Opinion-Mining Methodology for Social Media Analytics

Yoosin Kim[1], Seung Ryul Jeong[2]
[1] College of Business, University of Texas at Arlington, Texas, US
E-mail: yoosink@uta.edu
[2] Graduate School of Business IT, Kookmin University, Seoul, Korea
E-mail: srjeong@kookmin.ac.kr
Corresponding Author: Seung Ryul Jeong

Abstract

Social media have emerged as new communication channels between consumers and companies that generate a large volume of unstructured text data. This social media content, which contains consumers' opinions and interests, is recognized as valuable material from which businesses can mine useful information; consequently, many researchers have reported on opinion-mining frameworks, methods, techniques, and tools for business intelligence over various industries. These studies sometimes focused on how to use opinion mining in business fields or emphasized methods of analyzing content to achieve results that are more accurate. They also considered how to visualize the results to ensure easier understanding. However, we found that such approaches are often technically complex and insufficiently user-friendly to help with business decisions and planning. Therefore, in this study we attempt to formulate a more comprehensive and practical methodology to conduct social media opinion mining and apply our methodology to a case study of the oldest instant noodle product in Korea. We also present graphical tools and visualized outputs that include volume and sentiment graphs, time-series graphs, a topic word cloud, a heat map, and a valence tree map with a classification. Our resources are from public-domain social media content such as blogs, forum messages, and news articles that we analyze with natural language processing, statistics, and graphics packages in the freeware R project environment. We believe our methodology and visualization outputs can provide a practical and reliable guide for immediate use, not just in the food industry but other industries as well.

Keywords: Opinion-Mining, Korean NLP, Social Media Analytics, Food Market

---

1. Introduction

Many online communication services such as forums, blogs, online communities, and social networking sites have emerged as new communication channels that link consumers with companies. These online communications, often called social media, directly influence marketing and sales; consequently, companies aggressively participate in social media communication and use it for marketing strategies and public relations [1]. Social media also generate a large volume of content such as customer reviews, online articles, and tweets; therefore, analyzing social media content is regarded as an opportunity to discover meaningful information and intelligence for business [2], [3]. In particular, opinion mining and sentiment analysis are frequently used for such analysis because they emphasize the extraction of sentiment polarity and authors' opinions such as positive, negative, and neural sentiment [4]. For example, if a company can mine consumer opinions about products, services, brand image, and reputation by analyzing this data, the information would enable it to manage sales marketing and business strategy more effectively and efficiently [5]. In this regard, "Opinion mining, as a sub-discipline within data mining and computational linguistics, is referred to as the computational techniques used to extract, classify, understand, and assess the opinions expressed in various online news sources, social media comments, and other user-generated content. Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and other emotional states in online text" [2]. Therefore, opinion mining that involves sentiment analysis can be defined as a series of processes used to identify sentiment, nuance, and the author's attitude as shown in text and can turn this information into meaningful information for use in decision-making.

Many researchers conducted studies to discover business intelligence by using opinion-mining methodology for social media analytics. Early-stage studies that analyzed customer opinion focused on the volume of customer reviews and ratings in order to identify the influence of consumer word-of-mouth (WOM) on sales and marketing [6], [7]. However, it was realized that analyzing only volume and rating scores might not be enough to explain social dynamics in online communication; therefore, the sentiments of content were used as a new resource in WOM analysis [8]. In this regard, various approaches for opinion mining were proposed and used in several industries including e-commerce, retail, entertainment, and even stock markets. Some of the subsequent studies focused on how to use opinion mining in business fields, others emphasized methods of analysis to achieve results that are more accurate, and a few considered how to visualize in order to make results easier to understand.

Although these studies proposed frameworks, methods, techniques, and outputs for extracting business intelligence from social media content, the approaches are often technically complex and difficult for typical business users to comprehend and use. Realistically, in order to handle the large amount of complex social media content effectively and make prompt, timely decisions, the analysis results must be presented in a simple, easy-to-understand, and visual manner.

To overcome the weaknesses in current research and fill the gaps, this study proposes a comprehensive and practical methodology for social media opinion mining (SMOM). In this regard, the paper has three objectives. First, we describe the entire cycle of the SMOM framework from the initial data gathering stage to the final presentation. In this way, business analysts and marketers who are considering the use of opinion mining with social

media have a guide to the whole process. Second, we apply SMOM methodology to a case study of a Korean instant noodle company. This case study offers real-world business insights drawn from market sensing that show practical-minded business people how they can use these types of result for timely decision-making. Finally, we present several useful outputs such as a domain-specific lexicon, volume and sentiment graphs, a topic word cloud, heat maps, and a valence tree map to provide vivid, full-colored examples. All deliverables are generated by open library software packages of the R project; as a result, potential users can adopt these tools and techniques immediately.

2. Related Works

Traditional market research methods such as interviews, surveys, and observations usually investigated customers' opinions from sample groups instead of a target consumer population. However, following the emergence of the internet, online communication has become a common channel between customer and company, and has changed markets and organizations [3], [9]. In this regard, consumer WOM, which refers to personal communication among people, has been recognized as a significant source that enables companies to understand customers' interests, sentiments, and opinions about products and services. In contrast, earlier studies regarding WOM analysis simply considered reviews and popularity ratings generated by customers.

However, analyzing only the volume of customer reviews and rating scores might not be enough to explain social dynamics in online communication; therefore, the sentiments of user-generated content were applied as new parameters in WOM analysis [8]. Since then, many researchers proposed various approaches to opinion mining and conducted studies in several industries including e-commerce, retail, entertainment, and stock markets. These studies can be divided into three points of view according to the researchers' emphasis: the application view, which focuses on how to use opinion mining in business fields; the technical view, which considers analytical methods to achieve more accurate results; and the presentation view, which uses visualization for easier understanding.

2.1 The Application View: How to Use Opinion Mining

Studies that adopted the application view emphasized that opinion mining is a useful method for discovering marketing intelligence and business insights. They applied opinion mining for topic and information extraction [3], reputation comparison [10], sales forecasting [11], and market prediction [12] in various business fields.

One study suggested a market intelligence framework and conducted a case study of Wal-Mart in order to evaluate business constituents [3]. The proposed system architecture had four areas: data collection from online news and web forums; topic extraction using the mutual information method; sentiment identification using SentiWordNet; and opinion analysis, which includes traffic dynamics, topic and sentiment evolution, active topics and sentiments, and opinion leader analysis. The case study collected multiple web forums' data according to four types of opinion category: investor, employee, customer, and media. The research presented a message traffic graph, a moving average of sentiment scores, and a top five active authors list with negative sentiment issues in healthcare, customer service, unfair labor practices, and union membership. In conclusion, the researcher suggested that Wal-Mart must recognize marketing intelligence, the items that are important, and who needs to pay special attention.

Another study demonstrated opinion measurement about WOM information with regard to managerial decisions in the Hollywood movie industry [13]. This research collected online communication content for 257 movies from the message board of Yahoo Movies during 2005-2006 and extracted five measures using SentiWordNet and OpinionFinder: volume, valence, subjectivity, number of sentences, and number of valence words. The results of comparing the five measures to the movie lifecycle from the preproduction period to several weeks after release showed that the volume of messages had a significant correlation to forecast movie box office sales, but the valence of WOM did not show a meaningful relation between WOM and movie sales.

Another research study about movie sales forecasting suggested meaningful results that included the significant influence of the valence of online WOM on consumers' willingness to watch a movie [11]. The researchers gathered 4,166,623 tweets of 63 movies from June 2009 to February 2010 and conducted a filtering and categorizing process. About 380,000 filtered tweets, 9.12 % of total volume, were finally selected and categorized by a support vector machine and Naïve Bayesian classifier. The four categories were intention, positive, negative, and neutral. According to the results, "An increase by 1% in the ratio of positive tweets is associated with a $125,881 increase of movie revenue, but an increase by 1% in the ratio of negative tweets is associated with a $137,451 decrease of movie revenue" (p. 868).

Another study related to market prediction measured the correlation between the Dow Jones Industrial Average (DJIA) and online mood states on tweets [12]. This research collected a public data set of 9,853,498 tweets posted by approximately 2.7 million users for 10 months from February 28 to December 19, 2000 and tagged the mood of each tweet using a two-mode tracking tool. The tool's first mode calculated positive and negative sentiment and the second mode measured six mood dimensions: calm, alert, sure, vital, kind, and happy. The results suggested that public mood is correlated to the DJIA and the accuracy of standard stock market prediction models is able to improve significantly in certain mood dimensions such as calm.

2.2 The Method and Technique View: How to Analyze to Achieve Greater Accuracy

While prior studies focused on how to use opinion mining, some researchers emphasized the use of methods and techniques to improve opinion-mining capabilities such as accuracy. A case study of the film market proposed new methodologies for mining online reviews in order to predict movie sales performance [14]. The researchers attempted to generate several mining process models: sentiment probabilistic latent semantic analysis (S-PLSA) for summarizing deeper sentiment than simple classification as negative or positive; an autoregressive sentiment aware (ARSA) model for predicting sales performance by using sentiment analysis results and past performance values; and an autoregressive sentiment and quality aware (ARSQA) model for considering review quality. The proposed models demonstrated the significant impact of review sentiment and volume on the accuracy and effectiveness of the analysis results.

Another study into opinion mining's capabilities used objective words in SentiWordNet [15]. The research proposed that such words in SentiWordNet represented a public sentiment lexicon and were rarely used for sentiment mining; therefore, it conducted research to measure the sentimental relevance between objective words and sentiment sentences using support vector machines (SVM). The study's results showed a 4.1% improvement in accuracy from the traditional SentiWordNet to the revised SentiWordNet. However, because

approaches using single words or a flat structure could face several challenges such as particular linguistic and nonlinguistic knowledge, text styles, and domains, a further study highlighted the development of corpora for sentiment analysis [16]. The authors tried to find new concept-level approaches to the use of semantic and affective resources for annotation by investigating the Italian Twitter corpus.

Elsewhere, a study proposed a probability model to measure the sentiment of financial news articles and predicted movements of the Korea Composite Stock Price Index (KOSPI) [17]. The authors suggested a text opinion-mining method that included data collection, natural language processing (NLP), a generated domain-specific sentiment dictionary, and relevant news articles' opinions compared with rises and falls of the KOSPI. They achieved a 63.0% F1 score on the threshold of validation data. Further, in order to extract opinions and sentiments more accurately from news content, a study generated a lexicon of stock domain-specific sentimental words and achieved higher accuracy than with a general dictionary [18]. Finally, a contextual-meaning study proposed an algorithm to build automatically a word-level emotional dictionary for social emotion detection [19]. The authors stated that a dictionary generated for a specific purpose is more efficient at predicting the emotional distribution of news articles.

2.3 The Presentation View: How to Visualize for Easy Comprehension

The other view of opinion-mining research focuses on presentation in order to improve recognition using visualization. As an effective means of communication, a visual metaphor is able to provide users with an integrated view of analysis results and enables them to discover significant opinion patterns easily [20]. Thus, researchers have investigated visualization methods and systems in order to synthesize analysis results for marketers, analysts, and other users.

One study generated a visual analytic tool to support journalists and other professionals in the media industry who want to mine valuable news items from a large volume of social media content [21]. The researchers collected 101,285 tweets regarding the US presidential State of the Union address of 2010 and presented the analysis process and results. In this regard, they designed research to collect, analyze, aggregate, and visualize social media content about one broadcast event. They suggested using a visualized user interface tool with an integrated view that included a keyword search and filtering section, video and Twitter messages, topic sections, a message volume graph, and a sentiment trends bar.

Other researchers built an interactive visualization system, the OpinionSeer, to analyze online hotel customer reviews. The purpose was to achieve greater comprehension of customer opinions [20]. The researchers categorized each opinion by document level and feature level (e.g., room, service, and price), and attempted to reveal any uncertainty hiding in customer reviews through visualization. The system had three major components: opinion mining in order to extract customer sentiment from hotel reviews; subjective logic in order to define review features; and data visualization techniques. The visualization system was specifically developed to recognize opinion patterns easily and included a new visual layout such as an opinion wheel that used scatter plots and radial graphs.

2.4 Summary

Consequently, opinion mining can be categorized as three different points of view in **Fig. 1** according to what is emphasized in the research. The application view focuses on how to use

opinion mining; the technical view considers how to improve analytical capabilities; and the presentation view uses visualization to improve effective comprehension.
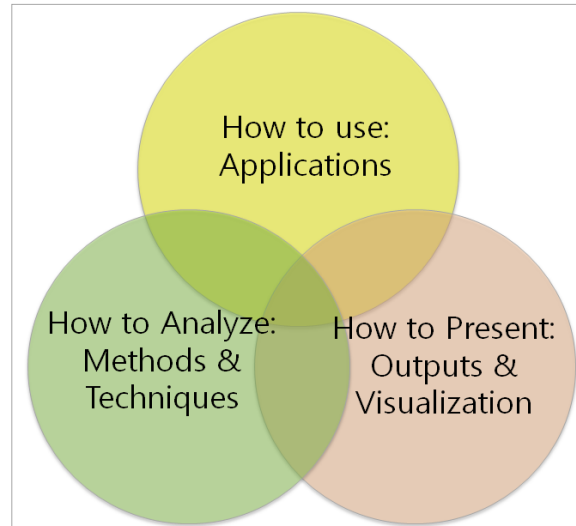

Fig. 1. Three Views of Opinion-Mining Research

Studies that researched the application view highlighted how to use opinion mining in business fields in order to discover marketing intelligence and business insights. They analyzed the opinions and complaints of business actors such as consumers, investors, and employees; forecasted movie sales performance; and generated new prediction parameters for stock market investments. Researchers who used the technical view focused on methods and techniques to find more effective and efficient approaches for improving aspects of analysis performance such as accuracy. They generated moderate algorithms and models, used a specific-purpose dictionary, and applied a new approach. The remaining researchers used visualization for the effective communication of opinion-mining outputs. They believed that a visual metaphor enables business decision-makers to understand analysis results easily and helps to discover significant opinion patterns. In this regard, they developed new visualization systems and suggested the use of visualized user interfaces and figures that provide an integrated view of analysis results.

Even though these studies have proposed frameworks, methods, and techniques for extracting business intelligence from social media content, there is still a lack of practical social analytics. First, the methodology explaining the overall process of opinion mining is scarce and does not offer a practical guide for conducting social media analysis. For example, the MI2 framework omitted social networking sites such as Facebook and Twitter, and its analytics and outputs were not given in sufficient detail to use for practical analysis [3]. Second, the technical view's approaches are often technically complex and difficult for typical business users to understand and use. There is also a lack of figures, tables, graphics, and other visualized images to use for practical opinion mining. This is because the researchers usually focused on mathematical formulae and artificial intelligence methods in order to increase accuracy rather than present an overall process and visualized outputs. Finally, studies rarely attempted to use visualization with regard to opinion mining for social media analytics. Some studies proposed the use of visual layouts and user interface tools with integrated views; however, data were limited to customer reviews without social

content such as blogs and tweets [20]. In addition, visualization often included too much information to enable effective comprehension of analysis results [21]. In the real business world, in order to handle the large amount of complex social media content effectively and make prompt, timely decisions, analysis results must be presented in a simple, easy-to-understand, and visual manner. To overcome the aforementioned weaknesses and fill gaps in the current research, this study proposes a comprehensive and practical approach to opinion mining from social media content.

3. The Proposed Approach: Social Media Opinion-Mining Methodology

As aforementioned, opinion mining is deployed to extract, classify, understand, and assess the opinions implicit in text content. Further, sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and emotional states toward entities, events, and their attributes in such content [22], [23]. Therefore, a social media opinion-mining methodology should have processes that involve computational techniques to aggregate, extract, analyze, and present the sentiment and attitude of authors in social media content. In this study, we propose a methodology, shown in **Fig. 2**, for social media opinion mining and conduct this approach into a real business case.

1. Connect to target social media channels and collect data from them

2. Qualify the collected data using natural language processing

3. Apply opinion-mining analytics to the qualified data set

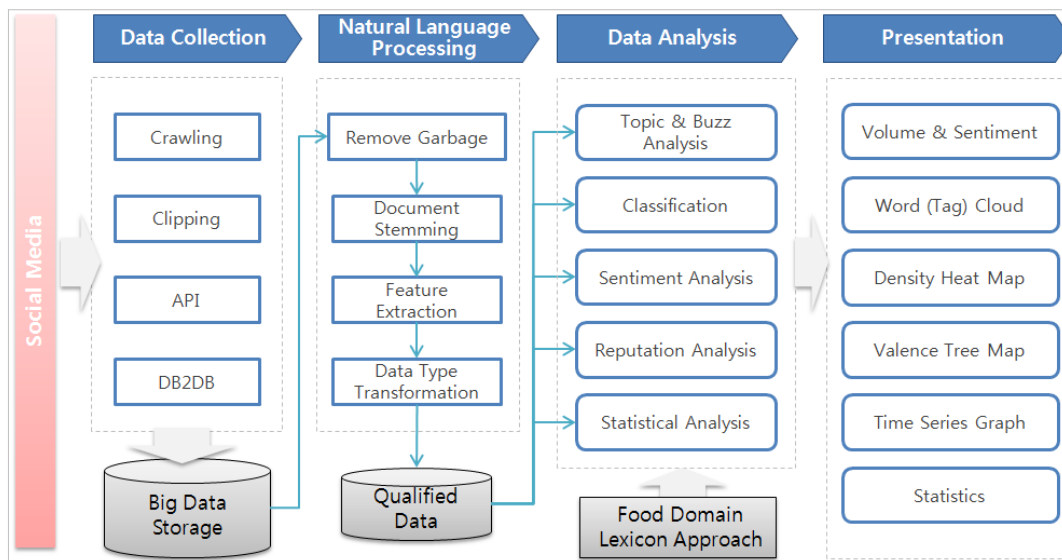4. Visualize and present opinion-mining results

Fig. 2. Overview of Social Opinion Mining Methodology

3.1 Phase 1: Social Data Aggregation

The first phase for mining social media opinion involves the choice of target social media channels and the collection of data from them. There are many kinds of online communication channel, and the ways to collect data differ depending on the type of social

media. For example, social networking sites such as Twitter and Facebook provide open API for accessing and gathering their data. Portal websites such as Yahoo.com and Naver.com do not support open API but an analyst can use search engine tools and techniques such as web crawlers, web scraping, and search engine robot software programs. In addition, social data can be purchased from social media data providers or obtained directly by applying database-to-database (DB2DB) interfacing modules to social media that allow data collection. Further, because data-gathering methods differ according to the type of social media, the nature of collected data also varies. Some social media data are stacked as log files in storage while others are managed in relational database. Basically, though, such data are unstructured text generated by users, the volume of which is considerable. In addition, social media are sometimes filled with too much noise such as advertisements and meaningless online emoticons that could distort opinion-mining results. A study had removed ninety percent of gathered tweet data through filtering in the research [11]. If the analyst neglects such garbage data, the results of social media analysis will not provide meaningful and useful business insights. Therefore, aggregated big data should be preprocessed in order to generate useful materials for meaningful analysis.

3.2 Phase 2: Data Qualification for Analytics

In the second phase, after aggregating the unstructured text data set, a rigorous data-qualifying procedure using NLP should be conducted. NLP is a computational technique that manipulates, understands, interprets, and presents natural language text for linguistic analysis. In this phase, NLP is responsible for preprocessing activities: parsing sentences, removing disabled letters, extracting features, and tagging specific characters. For example, meaningless characters such as html tags, punctuation, numbers, and emoticons are eliminated. In addition, stop words, which are invalid words such as prepositions, pronouns, and certain words that are defined as worthless, are removed in this cleansing process. The subsequent qualified data is then transformed into an analysis data format such as relational data or structured data. The format includes manipulated text and identification information such as created date, author name, content identification, counts, reviews, and favorites. For example, in the R project software program, social media content can be extracted as list structured text data from a social data file and then exchanged and treated as list, matrix, and vector types. Following this, through NLP the qualified data are stored as a data frame structure combined with identification data. In addition, domain-specific lexicon resources such as a sentiment dictionary and stop words can be generated to improve opinion-mining accuracy.

3.3 Phase 3: Applying Opinion-Mining Analytics

The next phase applies various analytics to mine market intelligence and business insights. The qualified data set includes not only information from user-generated content but also various identification data. Depending on the purpose of the analysis, the analyst and researcher can select suitable mining tools. For example, topic extraction and buzz analysis are usually related to market trends analysis, which interests many people. On the other hand, sentiment analysis is utilized to evaluate the reputations of products, services, and companies, and applied to establish customer responses to marketing activities. If domain-specific language resources such as lexicons or thesauruses have been generated and used in this phase, the analyst can expect more comprehensive and reasonable results [17]. For instance, sentiment analysis results categorized by business domain-specific lexicons can provide a detailed map about the characters that are negative or positive.

3.4 Phase 4: Presenting Analysis Results with Visualized Deliverables

The last and concluding phase of the methodology is to present opinion-mining results using visualized outputs such as graphs, tables, and matrixes. As aforementioned, many studies paid little attention to the effective and efficient communication of opinion-mining results to business users. However, effective visualization is able to explain a considerable portion of analysis results with an integrated visual figure that has no additional descriptions. Therefore, the major focus and purpose of this phase is to make results simple, clear, and easy to understand rather than complex and ostentatious so that business users can easily comprehend their meaning and use them for decision-making. For example, a tag cloud in topic analysis can visualize topic volume with an intuitive visible font color and size; a sentiment heat map can reveal customers' positive or negative opinions by using two contrasting colors such as red and green; and a valence tree map can provide straightforward visualization by using both volume and sentiment in a hierarchical categorization.

4. Analysis and Results

To illustrate our proposed methodology, we conducted a case study of an instant noodle company, the SY Food Corporation, in South Korea. The market size of the instant noodle "ramen" business in Korea was over US$2 billion in 2013. In particular, SY ramen, a representative product of SY Food, was released as the first ramen in Korea in 1963 and is still ranked on the top 10 ramen list in terms of revenue according to AC Nielsen Korea.

4.1 Data Collection and Preprocessing

We collected 14,204 items of social media content including blogs, forum (café) messages, and media news articles from January 2012 to June 2013. The collected data were user-generated text content together with author names, user IDs, release dates, URL addresses, etc. This content was gathered from online community websites such as Naver.com and Daum.com in South Korea with a web crawler that used a search keyword of a product name, "SY ramen." The types and volume of the collected data are shown in **Table 1**.

Table 1. Types and Volume of Data Set

18 Months	Blogs (Naver)	Blogs (Daum)	Forum (Naver)	News	Others	Total
Volume	5,524	1,647	3,917	1,749	1,367	14,204
Ratio	38.9%	11.6%	27.6%	12.3%	9.6%	100%

According to Table 1, the volume of blogs from Naver.com, the foremost Korean portal website, was the biggest in the data set while the volume of news articles from mass media was relatively low. However, according to **Fig. 3**, which shows the movement of data volume along the time line, the volume of news rapidly increased in certain periods such as March and August in 2012 while the volumes of blogs and forum messages were relatively fixed. Thus, it could be said that online consumers revealed various interests and opinions about SY ramen, but media news articles responded significantly to social events such as government compliance and regulation.
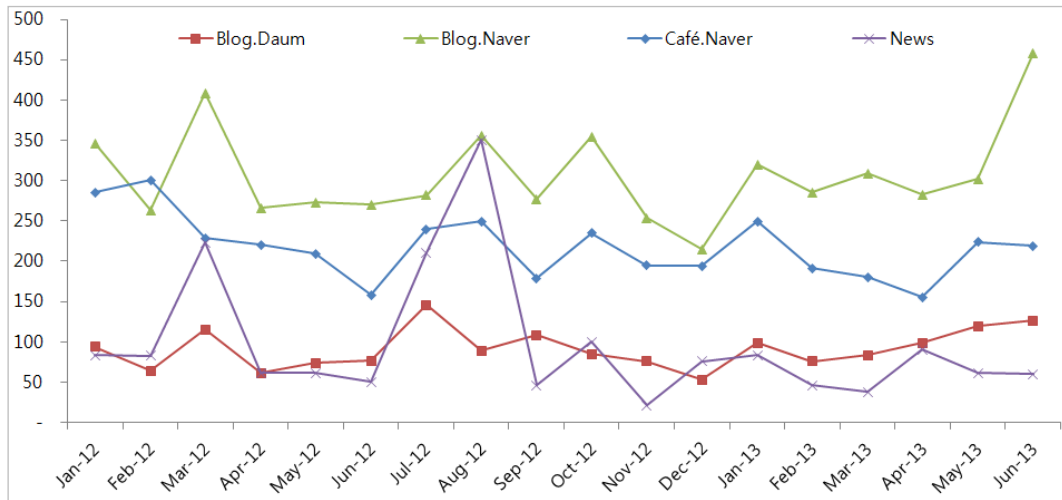
Fig. 3. Data Volume by Social Media Source on a Time Line

After gathering social media data, we generated domain-specific language resources such as a domain sentiment dictionary, and stop words for the instant noodle business. Since the Korean language does not have a public SentiWordNet for opinion mining, these tasks were conducted by following the opinion-mining model of [17], which generated a domain-specific sentiment dictionary for stock market forecasts. Next, the qualified data through pre-processing were applied into topic extraction, sentiment analysis, and other mining analytics. Before beginning them, if contents breakdown with subcategories, we can describe analysis results in more detail. In order to achieve a greater understanding of customers' opinions, a study divided features of hotel reviews contents with room, service and price [20]. Such classification provides a frame that enables closer observation of social media volume and sentiment status from real business perspectives. In this study, we tried to classify the contents within general categories which consist of the four marketing Ps (product, price, promotion, and place), environment, and management of the instant noodle business.

4.2 Sentiment Analysis

From this section, we introduce the visualized opinion-mining outputs with a few statistics. The first output, as shown in **Table 2**, is the sentiment analysis result which reports volume and ratio of polarity: positive, negative and neutral. We can see positive sentiment ratio in whole content is around 26.5% and negative sentiment is about 11%.

Table 2. Result of Sentiment Analysis

Sentiment	Count	%
Total	14,204	100.0%
Positive	3,769	26.5%
Negative	1,574	11.1%
Neutral	8,861	62.4%

Next deliverable, as displayed in **Fig. 4**, is a simple graph showing the movement in the daily averaged sentiment scores over the research time line. In this figure, the range for sentiment scores is between +1 (extremely positive) and -1 (extremely negative). Sentiment scores are almost all over 0 (neutral sentiment), which means that customers' opinions are relatively positive.
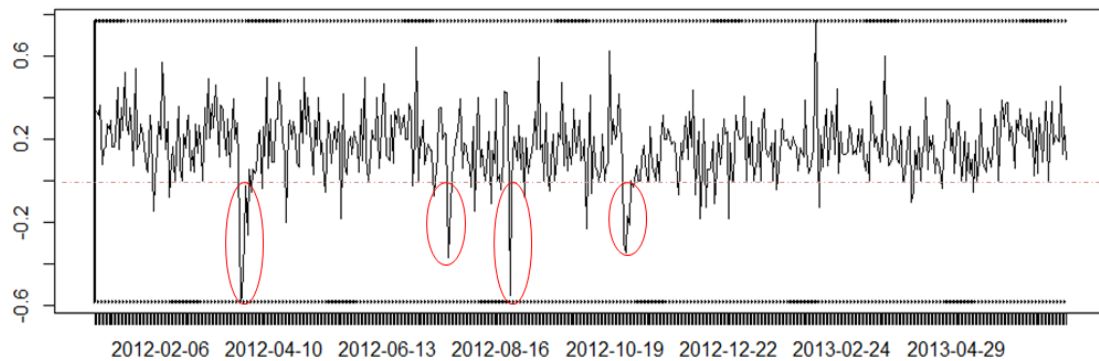
Fig. 4. Daily Sentiment Flow

To see whether the gap of sentiment exists between customers and conventional media, we divided the content source into user-generated content (UGC) such as blogs and forums and media-generated content (MGC) such as news article. According to T-test result, in **Table 3**, there was no significant difference between two sources.

Table 3. Sentiment T-Test by Sources

		Mean	SD	t	p
Sentiment	UGC	0.159	0.138	-0.6936	0.488
	MGC	0.143	0.506		

However, we can find that two social media sources show different movement in **Fig. 5**. The sentiment flow of consumer-generated content remains positive and stable area around a score of 0.2 with rare noticeable changes over time, but exceptional dropping points like March, 2012. Interestingly, this period was indeed a time of crisis for instant noodle companies because the Fair Trade Commission (FTC) in Korea imposed a US$100 million fine in that month as a penalty for price collusion among them. On the other hand, the pattern of media channel shows sharp fluctuations from extreme positive to negative. We can consider mass media sensitively reacts to social events rather than preference about the product.
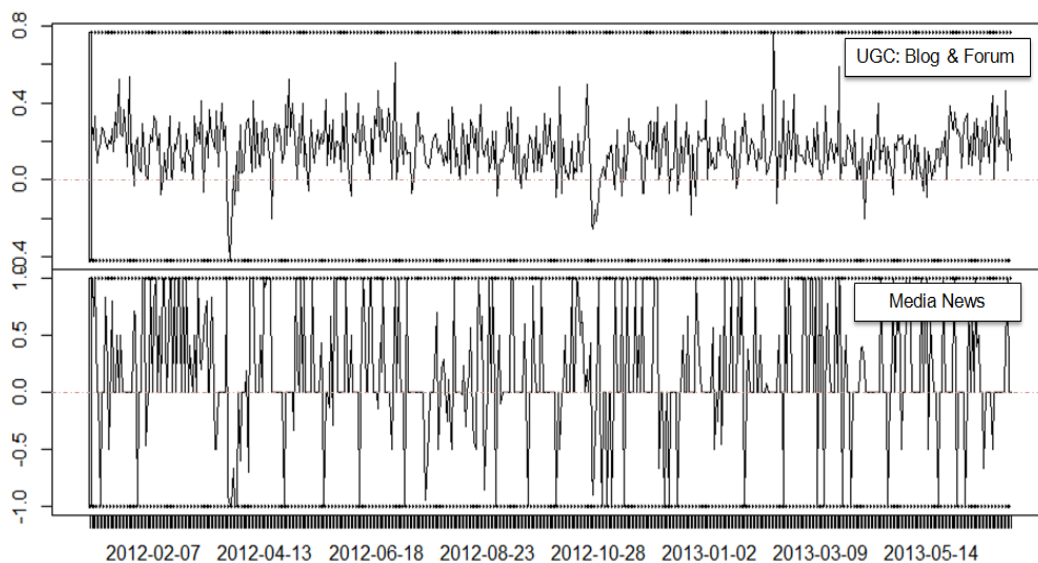

Fig. 5. Daily Sentiment Flow: UGC vs. Media News

4.3 Topic Extraction

Fig. 6 presents a word cloud (or tag cloud) that extracted hot issues and high frequency topic keywords during the targeted period. The size of a word in the cloud reflects topic volume while color emphasizes the topic. Greater insight can be obtained from this type of word cloud; for example, **Fig. 6** indicates that the most significant issues about SY ramen were "Penalty" and "FTC (Fair Trade Commission." This is understandable since the FTC imposed a large penalty on instant noodle companies. Other significant topics were new products such as Ggoggomen™ and Nagasaki Noodle™ because these were highly successful and set new trends in the noodle market. Analysts and business users would expect these to be major issues for companies.
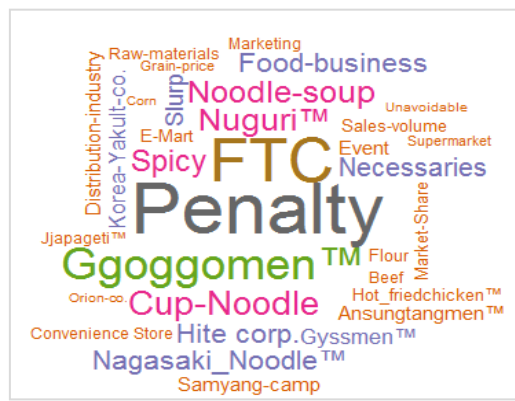

Fig. 6. Word Cloud of Topic Extraction

4.4 Feature Classification

In **Table 4**, the feature classification result is reported with count and ratio of classified contents; mean and standard deviation of each category; the result of one-way ANOVA test. First, ANOVA test with sentiment score as the dependent variable and feature categories as the independent variable shows that sentiment among feature categories is significantly different (F = 99.12, p < .01). Comparing sentiment on categories, "soup" feature ranked the highest positive (m =0.437, sd = 0.637), and the rest ordered as price noodle, promotion, place, distribution, recipe, competitor, design, top management, material and environment.

Table 4. ANOVA Test of Sentiment on Category

No.	Category	n	Ratio	Mean	SD	F	p
1	Soup	2,052	14.4%	0.437	0.637		
2	Price	471	3.3%	0.406	0.657		
3	Noodle	368	2.6%	0.266	0.599		
4	Promotion	780	5.5%	0.231	0.602		
5	Place	420	3.0%	0.210	0.581		
6	Distribution	79	17.4%	0.203	0.644	99.12	0.000*
7	Recipe	1,346	9.5%	0.195	0.551		
8	Competitor	2,054	14.5%	0.131	0.613		
9	Design	165	1.2%	0.079	0.492		
10	Top Mgmt	333	2.3%	0.030	0.584		
11	Material	766	5.4%	-0.090	0.663		
12	Environment	481	3.4%	-0.432	0.712		

* denote significance levels at 1 %, respectively

We can see the same result with a visualized output. Heat map, shown in **Fig. 7**, visualizes sentiment status with additional information by using hot and cold colors in a category and time matrix. In this grid, three colors (black, grey, and yellow) are used to present differences between negative and positive sentiments. Black indicates negative sentiment and yellow reflects positive sentiment. If the color of a cell in the grid is closer to yellow, this means that the cell has more positive than negative sentiment opinions. On the other hand, if the color is nearly black, the cell contains mostly negative sentiment. In **Fig. 7**, all soup category cells are close to yellow, thus intuitively suggesting that the consumer-intimate opinion about SY ramen soup is generally very positive. In contrast, some cells in the material, management, and environment categories are mostly in black, indicating potential problems and a need for the company to pay more attention to these areas.
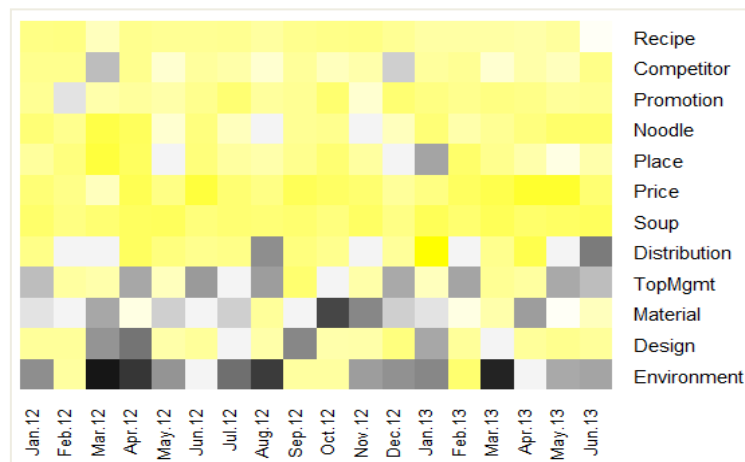

Fig. 7. Sentiment Heat Map on Categories

4.5 Tree Map consisting Volume and Sentiment

Heat maps show the density of either volume or sentiment according to categories over a series of periods, whereas tree maps present both volume and sentiment at the same time. Thus, a valence tree map, one of the most comprehensive and holistic visualization modes, can be very helpful for analysts and decision-makers because it enables them to understand the "big picture" of a business situation quickly alongside a hierarchical structure. A simple glance at such a map detects areas that are weak, strong, positive, negative, quiet, or loud.
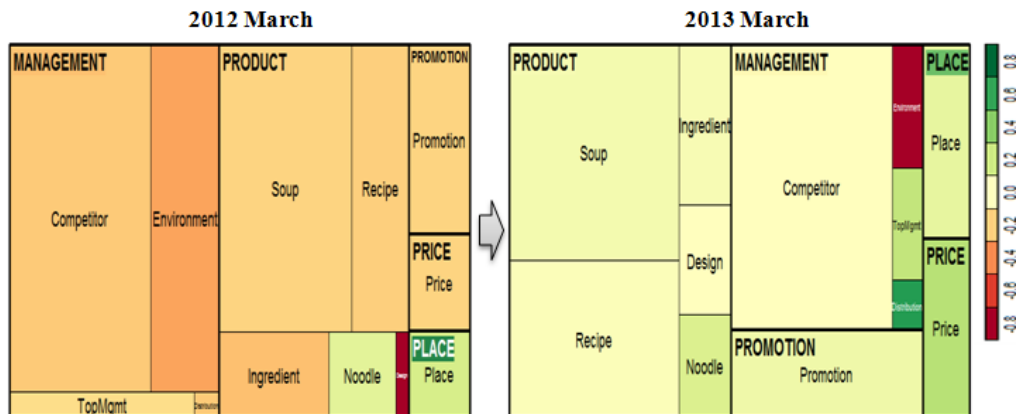

Fig. 8. Tree Map: Hierarchical View

As revealed in **Fig. 8**, in March 2012 SY Food faced very negative sentiment according to the buzz related to "penalty," "unfair," and "fine" in the management and environment categories, thereby affecting the company's reputation adversely. However **Fig. 8** also shows that the negative sentiment in social media had calmed by March 2013 and the crisis had passed. In addition, the map shows that the most significant interest about the instant noodle ramen is product features such as soup taste, noodle and recipe.

5. Conclusion

Social media, such as online communities, forums, blogs, and social networking sites, have rapidly become important new communication channels between customers and companies. They continuously generate a large amount of unstructured text data called social media content, social data, or social media data. Social media content that contains customers' opinions and interests is recognized as valuable material from which businesses can find useful information. Thus, many researchers have investigated opinion-mining frameworks, methods, techniques, and tools for mining business intelligence in various industries. These studies focused on how to use opinion mining in specific business fields; explored methods to analyze data and achieve results with greater accuracy; and considered visualization tools to enable users to understand data more easily. However, such approaches are often too technically complex and insufficiently user-friendly to help with business decisions and planning. To overcome these problems, we attempted to formulate a more comprehensive and practical methodology for social media opinion mining and applied this methodology to a case study of the oldest instant noodle product in Korea. We also presented graphical tools and visualized outputs that included volume and sentiment graphs, a business domain-specific lexicon, a topic and issue word cloud, a volume heat map, a sentiment heat map, and a valence tree map with a hierarchical structure. Our resources came from public domain social media content such as blogs, forum messages, and news articles, and we analyzed the data with the NLP, statistics, and graphics packages in the freeware R project environment. The resources and tools used here can be adopted easily by businesses for their market intelligence operations.

Our study has several implications and contributions. First, our proposed methodology is practical and comprehensive because our approach covers the entire cycle of opinion-mining activities from initial data collection to the final visualization of social media opinion mining. In this regard, we illustrated the whole cycle by applying the methodology to a case study of a Korean instant noodle company. Second, the various visualization outputs we presented with real social media content could serve as real-world examples of our approach for potential adopters who are considering the use of opinion-mining analytics in their business environments. Third, we believe that our approach can provide a practical and reliable guide to opinion mining with visualized results that are immediately useful not just in the food industry but in other industries as well.

Nonetheless, our study has several limitations. This paper focused on opinion-mining methodology to analyze social media content and used a case study of just one instant noodle company. In addition, our research used social media content such as blogs, café (forum) messages, and news articles in the blogosphere but did not use social networking sites such as Twitter and Facebook. Although social networking sites have a great deal of garbage data and private chat, the opinions contained in crowd buzz can help to identify quickly the

hottest topics and the source of "big mouths." Thus, in future research, we plan to include several competitors from the same market and reveal competitive intelligence in the dynamic social media world. In addition, we will attempt to apply our approach to other domains such as health care, entertainment, finance, and education. Because each industry has a unique culture and structure, the analysis results and interpretations are likely to vary according to the domain.

References

[1] J. H. Kietzmann, K. Hermkens, I. P. McCarthy, and B. S. Silvestre, "Social media? Get serious! Understanding the functional building blocks of social media," *Bus. Horiz.*, vol. 54, no. 3, pp. 241–251, May 2011. Article (CrossRef Link).

[2] H. Chen and D. Zimbra, "AI and Opinion Mining," *IEEE Intell. Syst.*, vol. 25, no. 3, pp. 74–76, May 2010. Article (CrossRef Link).

[3] H. Chen, "Business and Market Intelligence 2.0, Part 2," *IEEE Intell. Syst.*, vol. 25, no. 2, pp. 2–5, Mar. 2010. Article (CrossRef Link).

[4] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New Avenues in Opinion Mining and Sentiment Analysis," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 15–21, Mar. 2013. Article (CrossRef Link).

[5] M. Chau and J. Xu, "Business Intelligence in Blogs: Understanding Consumer Interactions and Communities," *MIS Q.*, vol. 36, no. 4, pp. 1189–1216, 2012. http://dl.acm.org/citation.cfm?id=2481684.

[6] J. A. Chevalier and D. Mayzlin, "The Effect of Word of Mouth on Sales: Online Book Reviews," *J. Mark. Res.*, vol. 43, no. 3, pp. 345–354, Aug. 2006. Article (CrossRef Link).

[7] Y. Liu, "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *J. Mark.*, vol. 70, no. 3, pp. 74–89, Jul. 2006. Article (CrossRef Link).

[8] G. P. Sonnier, L. McAlister, and O. J. Rutz, "A Dynamic Model of the Effect of Online Communications on Firm Sales," *Mark. Sci.*, vol. 30, no. 4, pp. 702–716, Jul. 2011. Article (CrossRef Link).

[9] R. F. Lusch, Y. Liu, and Y. Chen, "The Phase Transition of Markets and Organizations: The New Intelligence and Entrenreneurial Frontier.," *IEEE Intell. Syst.*, vol. 25, no. 1, pp. 5–8, 2010. Article (CrossRef Link).

[10] B. Liu, "Sentiment Analysis: A Mutlifaceted Problem," *IEEE Intell. Syst.*, vol. 25, no. 3, pp. 76–80, 2010. Article (CrossRef Link).

[11] H. Rui, Y. Liu, and A. Whinston, "Whose and what chatter matters? The effect of tweets on movie sales," *Decis. Support Syst.*, vol. 55, no. 4, pp. 863–870, Nov. 2013. Article (CrossRef Link).

[12] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011. Article (CrossRef Link).

[13] Y. Liu, 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 Intell. Syst.*, vol. 25, no. 1, pp. 8–12, 2010. Article (CrossRef Link).

[14] X. Yu, Y. Liu, X. Huang, and A. An, "Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 4, pp. 720–734, Apr. 2012. Article (CrossRef Link).

[15] C. Hung and H.-K. Lin, "Using Objective Words in SentiWordNet to Improve Word-of-Mouth Sentiment Classification," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 47–54, Mar. 2013. Article (CrossRef Link).

[16] C. Bosco, V. Patti, and A. Bolioli, "Developing Corpora for Sentiment Analysis: The Case of Irony and Senti-TUT," *IEEE Intell. Syst.*, vol. 28, no. 2, pp. 55–63, Mar. 2013. Article (CrossRef Link).

[17] Y. Kim, S. R. Jeong, and I. Ghani, "Text Opinion Mining to Analyze News for Stock Market Prediction," *Int. J. Adv. Soft Comput. Its Appl.*, vol. 6, no. 1, 2014. http://home.ijasca.com/data/documents/Paper-ID-424-IJASCA_Formated.pdf

[18] Y. Yu, Y. Kim, N. Kim, and S. R. Jeong, "Predicting the Direction of the Stock Index by Using a Domain-Specific Sentiment Dictionary," *J. Intell. Inforamtion Syst.*, vol. 19, no. 1, pp. 95–110, 2013. http://jiisonline.evehost.co.kr/files/DLA/6-19-1.pdf.

[19] Y. Rao, J. Lei, L. Wenyin, Q. Li, and M. Chen, "Building emotional dictionary for sentiment analysis of online news," in *World Wide Web*, 2013. Article (CrossRef Link).

[20] Y. Wu, F. Wei, S. Liu, N. Au, W. Cui, H. Zhou, and H. Qu, "OpinionSeer: interactive visualization of hotel customer feedback.," *IEEE Trans. Vis. Comput. Graph.*, vol. 16, no. 6, pp. 1109–18, 2010. Article (CrossRef Link).

[21] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine, "Diamonds in the rough: Social media visual analytics for journalistic inquiry," in *2010 IEEE Symposium on Visual Analytics Science and Technology*, 2010, pp. 115–122. Article (CrossRef Link).

[22] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business Intelligence and Analytics: From Big Data To Big Impact," *MIS Q.*, vol. 36, no. 4, pp. 1165–1188, 2012. http://dl.acm.org/citation.cfm?id=2481683.

[23] B. Pang and L. Lee, *Opinion Mining and Sentiment Analysis*. 2008. Article (CrossRef Link).

Yoosin Kim is a visiting research fellow at the College of Business, the University of Texas at Arlington. He received a Ph.D. in MIS from 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.

Seung Ryul Jeong 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.