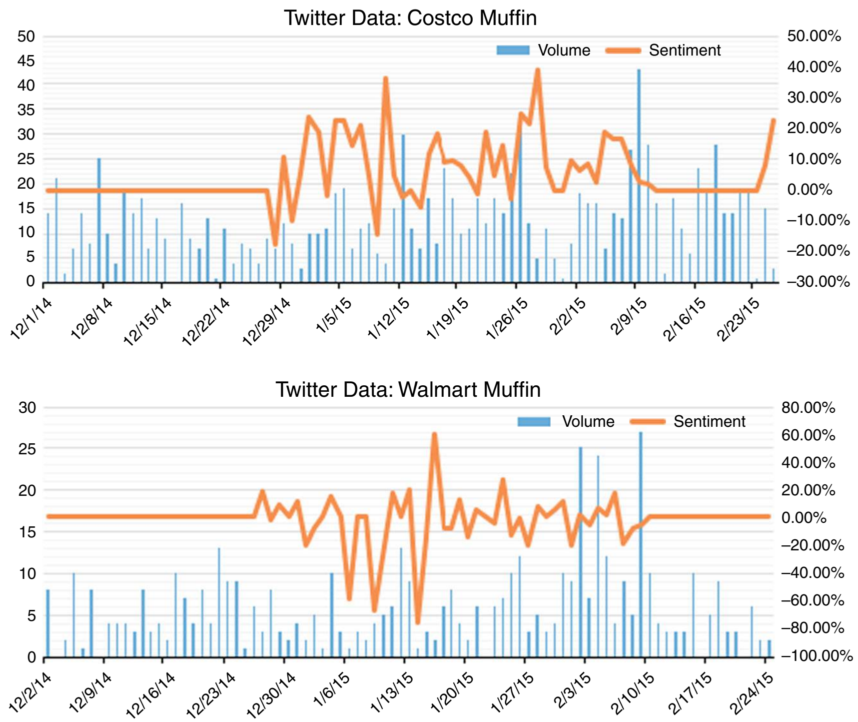

Figure 4. Twitter data comparision of muffin for Costco and Walmart

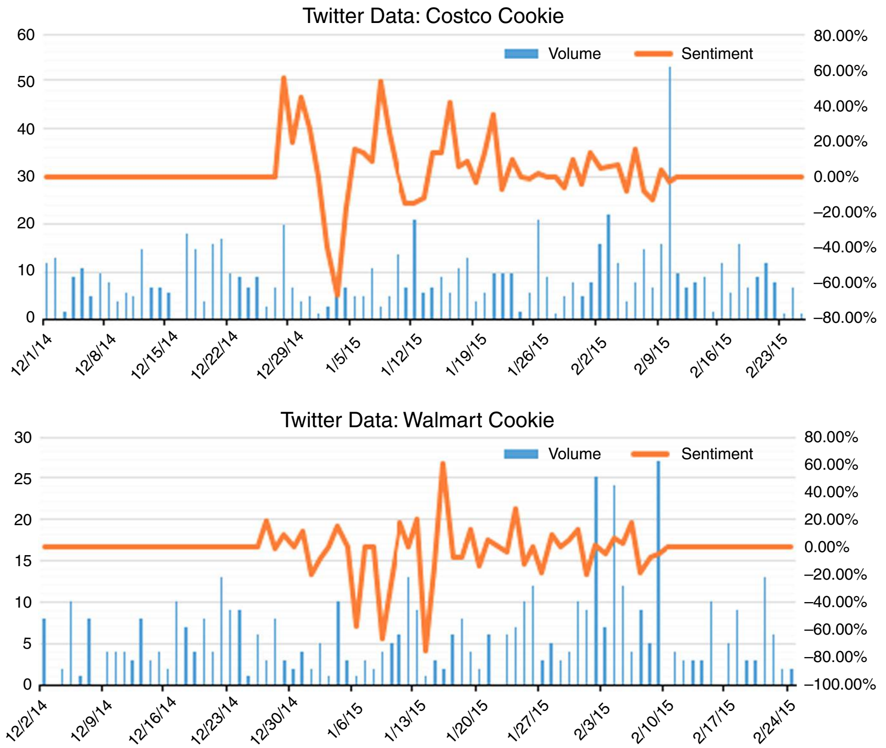

Figure 5.
Twitter data comparision of cookie for Costco and Walmart

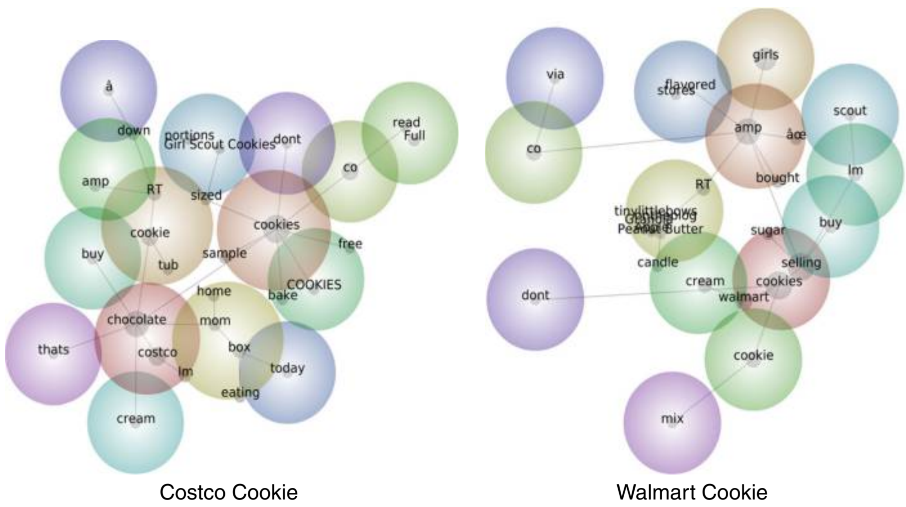

Figure 6.
Generated cluster diagrams for cookie-related Costco and Walmart tweets

generated clusters, we had an in-depth discussion to finalize the clusters. A third researcher was invited to join in the discussion and provided additional input when there were conflicts and disagreement about certain themes and clusters. As a result of the discussion, we merged some of the clusters and sub-clusters based on our knowledge and finalized the clusters manually. As a result, we reduced the final number of the clusters to several major clusters to make the results more meaningful. Due to page length limitation, we used Table IV as an example to list the three major clusters we identified from cookie-related Costco tweets.

To give readers a better idea about the tweets, we listed some sample tweets related to Costco cookie in our data set below:

- Haha they are my Costco cookies that I love, but my Costco stopped carrying them. Hopefully only temporary!
- So I made cookies out of the ghirardelli brownie mix Costco sells. It was a phenomenal idea. You can thank me later.
- Well, time for more @TatesBakeShop chocolate chip cookies. So glad @Costco stocks the big box!
- Costco cookies and a cup of tea is a match made in heaven.
- Costco cookies are ridiculous.
- choice cause of death: overdose on Costco chocolate chip cookies.
- Costco cookies are addicting.
- Costco chocolate chip cookies are dangerous items to have at home within hands reach, but worth it.

Discussion

The results of the case study revealed the value of analyzing social media mentions and conducting sentiment analysis and comparison on individual product level. As the number of social media mentions in our data set is very large, these tweets contain data about numerous products sold by Walmart and Costco, which provides an opportunity for comparison on the product level. As Berman (2013) points out, the most common purpose of large data set is to produce small data and there is almost always a filtering process to reduce large data into smaller data. In this case study, we queried the data set using keyword and focussed our attention to a smaller data set related to four products: muffin, cookie, pizza and chicken to get more meaningful results. For example, we found that although customers mentioned Walmart more often than Costco on Twitter during that period, people tend to talk about Costco's muffin and cookie more than Walmart's muffin and cookie on Twitter.

Our case study shows that analyzing the overall social media mentions and sentiment trend alone is not sufficient. It is also necessary for organizations to look into individual products and identify possible opportunities and issues on individual product level to enhance their competitiveness in different areas. For example, as there is a higher number of positive comments and negative comments on Costco's muffin and cookie

Concept clusters (Costco)	Main content
Discussion on cookie's quality and variety	Ingredient (cream, chocolate, etc.), variety (American, European, etc.)
Discussion on cookie's flavors	Taste good, different flavors, chocolate chips cookies
Sharing feelings	Costco cookies are so good; Costco cookies are the best cookies

Table IV. Main clusters and themes on cookie-related Costco tweets than on Walmart's muffin and cookie, Costco or Walmart could dig into these positive and negative comments and figure out what is going on using methods such as text mining or content analysis depending on the size of the data set. In particular, compared to Walmart, Costco received a higher percentage of negative comments for muffin and a lower percentage of negative comments for cookie. We examined these comments and found that tweets associated with Costco tend to be more detailed with diverse opinions on flavor, ingredient, appearance and variety of the muffin and cookie. By comparing these consumers' opinions on flavor, ingredient, appearance and variety, Costco could discover the underlying causes or insights to address the discrepancy in customer sentiment across products and improve their competitiveness against competitors on the product level. For example, several customers mentioned in their tweets that they were allergic to the chocolate chip in their blueberry muffin and asked what Costco could do. This case study indicates that conducting big social data competitive analytics made it possible to realize new business value from massive amounts of social media data generated by millions of users. Furthermore, organizations often have weekly or monthly sales data, daily stock price data or other business metrics which can be tested to see if there are any correlations with social media mentions and sentiments on both overall and product level. Doing this will bring even more business insights to organizations.

Our work also contains practical contributions by providing a framework for companies to engage in social media competitive intelligence. Managers should be more proactive to learn as much as possible and as soon as possible about customers' opinion on their products and services in general as well as customers' opinion on identical products and services from their competitors. In a competitive market place, up-to-date information can make a big difference between keeping pace, getting ahead or being left behind (Helm, 2011). Thus, managers consider social media competitive intelligence activities as a formal regular business process to empowers themselves to anticipate and face future challenges, enhance their capability and ability to maintain a competitive edge over their competitors (Ross *et al.*, 2012). As the result of competitive intelligence activities, businesses could develop new products and services, optimize business processes, enhance value creation and foster innovations. The increase in business knowledge and innovations will lead to an increase in economic activities which will have significant impact on today's knowledge economy.

In addition to analyzing the captured data-at-rest, the proposed framework and the case study results also indicate that there is a strong need for creating a social media data application that can conduct real-time social media comparative analytics for social media data-in-motion so that organizations can immediately respond to changing events. Figures 1-5 illustrate the dynamic nature of social media data and confirm that customer sentiment changes quite frequently over time. Thus, stream-processing and stream-analysis techniques need to be leveraged by organizations to promptly understand changing sentiments and concepts, gain time-sensitive competitive advantages, understand what is happening in real time and make more efficient decisions than their peers.

Conclusion and future research

Competitive analytics and intelligence has a great potential to produce useful information, actionable knowledge and critical insights for companies to enhance competitiveness and solve business problems. The knowledge-intensive business activities caused by technical advances in competitive analytics and intelligence will generate tangible and intangible business values and contribute to a knowledge-based economy (Amidon *et al.*, 2005). The explosive increase of social media data creates a major opportunity for companies to leverage data analytics solutions to harness customer perceptions and better understand what people are saying on a topic, a product or about a company (Stieglitz and Dang-Xuan, 2013). The proposed framework and the results of the case study suggest a business opportunity for developing a social media data application for competitive analytics and intelligence purpose because companies in general want to know how consumers feel about their products and services and those of their competitors. So far social media analytics companies such as VOZIQ has made very good progress in this area and offers tools to help companies benchmark again peers using social media data-at-rest.

A lot of challenges still exist in the social media analytics field. For example, although sentiment analysis has made good progress, it still has issues with sarcastic and ironic sentences and the sentiment analysis scores may contain many errors (Duan *et al.*, 2013; Pang and Lee, 2004). Many people also write spam reviews on social media to promote their own products by giving undeserving positive opinions, or defame their competitors' products by giving false negative opinions (Vohra and Teraiya, 2013). There are also various types of unwanted and malicious spam messages on social media (Jin *et al.*, 2011) that need to be efficiently and effectively detected and filtered. More future research on sentiment analysis and machine learning methods is needed to address these challenges and improve the quality of social media data mining and analytics.

References

Amidon, D.M., Formica, P. and Mercier-Laurent, E. (2005), *Knowledge Economics: Emerging Principles, Practices and Policies*, Faculty of Economics and Business Administration, University of Tartu, Tartu.

Barbier, G. and Liu, H. (2011), "Data mining in social media", *Social Network Data Analytics*, pp. 327-352, available at: http://link.springer.com/chapter/10.1007/978-1-4419-8462-3_12

Benhardus, J. and Kalita, J. (2013), "Streaming trend detection in Twitter", *International Journal of Web Based Communities*, Vol. 9 No. 1, pp. 122-139.

Berman, J.J. (2013), *Principles of Big Data: Preparing, Sharing, and Analyzing Complex Information*, Newnes, Boston, WA.

Bifet, A. and Frank, E. (2010), "Sentiment knowledge discovery in Twitter streaming data", in Pfahringer, B., Holmes, G. and Hoffmann, A. (Eds), *Discovery Science*, Springer, Berlin and Heidelberg, pp. 1-15.

Bollen, J., Mao, H. and Zeng, X. (2011), "Twitter mood predicts the stock market", *Journal of Computational Science*, Vol. 2 No. 1, pp. 1-8.

Bose, R. (2008), "Competitive intelligence process and tools for intelligence analysis", *Industrial Management & Data Systems*, Vol. 108 No. 4, pp. 510-528.

Chen, H. and Das, S.R. (2010), "Business and market intelligence 2.0", *Part 2, IEEE Intelligent Systems*, Vol. 25 No. 2, pp. 74-78.

Chen, H. and Yang, C.C. (2011), "Special issue on social media analytics: understanding the pulse of the society", *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, Vol. 5 No. 41, pp. 826-827.

Duan, W., Cao, Q., Yu, Y. and Levy, S. (2013), "Mining online user-generated content: using sentiment analysis technique to study hotel service quality", *Proceedings of the 46th Hawaii International Conference on System Sciences*, pp. 3119-3128.

Earle, P.S., Bowden, D.C. and Guy, M. (2012), "Twitter earthquake detection: earthquake monitoring in a social world", *Annals of Geophysics*, Vol. 54 No. 6, available at: www.annalsofgeophysics.eu/index.php/annals/article/view/5364 (accessed March 19, 2015).

Fan, W. and Yan, X. (2015), "I&M special issue on novel applications of social media analytics", *Information & Management*, Vol. 52 No. 7, pp. 761-763. doi:10.1016/j.im.2015.07.007.

Festinger, L. (1954), "A theory of social comparison processes", *Human Relations*, Vol. 7 No. 2, pp. 117-140.

Garcia, S.M., Tor, A. and Schiff, T.M. (2013), "The psychology of competition: a social comparison perspective", *Perspectives on Psychological Science*, Vol. 8 No. 6, pp. 634-650.

Gupta, R., Gupta, H. and Mohania, M. (2012), "Cloud computing and big data analytics: what is new from databases perspective?", in Srinivasa, S. and Bhatnagar, V. (Eds), *Big Data Analytics*, Springer, Berlin and Heidelberg, pp. 42-61.

He, W., Zha, S.H. and Li, L. (2013), "Social media competitive analysis and text mining: a case study in the pizza industry", *International Journal of Information Management*, Vol. 33 No. 3, pp. 464-472.

He, W., Wu, H., Yan, G., Akula, V. and Shen, J. (2015), "A novel social media competitive analytics framework with sentiment benchmarks", *Information & Management*, Vol. 52 No. 7, pp. 801-812. doi:10.1016/j.im.2015.04.006.

Helm, B. (2011), "How to use competitive intelligence to gain an advantage", available at: www.inc.com/magazine/20110401/how-to-use-competitive-intelligence-to-gain-an-advantage.html (accessed August 21, 2015).

Jin, X., Lin, C., Luo, J. and Han, J. (2011), "A data mining-based spam detection system for social media networks", *Proceedings of the VLDB Endowment*, Vol. 4 No. 12, pp. 1458-1461.

Kang, J.S., Kuznetsova, P., Luca, M. and Choi, Y. (2013), "Where not to eat? Improving public policy by predicting hygiene inspections using online reviews", *EMNLP*, pp. 1443-1448.

Lansdall-Welfare, T., Lampos, V. and Cristianini, N. (2012), "Effects of the recession on public mood in the UK", *Proceedings of the 21st International Conference Companion on World Wide Web. ACM*, *April*, pp. 1221-1226.

Liau, B.Y. and Tan, P.P. (2014), "Gaining customer knowledge in low cost airlines through text mining", *Industrial Management & Data Systems*, Vol. 114 No. 9, pp. 1344-1359.

Ngai, E.W., Moon, K.L.K., Lam, S.S., Chin, E.S. and Tao, S.S. (2015), "Social media models, technologies, and applications: an academic review and case study", *Industrial Management & Data Systems*, Vol. 115 No. 5, pp. 769-802.

NYSE (2013), "NYSE Technologies and SMA Announce Agreement to distribute social media analysis data via SFTI", available at: www1.nyse.com/press/1360840642983.html (accessed March 13, 2015).

Pang, B. and Lee, L. (2004), "A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts", *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, *Barcelona*, pp. 271-278.

Pearson, T. and Wegener, R. (2013), "Big data: the organizational challenge", available at: www.bain.com/publications/articles/big_data_the_organizational_challenge.aspx (accessed March 13, 2015).

RemitDATA (2013), "The evolving comparative analytics market: benchmarking key business metrics against peers to reduce risk, pinpoint areas for improvement, and optimize performance", available at: www.mgma.com/Libraries/Assets/Practice%20Resources/Vendors%20and%20Partners/IRC/The-Evolving-Comparative-by-RemitDATA.pdf (accessed March 21, 2015).

Risius, M. and Beck, R. (2015), "Effectiveness of corporate social media activities in increasing relational outcomes", *Information & Management*, doi:10.1016/j.im.2015.06.004.

Ross, P., McGowan, C. and Styger, L. (2012), "A comparison of theory and practice in market intelligence gathering for Australian micro businesses and SMEs", *19th International Business Research Conference: Research for Re-thinking, Social Science Research Network*, pp. 1-17.

Sanderson, M. (2013), "Maximize performance with BI and big data. Comparative analytics enables organizations to benchmark performance against their peers", *Health Management Technology*, Vol. 34 No. 1 p. 18.

Stieglitz, S. and Dang-Xuan, L. (2013), "Social media and political communication: a social media analytics framework", *Social Network Analysis and Mining*, Vol. 3 No. 4, pp. 1277-1291.

Tan, W., Blake, M.B., Saleh, I. and Dustdar, S. (2013), "Social-network-sourced big data analytics", *IEEE Internet Computing*, Vol. 17 No. 5, pp. 62-69.

Vohra, M.S. and Teraiya, J. (2013), "Applications and challenges for sentiment analysis: a survey", *International Journal of Engineering Research and Technology*, Vol. 2 No. 2.

Wang, H. and Wang, W. (2014), "Product weakness finder: an opinion-aware system through sentiment analysis", *Industrial Management & Data Systems*, Vol. 114 No. 8, pp. 1301-1320.

Zeng, D., Chen, H., Lusch, R. and Li, S.H. (2010), "Social media analytics and intelligence", *IEEE Intell. Syst.*, Vol. 25 No. 6, pp. 13-16.

Zikopoulos, P., Parasuraman, K., Deutsch, T., Giles, J. and Corrigan, D. (2012), *Harness the Power of Big Data the IBM Big Data Platform*, McGraw Hill Professional, New York, NY.

About the authors

Dr Wu He is an Assistant Professor in the Department of Information Technology and Decision Sciences at the Old Dominion University. Wu He received his PhD at the University of Missouri. His research interests include cyber security, data mining, social media, knowledge management and e-learning technologies. He has received research funding from the National Science Foundation. He has published more than 60 articles in peer-reviewed journals such as *the Journal of the Association for Information Science and Technology*, *Information & Management*, *International Journal of Information Management*, *IEEE Transactions on Industrial Informatics*, *Internet Research*, *Information Technology and Management*, and *Expert Systems with Applications*. Dr Wu He is the corresponding author and can be contacted at: whe@odu.edu

Jiancheng Shen is currently a PhD Candidate in Finance at the Old Dominion University, USA. His research interests include neuro-finance, media and financial markets, high-frequency financial data analysis and international economics.

Xin Tian is currently a PhD Candidate in Information Technology at the Old Dominion University, USA. Her research interests include cyber security, social media and database administration.

Dr Yaohang Li is an Associate Professor in the Department of Computer Science at the Old Dominion University. His research interests are in computational biology and scientific computing. He received the PhD and MS Degrees in Computer Science from the Florida State University in 2003 and 2000, respectively. After graduation, he worked at the Oak Ridge National Laboratory as a Research Associate for a short period of time. Before joining ODU, he was an Associate Professor in the Computer Science Department at the North Carolina A&T State University.

Dr Vasudeva Akula is the Co-founder and Principal Business Consultant at the VOZIQ where he creates and manages strategy for clients in establishing voice of customer-based business performance improvement programs. In his 20 year career spanning banking and financial services, insurance, retail, technology, telecom and wireless, utility industries, he played key roles in helping Fortune 500 companies successfully deploy and realize significant ROI out of customer intelligence solutions.

Dr Gongjun Yan received his PhD in Computer Science from the Old Dominion University. He is currently an Assistant Professor in the University of Southern Indiana. His main research areas include algorithms, security, privacy, routing and health care in internet, vehicular ad-hoc networks, sensor networks and wireless communication. In years, Dr Yan applies mathematical analysis to model behavior of complex systems and integrates existing techniques to provide comprehensive solutions.

Professor Ran Tao is a Faculty Member at the School of Computer Science & Technology in the Donghua University, Shanghai, China. He is a Visiting Scholar in the Department of Information Technology and Decision Sciences at the Old Dominion University from 2014 to 2015. His research interests include cloud computing, data mining and big data.

This article has been cited by:

1. Ting-Peng Liang, Yu-Hsi Liu. 2018. Research Landscape of Business Intelligence and Big Data analytics: A bibliometrics study. *Expert Systems with Applications* **111**, 2-10. [Crossref]
2. HsuLi-Chun, Li-Chun Hsu, WangKai-Yu, Kai-Yu Wang, ChihWen-Hai, Wen-Hai Chih. 2018. Investigating virtual community participation and promotion from a social influence perspective. *Industrial Management & Data Systems* **118**:6, 1229-1250. [Abstract] [Full Text] [PDF]
3. Anant Saxena, K. R. Chaturvedi, Sapna Rakesh. 2018. Analysing Customers Reactions on Social Media Promotional Campaigns. *Paradigm* **22**:1, 80-99. [Crossref]
4. Mark van Rijmenam, Tatiana Erekhinskaya, Jochen Schweitzer, Mary-Anne Williams. 2018. Avoid being the Turkey: How big data analytics changes the game of strategy in times of ambiguity and uncertainty. *Long Range Planning* . [Crossref]
5. Anna Raffoni, Franco Visani, Monica Bartolini, Riccardo Silvi. 2018. Business Performance Analytics: exploring the potential for Performance Management Systems. *Production Planning & Control* **29**:1, 51-67. [Crossref]
6. Pekka Pääkkönen, Juha Jokitulppo. 2017. Quality management architecture for social media data. *Journal of Big Data* **4**:1. . [Crossref]
7. Roberta Akemi Sinoara, Jo?o Antunes, Solange Oliveira Rezende. 2017. Text mining and semantics: a systematic mapping study. *Journal of the Brazilian Computer Society* **23**:1. . [Crossref]
8. HeWu, Wu He, TianXin, Xin Tian, TaoRan, Ran Tao, ZhangWeidong, Weidong Zhang, YanGongjun, Gongjun Yan, AkulaVasudeva, Vasudeva Akula. 2017. Application of social media analytics: a case of analyzing online hotel reviews. *Online Information Review* **41**:7, 921-935. [Abstract] [Full Text] [PDF]
9. Yixiu Yu, Xu Li, Canhua Qiu. 2017. Understanding Chinese Domestic Travelers' Trip Patterns to Mount Fanjing. *Journal of China Tourism Research* **8**, 1-20. [Crossref]
10. Wu He, Xin Tian, Andy Hung, Vasudeva Akula, Weidong Zhang. 2017. Measuring and comparing service quality metrics through social media analytics: a case study. *Information Systems and e-Business Management* **55**. . [Crossref]
11. Daniel Pérez-González, Sara Trigueros-Preciado, Simona Popa. 2017. Social Media Technologies' Use for the Competitive Information and Knowledge Sharing, and Its Effects on Industrial SMEs' Innovation. *Information Systems Management* **34**:3, 291-301. [Crossref]
12. Antonio Teixeira, Raul M. S. Laureano. Data extraction and preparation to perform a sentiment analysis using open source tools: The example of a Facebook fashion brand page 1-6. [Crossref]
13. WangHongwei, Hongwei Wang, GaoSong, Song Gao, YinPei, Pei Yin, LiuJames Nga-Kwok, James Nga-Kwok Liu. 2017. Competitiveness analysis through comparative relation mining. *Industrial Management & Data Systems* **117**:4, 672-687. [Abstract] [Full Text] [PDF]
14. YuYixiu, Yixiu Yu, LiXu, Xu Li, JaiTun-Min (Catherine), Tun-Min (Catherine) Jai. 2017. The impact of green experience on customer satisfaction: evidence from TripAdvisor. *International Journal of Contemporary Hospitality Management* **29**:5, 1340-1361. [Abstract] [Full Text] [PDF]
15. HeWu, Wu He, WangFeng-Kwei, Feng-Kwei Wang, AkulaVasudeva, Vasudeva Akula. 2017. Managing extracted knowledge from big social media data for business decision making. *Journal of Knowledge Management* **21**:2, 275-294. [Abstract] [Full Text] [PDF]

16. Wessam Elhenfnawy, Jesse Wright, Kumara Kallepalli, Kevin Racheal, Ajay Gupta, Rohit Parimi, Parth Shah, Yaohang Li. What Differentiates News Articles with Short and Long Shelf Lives? A Case Study on News Articles at Bloomberg.com 131-136. [Crossref]
17. Ketonen-OksiSanna, Sanna Ketonen-Oksi, JussilaJari J., Jari J. Jussila, KärkkäinenHannu, Hannu Kärkkäinen. 2016. Social media based value creation and business models. *Industrial Management & Data Systems* **116**:8, 1820-1838. [Abstract] [Full Text] [PDF]