ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org

J Intell Inform Syst 2014 December: 20(4): 89~105 http://dx.doi.org/10.13088/jiis.2014.20.4.89

Visualizing the Results of Opinion Mining from Social Media Contents: Case Study of a Noodle Company

Yoosin Kim
College of Business, University of Texas at Arlington
(E-mail: yoosink@uta.edu)

Do Young Kwon
Graduate School of Business IT, Kookmin University
(E-mail: doyoung@kookmin.ac.kr)

Seung Ryul Jeong
Graduate School of Business IT, Kookmin University
(E-mail: srjeong@kookmin.ac.kr)

After emergence of Internet, social media with highly interactive Web 2.0 applications has provided very user friendly means for consumers and companies to communicate with each other. Users have routinely published contents involving their opinions and interests in social media such as blogs, forums, chatting rooms, and discussion boards, and the contents are released real-time in the Internet. For that reason, many researchers and marketers regard social media contents as the source of information for business analytics to develop business insights, and many studies have reported results on mining business intelligence from Social media content. In particular, opinion mining and sentiment analysis, as a technique to extract, classify, understand, and assess the opinions implicit in text contents, are frequently applied into social media content analysis because it emphasizes determining sentiment polarity and extracting authors' opinions. A number of frameworks, methods, techniques and tools have been presented by these researchers. However, we have found some weaknesses from their methods which are often technically complicated and are not sufficiently user-friendly for helping business decisions and planning.

In this study, we attempted to formulate a more comprehensive and practical approach to conduct opinion mining with visual deliverables. First, we described the entire cycle of practical opinion mining using Social media content from the initial data gathering stage to the final presentation session. Our proposed approach to opinion mining consists of four phases: collecting, qualifying, analyzing, and visualizing. In the first phase, analysts have to choose target social media. Each target media requires different ways for analysts to gain access. There are open-API, searching tools, DB2DB interface, purchasing contents, and so son. Second phase is pre-processing to generate useful materials for meaningful analysis. If we do not remove garbage data, results of social media analysis will not provide meaningful and useful business insights. To clean social media data, natural language processing techniques should be applied. The next step is the opinion mining phase where the cleansed social media content set is to be analyzed. The qualified data set includes not only user-generated contents but also content identification information such as creation date, author name, user id, content id, hit counts, review or reply, favorite, etc. Depending on the purpose of the analysis, researchers or data analysts can select a suitable mining tool. Topic extraction and buzz analysis are usually related to market trends analysis, while sentiment analysis is utilized to conduct reputation analysis. There are also various applications, such as stock prediction, product recommendation, sales forecasting, and so on. The last phase is visualization and presentation of analysis results. The major focus and purpose of this phase are to explain results of analysis and help users to comprehend its meaning. Therefore, to the extent possible, deliverables from this phase should be made simple, clear and easy to understand, rather than complex and flashy.

To illustrate our approach, we conducted a case study on a leading Korean instant noodle company. We targeted the leading company, NS Food, with 66.5% of market share; the firm has kept No. 1 position in the Korean “Ramen” business for several decades. We collected a total of 11,869 pieces of contents including blogs, forum contents and news articles. After collecting social media content data, we generated instant noodle business specific language resources for data manipulation and analysis using natural language processing. In addition, we tried to classify contents in more detail categories such as marketing features, environment, reputation, etc. In those phase, we used free ware software programs such as TM, KoNLP, ggplot2 and plyr packages in R project.

As the result, we presented several useful visualization outputs like domain specific lexicons, volume and sentiment graphs, topic word cloud, heat maps, valence tree map, and other visualized images to provide vivid, full-colored examples using open library software packages of the R project. Business actors can quickly detect areas by a swift glance that are weak, strong, positive, negative, quiet or loud. Heat map is able to explain movement of sentiment or volume in categories and time matrix which shows density of color on time periods. Valence tree map, one of the most comprehensive and holistic visualization models, should be very helpful for analysts and decision makers to quickly understand the “big picture” business situation with a hierarchical structure since tree-map can present buzz volume and sentiment with a visualized result in a certain period.

This case study offers real-world business insights from market sensing which would demonstrate to practical-minded business users how they can use these types of results for timely decision making in response to on-going changes in the market. We believe our approach can provide practical and reliable guide to opinion mining with visualized results that are immediately useful, not just in food industry but in other industries as well.

Keyword : Opinion Mining, Visualization, Social Media, Market Intelligence, Food Industry

Received : November 7, 2014 Revised : December 2, 2014 Accepted : December 3, 2014
Type of Submission : Fast Track Corresponding Author : Seung Ryul Jeong

1. Introduction

After emergence of Internet, Social media content, often called user-generate contents (UGC) or word of mouth (WOM), are now routinely published and made available on the Internet. Social media, as highly interactive Web 2.0 applications, including micro-blogs, blogs, forums, chatting rooms, and discussion boards, has provided very user friendly means for consumers and companiesto communicate with each other (Chau and Xu, 2012; Chen, 2010). Blogs, for example, contain countless personal stories and opinions, which are richly embedded with feelings, sentiments, and emotions. Bloggers also are linked to each other in social media platform, and share their interests and opinions. Customer-generated contents on various forums, blogs, and newsgroups offer corporate decision makers and analysts opportunities to listen to the voice of the market from business actors such as consumers, employees, investors, and mass-media (Chen, Chiang and Storey, 2012; Kim and Jeoung, 2013; Lusch et al., 2010; Mangold and Faulds, 2009). Recently, huge amount of Social media content from various blogosphere, portal websites and social networking sites, have been treated as unstructured text big-data, and subject to various analytics techniques and tools like opinion mining, sentiment analysis,topic and buzz analysis, and recommender system for the purpose of extracting business intelligence (Chen et al., 2012; Cruz and Lee, 2014; Kim et al., 2012; Liu et al., 2010; Pang and Lee, 2008).

Prior to the advent of big data analytics, some researchers had attempted to analyze the relevance of WOM materials for marketing and sales by examining Social media content like customer review comments and ratings (Chevalier and Mayzlin, 2006; Liu, 2006). Upon the emergence of the big data phenomenon from sources like tweets, blogs, reviews, comments and board messages, researchers have tried to analyze these user-generated contents in various industries, including book publishing, movies, music albums, hotels, and restaurants (Dhar and Chang, 2009; Jin et al., 2014; Rui et al., 2013; Ye et al., 2011; Zhang et al., 2010). To improve the relevance and perceived usefulness of analysis results, a few researchers have ventured further with the visualization techniques to synthesize the sentiment analysis result for end-users such as journalists and customers (Diakopoulos, Naaman, and Kivran-Swaine, 2010; Duanet al., 2012; Wu et al., 2010).

These earlier studies have advanced the state of art in opinion mining for extracting business intelligence from social media content. However, we recognize needs for improvement in a number of areas. First, researchers have put forward method for opinion mining to extract market intelligence, but there is a lack of easy-to-follow road map that provides practical guide for conducting opinion mining analytics (Chen, 2010; Liu et al., 2010). In this study, we will present a step-to-step life-cycle approach to fill this gap. Second,we need to go beyond analysis methods with more technical sophistication and greater accuracy (Chau and Xu, 2012; Chen, 2010; Kim et al., 2014; Liu et al., 2010; Rui et al., 2013) and pay more attention to practical use of analysis outputs for decision making in real-world business settings. Methods for visualization of outputs are helpful, but overly ornate and complicated display of outputs with a large number of variables, such as year, month, group, category, sentiment, volume, etc may confuse and bewilder the users. In this study, we will demonstrate that the output of opinion mining can indeed be presented in an intuitive way with visualization techniques that are easy for average users to comprehend and utilize.

To overcome the above weaknesses and fill gaps in current research, this study proposed a comprehensive and practical approach to opinion mining from Social media content. This paper thus has three objectives. First, we seek to describe the entire cycle of practical opinion mining using Social media content from the initial data gathering stage to the final presentation session. Secondly, we presented several useful outputs like domain specific lexicons, volume and sentiment graphs, topic word cloud, heat maps, valence tree map, and other visualized images to provide vivid, full-colored examples using open library software packages of the R project, so potential users can consider these tools and techniques for immediate adoption. Finally, we illustrated our method with a case study on a leading Korean instant noodle company. This case offersreal-world business insights from market sensing which would demonstrate to practical-minded business users how they can use these types of results for timely decision making in response to on-going changes in the market.

2. Related precious work

According to various researcher and practitioner sources (Kaplan and Haenlein, 2010; Mangold and Faulds, 2009), the concept of social media may be described as a group of online communication channels for publishing, delivering, interacting, and sharing their interest, opinion, sentiment, emotion, and personal thoughts, using Internet-based applications that build on the ideological and technological foundations of Web 2.0. These Web 2.0 tools include micro-blogs, blogs, chat rooms, forums, news boards, etc. that carry tweets, chats, comments, reviews, news articles and other online contents. These online communication channels include not only user generated contents like private messages and memo in personal networking using blogs, forums, and chatters, but also word of mouth such as consumer reviews and comments communicated between customers and companies in online communities (Mangold and Faulds, 2009).

With the emergence of big data, the era of advanced business intelligence and analytics has suddenly descended on many organizations, including for-profit businesses, governments, nonprofit entities, and healthcare organizations, many of whom have been very proactive in applying advanced tools and techniques to conduct domain-specific analytics (Chen et al., 2012; Chen and Zimbra, 2010; Lusch et al., 2010). Social media content analysis, in particular, has been a prominent application of big data analytics in business domains such as e-commerce, health care, banking, entertainment, etc. These analytics efforts can help a company to obtain first-hand knowledge on market reception of its products and services, and even those of the competitors', through torrents of customer feedbacks in on-line Social media content. This knowledge should enables business analysts and decision makers to develop insights on consumer opinions,discover new ideas for marketing, improve customer satisfaction, and ultimately increase returns on business investments (Chau and Xu, 2012; Chen et al., 2012).

Huge amount of Social media content are generated from various blogosphere, portal websites and social networking site, and opinion mining using those unstructured text data is widely used in research and market intelligence. Because these contents have the potential to influence product sales and analyzing them can help to fine tune market strategy through better and more accurate measures of customer responses (Chen and Zimbra, 2010; Liu et al., 2010; Lusch et al., 2010). A study demonstrated the potential of opinion mining for extracting market intelligence from Wal-Mart's Social media content through visualized graphics showing message traffic patterns and averages of sentiment scores, along with tables indicating top five active authors as output (Chen and Zimbra, 2010). Another study presented a visual comparison to summarize customers' opinions on various cell-phone features, such as picture, battery, camera, size and weight, from competing brands (Liu et al., 2010). Several studies reported results from analyzing social media data from tweets, blogs, news article, customer review comments, and favorite ratings, covering a wide variety of industries and domains such as hotel reservations, restaurant recommendations, books, music CD, and movies (Dhar and Chang, 2009; Liu, 2006; Rui et al., 2013; Ye et al., 2011; Zhang et al., 2010).

In addition to reporting results, a few published studies have paid specific attention to effective communication and presentation of big- data opinion mining analytics output via visualization. These researchers suggested several techniques and systems: a visualization tool to help journalists and media professionals extracting value from news article (Diakopoulos et al., 2010), a system using visual metaphors to provide an integrated view of multiple correlations to find opinion patterns and visual comparison of user groups' feedback (Wu et al., 2010), and a visualizing system synthesizing the analysis results and presenting it to the customer experience (Duan et al., 2012). Another paper proposed visualization framework, TexVizu, for common elements between text analysis and visualization, and conducted the framework into a case study visualizing public sentiment about politicians in tweets and news (Choi et al., 2014).

Even though many studies provided very helpful methods, frameworks, and techniques for visualization of text mining, part of them are often ornate and complicated display of outputs with too much information such as year, month, group, category, sentiment, volume, etc to understand those visualized deliverables. For that reason, we will propose a comprehensive and practical approach to opinion mining from Social media content and demonstrate outputs of opinion mining through real business case study.

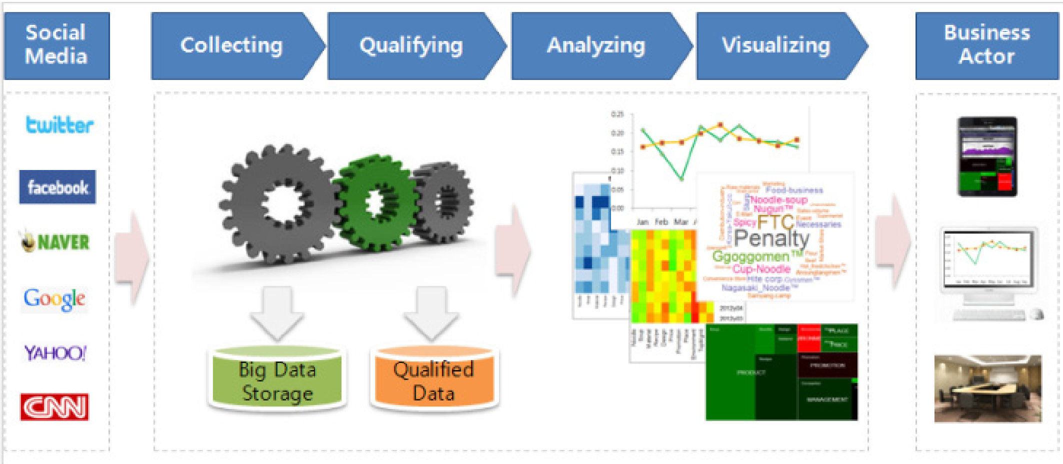
3. Proposed Approach

As mentioned previously, opinion mining as a technique for social media content analysis is deployed to extract, classify, understand, and assess the opinions implicit in text contents like social media data: reviews, tweets, blogs, messages and news articles. Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and emotional states towards entities, events, and their attributes in text content (Chen et al., 2012; Pang and Lee, 2008).

3.1 Phases of the Opinion Mining Process

Our proposed approach to opinion mining consists of the following four phases as shown Figure 1: collecting, qualifying, analyzing, and visualizing.

In the first phase, analysts have to choose target social media. Different target media requires different ways for analysts to gain access. Social network services (SNS) such as Twitter and Facebook offer open-API for accessing and



〈Figure 1〉 Overview of Process Stages

extracting data. For portal websites, blogs and forums that do not provide open-API, tools such as web-crawler, web-scrapper or search robot software program can be utilized. In addition, data may be gathered by developing DB2DB interface or purchasing contents from social media data provider. This data access phase will give rise to local storage of an enormous amount of unstructured text big-data that needs to be cleansed.

These unstructured big-data text files must go through pre-processing to generate useful materials for meaningful analysis. Raw tweets data from Twitter, for example, sometimes has filled with noises in large part of its content. If we do not remove such garbage data, results of social media analysis will not provide meaningful and useful business insights. Therefore, in the second phase, a rigorous data qualifying procedure must be carried out. To clean social media data in unstructured text, natural language processing (NLP) techniques should be applied. Here, NLP involves activities for word parsing, disabled characters removing; emoticons, html tags, punctuations, stop words elimination and feature tagging, etc. Qualified data through the cleansing process is transformedto analysis data format, and ready for opinion mining analytics like topic extracting, content categorizing, sentiment analysis, and opinion forecasting.

The next step is the opinion mining phase where the cleansed social media content set is to be analyzed. The qualified data set includes not only user-generated contents but also content identification information such as creation date, author name, user id, content id, hit counts, review or reply, favorite, etc. Depending on the purpose of the analysis, researchers or data analysts can select a suitable mining tool. Topic extraction and buzz analysis are usually related to market trends analysis, while sentiment analysis is utilized to conduct reputation analysis. Several techniques and algorithms, such as machine learningmethods and lexicon based sentiment analysis, could be deployed jointly to demonstrate stock prediction, product recommendation, sales forecasting, and other applications. In lexicon based approach, domain specific language resources have been generated and operated, one can expect greater analysis (Kim et al., 2014; Rao, Lei, Wenyin, Li, and Chen, 2013).

The last and concluding phase of the social media content analysis is visualization and presentation of analysis results. The major focus and purpose of this phase is to explain results of analysis and help users to comprehend its meaning and able to use it to support decision making. Therefore, to the extent possible, deliverables from this phase should be made simple, clear and easy to understand, rather than complex and "flashy". For example, when a tag cloud is used in issue and topic analysis output, it conveys the topics' volume with an intuitive visible font color and size. Sentiment heat-map is great in presenting customers' opinion status (positive or negative), using density degree of contrast with 2 colors. Tree map revealing both volume and sentiment in a hierarchical categorization is one of the most straightforward format for presenting analytics deliverables.

3.2 Data Collection

To illustrate our proposed phases for conducting opinion mining, we collected case study data from instant food company. Market size of the instant noodle "Ramen" business in Korea was over $ 2 billion in 2013. The leading firm, NS Food, with 66.5% of market share (FoodJournal, 2013), has kept No. 1 position in the Korean "Ramen" business for several decades. N-Ramen has been a representative product of the company for over 10 years. As shown in Table 1, we collected a total of 11,869 pieces of contents including blogs, forum contents and news articles.

〈Table 1〉 Collected Data Set

9 Months	Blog	News	Total
NS Food N-Ramen	10,591 (89%)	1,278 (11%)	11,869

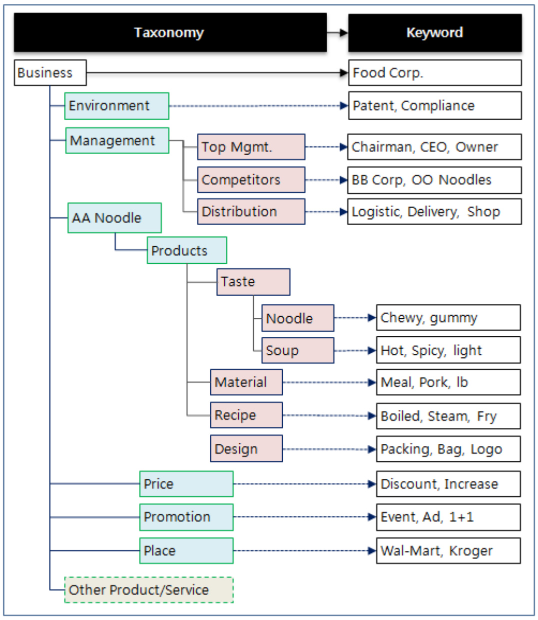
Social media data for the experiment were collected from portal websites Naver.com, Daum.com, and other related sites in South Korea using a web-crawler software program. Searches with this web-search robot are done by using the ramen product name as unique keywords. The data were user-generated contents like blogs, café (forum) messages, and news articles released in target portal websites between January and September in the year 2012. Our web-crawler queried the web-portal for extracting the data related to not only the prime content, but also writer's name or user ID, source site, create-date, and URL address.

3.3 Data Cleansing and Categorizing

After collecting social media content data, we generated instant noodle business specific language resources for data manipulation and analysis. We identified several sets of words for stop-words, sentiment dictionary, and content categorization. These tasks are made more difficult since the Korean language, unlike English, is an agglutinative language with very complicated structure, and doesn't have standard word stems for sentiment. Table 2 shows sample words of instant noodle domain sentiment dictionary from social contents. For this domain specific lexicon, we selected 2,000 sample contents and classified them with positive, neutral, and negative sentiment in a manual way. After classification, NLP was conducted for sentiment words within each group and sentiment were tagged on extracted words by higher frequency. In addition, alarming words, which are recognized as the extremely risk words regardless frequency, are related to company reputation and deserve management attention, because customers are much sensitive and react to those terms such as worm, poison, and cancer-causing, are to customer feedback in food industry.

〈Table 2〉 Sample of Sentiment words

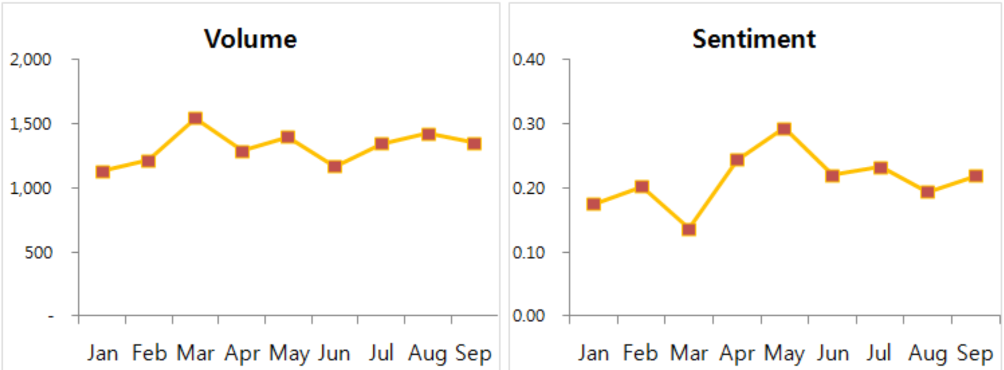
Positive	Negative	Alarming
Cool	Bad	Accusation
Delicious	Complain	Benzopyrene
Excellent	Freshness	Boycott
Good	Hate	Cancer-causing
Happy	Sick-of	Carcinogen
Like	Poor	Poisoning
Love	Tasteless	vermin
Sweet	Terrible	Worm



〈Figure 2〉 Food Domain Specific Lexicon

We tried to classify contents in more detail categories such as marketing features, environment, reputation, etc. Categorizing provides a frame to look more closely at social media volume and sentiment status with real business perspectives. Figure 2 shows part of the lexicons we identified for the instant noodle business domain to categorize the contents. Categorizing provides a frame to look more closely at social media volume and sentiment status with real business perspectives. In this study, marketing 4 P (Product, Price, Promotion, and Place) and internal/external management environment were applied for the frame.

For this phase, we upgraded the language resources based on the research method for stop-words and sentiment dictionary (Kim et al.,



〈Figure 3〉 Movement of Volume and Sentiment

2014; Yu et al., 2013), and selected words with high frequency from collected data. To accomplish this, TM, KoNLP, ggplot2 and plyr packages in R project were utilized. R is a free open source programming language for statistical computing and graphics.

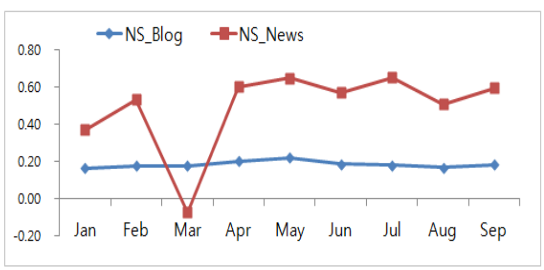
4. Analysis and visualization

In this section, we introduce visualized deliverables through opinion mining on the firm's Social media content. The first set of deliverables, as displayed in Figure 3, consists of two relatively simple graphs showing the changes in volume and averaged sentiment scores over time.

In Figure 3, the range for sentiment scores is between +1 (extremely positive) and -1 (extremely negative). NS food's sentiment scores are always over 0 (neutral sentiment), and it means customers' opinion is relatively positive.

While Figure 3 shows aggregated volume and sentiment scores, Figure 4 present the scores from different social media sources. At first, we can see that consumer sentiment from user-generated contents like blogs and forums shows a very stable pattern with no noticeable changes over time. The sentiment from blogs remains positive and stable around a score of 0.2. In contrast, the patterns from news articles show a sharp.

Interestingly, the Fair Trade Commission (FTC) in Korea imposed a 100 million USD fine in that month as a penalty for price collusion among noodle companies. This has made public sentiment to take a drastic negative turn. However, our analysis revealed that online customer sentiment



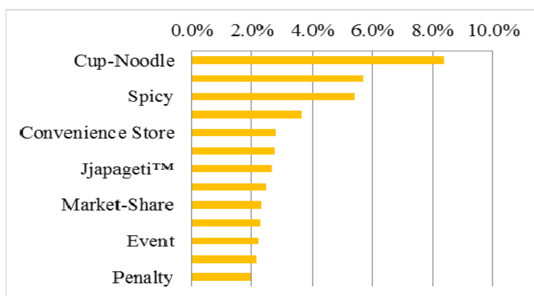
〈Figure 4〉 Movement of Sentiment by Media on blogs toward NS food's noodle product remains positive in March and at the same level as the previous month. It could be interpreted as customer loyalty of N-Ramen is very robust. Although, March was a worst month for noodle companies due to FTC's penalty ruling, and our sentiment analysis clearly portrays this situation. Sentiment scores from news content plummeted to an all-time low, and scores from user-generated contents saw a drop too. Excluding this worst event, sentiment scores from all data sources were in positive territories, with results from news content in the high positive area.

A word cloud (or tag cloud), which is presented in Figure 5, communicates the above situation but from a different angle. Hot issue and topic keywords in the figure are extracted with high frequency during a targeted period. Size and color of the words in the cloud reflect volume of a topic. We can obtain some insights from this word cloud. For example, Figure 5 indicates that the biggest issues about NS food were "Cup-Noodle" (same product of different shape), and "Spicy" (its representative taste). Other big topics were new products, such as Ggoggomen and Nagasaki Noodle, since these new noodles were big hits and were setting new trend in the noodle market. Negative topics like FTC and Penalty were relatively small and obscured by positive sentiment topics.

While the tag cloud intuitively reveals difference in perception between topics, Figure 6 shows the precise frequency ratios for the various topic keywords. We can see that negative topics like FTC and Penalty actually has frequencies below those for positive tags such as spicy taste of N-Ramen and hot items released recently.

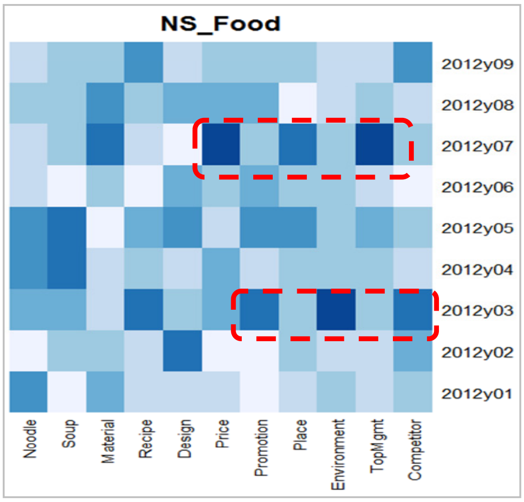
Next, we present a heat map grid, as shown in Figure 7, which display density of social media content traffic. The density of the blue color in the heat map indicates the density of volume traffic. Therefore, movement of volume can be easily recognized by change of color between months and categories. For example, the price category in Julywas colored in heavy blue, and much darker

〈Figure 5〉 Tag Cloud



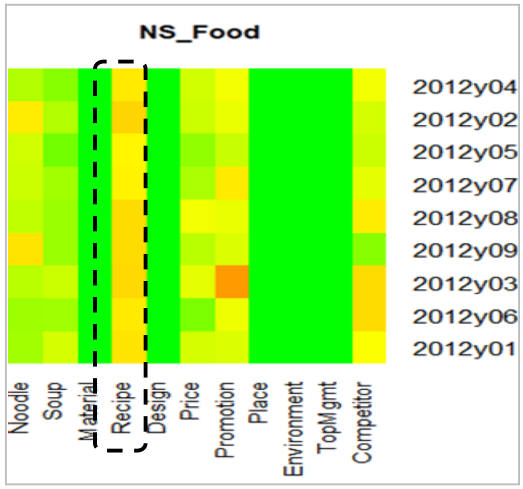
〈Figure 6〉 FreqRatio Bar

than other cells, and this indicates very heavy traffic for discussion on price. In fact, the amount of social media content in July related to price was twice the amount compared to other periods. This is because NS food raised the price of N-Ramen in that period.



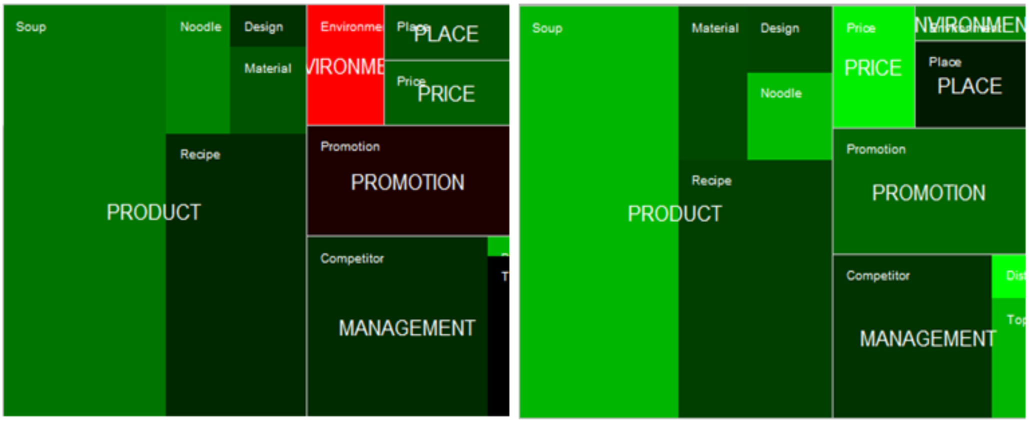
〈Figure 7〉 Volume Density Heat map

In Figure 8 below, sentiment statuses are displayed with hot or cold colors in a category and time matrix. In this grid, three Colors (Red, Yellow, and Green) are used to present differences between negative and positive sentiments. The red color indicates negative sentiment involving pessimistic views; and the green color reflects positive sentiment. If the color of a cell in the grid is closer to the green color, it means that the cell has more positive than negative sentiment. On the other hand, if the color is nearly dark red, the cell contains mostly negative sentiment. As seen in Figure 8, most cells are green or close to green, indicating that consumer-intimate reputation for NS food is very good and this is reflected by strongly positive sentiment.



〈Figure 8〉 Sentiment Density Heat map

While heat maps showed density of either volume or sentiment by categories over a series of time periods, tree maps can be used to presents both volume and sentiment at the same time. Heat map is able to explain movement of sentiment or volume in categories and time but limited to report both of them in one figure. Decision makers would like to see if they can afford to ignore minor opinion even though it is much negative sentiment. Valence tree map, one of the most comprehensive and holistic visualization modes, should be very helpful for analysts and decision makers to quickly understand the "big picture" business situation with a hierarchical structure since tree-map can present buzz volume and sentiment with a visualized result in a certain period. A swift glance can quickly



〈Figure 9〉 Valence Tree Map on March and June

detect areas that are weak, strong, positive, negative, quiet or loud. For generating tree-map, four types of information are needed; main-categories, sub-categories under main-categories, volume and sentiment score of each sub-category. Both of main and sub categories determine the size of the block, and sentiment score presents their sentiment with three colors of positive (e.g., green), negative (e.g., red), and neutral (black). There are several packages such as Rcpp and tremap in R project for generating a valence treemap figure.

As revealed in Figure 9, NS food in March faced a very negative sentiment by buzz related to "Penalty", "Unfair", and "Fine" in management and environment categories, and its reputation was adversely effected. But there were still very positive contents in the product categories. Several months later, crisis is gone and social media sentiment from the negative event becomes calm. The biggest interest toward instant noodle product is soup taste, as can be seen in June tree map.

5. Conclusion

Social media, as highly interactive Web 2.0 applications, including micro-blogs, blogs, forums, chatting rooms, and discussion boards, has rapidly become an important new type of channels for communicationbetween customers and companies. From social media such as various blogosphere, portal websites and social networking site, huge amount of contents labeled as social media content, social data, or social media data are generated by consumers and can be tapped as unstructured text big data. Social media content involving customers' opinions and interests can be effectively utilized to generate business analytics that help managers to develop business insights, and many studies have reported results on mining business intelligence from Social media content.

A number of frameworks, methods, techniques and tools have been presented by researchers. However,these approaches are often technically complicated and are not sufficiently user-friendly for helping business decisions and planning. To overcome these limits, we attempted to formulate a more comprehensive and practical approach to conduct opinion mining with visible deliverables. We applied our approach to a case study on a leading instant noodle company in Korea.

We present graphical tools and visualized output associated with the various phases of the opinion mining process. These include business domain specific lexicons, volume and sentiment graphs, topic and issue word clouds, volume density heat map, sentiment density heat map, and valence tree map with hierarchical structure. Our entire resources are from public-domain Social media content, are collected using a web-crawler software program, and analyzed by the free software programming language Rfor NLP, statistical analytics and graphics. These resources and tools can be easily adopted by businesses for their market intelligence operations.

Our study has several implications and contributions. First, our proposed method is practical and comprehensive, since our approach covers the entire cycle of opinion mining activities, from the initial data collection to the final result visualization of Social media content. Second, the various visualization outputs we presented with real Social media content, may serve as live examples for potential adopters of our approach to launch opinion mining analytics in their own business environment. In addition, using open source software R is economical and cost effective. Third, we believe our approach can provide practical and reliable guide to opinion mining with visualized results that are immediately useful, not just in food industry but in other industries as well.

Nevertheless, there are several limitations to our study. Our research tapped Social media content like blogs, café (forum) messages and news article searched in blogosphere, but did not include micro-blog like Tweets and social networking sites like Facebook. Although these social networking oriented services have many garbage data and private chats, opinions involved in crowd buzz can be helpful to quickly catch the hottest topic and identify the source of "big mouths". Another limitation is to study just one company of major companies in the instantnoodle industry. If all major noodle manufacturers are analyzed, business users can expect competitive intelligence of ramen business. Another improvement is to apply our approach to other domains such as health-care, entertainment, finance, and education industry. Since each business has its unique culture and structure, analysis results and interpretations can vary within the domain.

References

Chau, M. and J. Xu, "Business Intelligence in Blogs: Understanding Consumer Interactions and Communities," *MIS QUARTERLY*, Vol.36, No.4(2012), 1189~1216.

Chen, H. and D. Zimbra, "AI and Opinion Mining," *IEEE Intelligent System*, Vol.25, No.3(2010), 74~80.

Chen, H., "Business and Market Intelligence 2.0, Part 2," *IEEE Intelligent System*, Vol.25, No.1(2010), 2~5.

Chen, H., R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data To Big Impact," *MIS QUARTERLY*, Vol.36, No.4(2012), 1165~1188.

Chevalier, J. A. and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, Vol.43, No.3(2006), 345~354.

Choi, K. S., K. S. Jeong, and S. D. Kim, "A MVC Framework for Visualizing Text Data", *Journal of Intelligence and Information Systems*, Vol.20, No.2(2014), 39~58.

Cruz, R. A. B. and H. J. Lee,"The Brand Personality Effect." *Journal of Intelligence and Information Systems*, Vol.20, No.1(2014), 67~101.

Dhar, V. and E. A. Chang, "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing*, Vol.23, No.4(2009), 300~307.

Diakopoulos, N., M. Naaman, and F. Kivran-Swaine, "Diamonds in the rough: Social media visual analytics for journalistic inquiry," *Proceedings of the 2010 IEEE Symposium on Visual Analytics Science and Technology*, (2010),115~122.

Duan, D., W. Qian, S. Pan, L. Shi, and C. Lin, "VISA: a Visual Sentiment Analysis System," *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction - VINCI '12*, (2012), 22~28.

FoodJournal, The Korean Food Distribution Yearbook 2013. 2013.

Jin, Y., J. Kim, and J. Kim, "Product Community Analysis Using Opinion Mining and Network Analysis: Movie Performance Prediction Case", *Journal of Intelligence and Information Systems*, Vol.20, No.1(2014), 49~65.

Kaplan, A. M. and M. Haenlein, "Users of the world, unite! The challenges and opportunities of Social Media," *Business horizons*, Vol.53, No.1(2010), 59~68.

Kim, Y., N. Kim, and S. R. Jeong, "Stock-Index Invest Model Using News Big Data Opinion Mining", *Journal of Intelligence and Information Systems*, Vol.18, No.2(2012), 143~156.

Kim, Y. and S. R. Jeong, "Intelligent VOC Analyzing System Using Opinion Mining", *Journal of Intelligence and Information Systems*, Vol.19, No.3(2013), 113~125.

Kim, Y., S. R. Jeong, and I. Ghani, "Text Opinion Mining to Analyze News for Stock Market Prediction," *International Journal of Advances in Soft Computing and its Application*, Vol.6, No.1(2014).

Liu, Y., "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, Vol.70, No.3(2006), 74~89.

Liu, Y., Y. Chen, R. F. Lusch, H. Chen, D. Zimbra, and S. Zeng, "User-Generated Content on Social Media: Predicting Market Success with Online Word-on-Mouth," *IEEE Intelligent. System*, Vol.25, No.1(2010), 8~12.

Lusch, R. F., Y. Liu, and Y. Chen, "The Phase Transition of Markets and Organizations: The

New Intelligentigence and Entrenreneurial Frontier," *IEEE Intelligent. System*, Vol.25, No.1(2010), 5~8.

Mangold, W. G. and D. J. Faulds, "Social media: The new hybrid element of the promotion mix," *Business Horizons*, Vol.52, No.4(2009), 357~365.

Pang, B. and L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and trends in information retrieval,* Vol.2, No,1-2(2008), 1~135.

Rao, Y., J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," *World Wide Web*, Vol.17, No.4(2014), 723~742.

Rui, H., Y. Liu, and A. Whinston, "Whose and what chatter matters? The effect of tweets on movie sales," *Decision Support System*, Vol.55, No.4(2013), 863~870.

Wu, Y., F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu, "OpinionSeer: interactive visualization of hotel customer feedback," *IEEE Transactions on Visualization and Computer Graphics*, Vol.16, No.6(2010), 1109~ 1118.

Ye, Q., R. Law, B. Gu, and W. Chen, "The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings," *Computers in Human Behavior*, Vol.27, No.2(2011), 634~639.

Yu, E., Y. Kim, N, Kim, and S. R. Jeong, "Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary", *Journal of Intelligence and Information Systems*, Vol.19, No.1(2013), 95~110.

Zhang, Z., Q. Ye, R. Law, and Y. Li, "The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews," *International Journal of Hospitality Management*, Vol.29, No.4(2010), 694~700.

국문요약

소셜미디어 콘텐츠의 오피니언 마이닝결과 시각화: N라면 사례 분석 연구

김유신* · 권도영** · 정승렬***

Web2.0의 등장과 함께 급속히 발전해온 온라인 포럼, 블로그, 트위터, 페이스북과 같은 소셜 미디어 서비스는 소비자와 소비자간의 의사소통을 넘어 이제 기업과 소비자 사이의 새로운 커뮤니케이션 매체로도 인식되고 있다. 때문에 기업뿐만 아니라 수많은 기관, 조직 등에서도 소셜미디어를 활용하여 소비자와 적극적인 의사소통을 전개하고 있으며, 나아가 소셜 미디어 콘텐츠에 담겨있는 소비자 고객들의 의견, 관심, 불만, 평판 등을 분석하고 이해하며 비즈니스에 적용하기 위해 이를 적극 분석하는 단계로 진화하고 있다. 이러한 연구의 한 분야로서 비정형 텍스트 콘텐츠와 같은 빅 데이터에서 저자의 감성이나 의견 등을 추출하는 오피니언 마이닝과 감성분석 기법이 소셜미디어 콘텐츠 분석에도 활발히 이용되고 있으며, 이미 여러 연구에서 이를 위한 방법론, 테크닉, 툴 등을 제시하고 있다. 그러나 아직 대량의 소셜미디어 데이터를 수집하여 언어처리를 거치고 의미를 해석하여 비즈니스 인사이트를 도출하는 전반의 과정을 제시한 연구가 많지 않으며, 그 결과를 의사결정자들이 쉽게 이해할 수 있는 시각화 기법으로 풀어내는 것 또한 드문 실정이다. 그러므로 본 연구에서는 소셜미디어 콘텐츠의 오피니언 마이닝을 위한 실무적인 분석방법을 제시하고 이를 통해 기업의사결정을 지원할 수 있는 시각화된 결과물을 제시하고자 하였다. 이를 위해 한국 인스턴트 식품 1위 기업의 대표 상품인 N-라면을 사례 연구의 대상으로 실제 블로그 데이터와 뉴스를 수집/분석하고 결과를 도출하였다. 또한 이런 과정에서 프리웨어 오픈 소스 R을 이용함으로써 비용부담 없이 어떤 조직에서도 적용할 수 있는 레퍼런스를 구현하였다. 그러므로 저자들은 본 연구의 분석방법과 결과물들이 식품산업뿐만 아니라 타 산업에서도 바로 적용 가능한 실용적 가이드와 참조자료가 될 것으로 기대한다.

주제어 : 빅데이터, 오피니언 마이닝, 시각화, 소셜미디어 분석, 식품 산업

논문접수일 : 2014년 11월 7일 논문수정일 : 2014년 12월 2일 게재확정일 : 2014년 12월 3일
투고유형 : 영문급행 교신저자 : 정승렬

* College of Business, University of Texas at Arlington
** 국민대학교 비즈니스 IT전문대학원
*** Corresponding author: Seung Ryul Jeong
77 Jeongneung-ro, Seongbuk-gu, Seoul 136-702, Korea
Tel: +82-2-910-4025, Fax: +82-2-910-4079, E-mail: srjeong@kookmin.ac.kr

저 자 소 개

Yoosin Kim
He is a visiting research fellow at the College of Business, the University of Texas at Arlington. He received a Ph.D. in MIS from the Kookmin University in Seoul, Korea. He had worked as a Data Scientist and a Business Analyst for a number of firms, including Accenture and SK C&C, in financial, medical, and e-commerce fields. His research interests include business analytics for consumer behaviour, market sensing, recommender systems, and business intelligence using Big Data

Do Young Kwon
She pursues a Master's degree at Koookmin University, Korea. Her research interests include opinion mining and business analytics. She has published several papers in various conference proceedings.

Seung Ryul Jeong
He is a Professor in the Graduate School of Business IT at Kookmin University, Korea. He holds a B.A. in Economics from Sogang University, Korea, an M.S. in MIS from University of Wisconsin, and a Ph.D. in MIS from the University of South Carolina, U.S.A. Professor Jeong has published extensively in the information systems field, with over 60 publications in refereed journals like Journal of MIS, Communications of the ACM, Information and Management, Journal of Systems and Software, among others.