ISSN 1598-0170 (Print)
ISSN 2287-1136 (Online)
http://www.jksii.or.kr

Competitive intelligence in Korean Ramen Market using Text Mining and Sentiment Analysis

Yoosin Kim[1] Seung Ryul Jeong[2*]

ABSTRACT

These days, online media, such as blogospheres, online communities, and social networking sites, provides the uncountable user-generated content (UGC) to discover market intelligence and business insight with. The business has been interested in consumers, and constantly requires the approach to identify consumers' opinions and competitive advantage in the competing market. Analyzing consumers' opinion about oneself and rivals can help decision makers to gain in-depth and fine-grained understanding on the human and social behavioral dynamics underlying the competition. In order to accomplish the comparison study for rival products and companies, we attempted to do competitive analysis using text mining with online UGC for two popular and competing ramens, a market leader and a market follower, in the Korean instant noodle market. Furthermore, to overcome the lack of the Korean sentiment lexicon, we developed the domain specific sentiment dictionary of Korean texts. We gathered 19,386 pieces of blogs and forum messages, developed the Korean sentiment dictionary, and defined the taxonomy for categorization. In the context of our study, we employed sentiment analysis to present consumers' opinion and statistical analysis to demonstrate the differences between the competitors. Our results show that the sentiment portrayed by the text mining clearly differentiate the two rival noodles and convincingly confirm that one is a market leader and the other is a follower. In this regard, we expect this comparison can help business decision makers to understand rich in-depth competitive intelligence hidden in the social media.

☞ keyword : Text Mining; Sentiment Analysis; Competitor Analysis; User-generated Content; Korean Instant Noodle Market

1. Introduction

After emerging online media such as blogosphere, online communities, and social networking sites, user-generated content (UGC) has been considered the valuable information for marketing insight [1][2][3]. For instance, blogs, one of the UGC, have become a critical source to mine business intelligence [4], because blogs contain personal stories which are richly embedded with interests, sentiments, and opinions, and bloggers share their interests and opinions with each other as well. If a company can mine consumer opinions about products, services, brand images, and reputations by analyzing their online behaviors, the information would help firms to manage sales, marketing, and execute business strategy more effectively and efficiently [4].

Many researchers and marketers have attempted to mine consumer behavior from social media data using text mining techniques such as opinion mining and sentiment analysis in various industries including movies [5][6], music albums [7], hotels [8][9], restaurants [10][11] and retail business [1]. Those studies have focused on just a single case or in a restricted context of a specific phenomenon. However business users want to obtain first-hand information about not only the market reception of own products and services but also competitors. Recently, there are some studies for competitive analysis such as comparison of pizza companies [12], comparing two brands of cellular phone [13], and a multiple case study about the football clubs' brand in UK [14]. Due to the lack of the comparison research for competing products and rival companies, we attempt to do competitive analysis and opinion mining in social media through a multiple case study comparing consumer behavior toward rival products. In addition, this study seeks to develop the domain specific sentiment dictionary for opinion mining of Korean texts because there is rarely research for Korean alphabet "Hangeul" in sentiment analysis [15]. We follow the guidelines for conducting a multiple case study

1 BigData Analytics Dept., Univ. of Seoul, Seoul, 02504, Korea
2 Graduate School of Business IT, Kookmin Univ., Seoul, 02727, Korea
* Corresponding author (srjeong@kookmin.ac.kr)
[Received 19 January 2018, Reviewed 24 January 2018, Accepted 3 February 2018]

http://dx.doi.org/10.7472/jksii.2018.19.1.155

prescribed by Yin [16], and select two popular and competing instant noodles in the same industry: a market leader and a market follower in the Korean instant noodle industry.

In the context of our study, we conduct opinion mining to present inherently “qualitative” consumers’ sentiment and thus employ quantitative measurements to demonstrate the differences between the competitors. We will incorporate techniques and tools such as data collection, generating a sentiment lexicon, defining taxonomy for categorizing, and polarity classification in opinion mining and sentiment analysis. Our results will show that, indeed, the sentiment portrayed by the opinion mining clearly differentiate the two rival noodles and convincingly confirm that one is a market leader and the other is a follower. In this regard, we expect this comparison can help business decision makers to understand rich in-depth competitive intelligence hidden in the social media.

2. Related Work

After emergence of Web 2.0 technologies, the social media such as forums, blogs, online communities, and social networking sites have become the common channel linking customers and companies. The content generated by users in social media is recognized as a significant source for understanding customers’ interest, sentiment, and opinion about a company’s products and service for the markets and the business organizations [1][17][18][19]. In advance, online communications now directly influences sales and marketing, and firms aggressively uses it for communicating with customers and understanding them [20]. These efforts can help a company to obtain first-hand knowledge on market reception of its products and services, and even those of the competitors [12]. This knowledge enables business analysts and decision makers to develop insights on consumer opinions, discover new ideas for marketing, improve customer satisfaction, ultimately increasing returns on business investments [4][21].

In addition, analyzing UGC in social mediacan help to obtain not only the market insight of oneself but also competitors’ information because social media is publicly opened to anyone who can access and crawl it. Competitive intelligence in social mediacan support that a company understands suppliers, competitors, environments, business trends within social dynamics [22]. Through comparing with competitors, the company can take comprehension about the market position, the competing condition, the customer reputation and opportunities against rivals [23]. Therefore, the competitive intelligence to monitor the competitive environment can provide a competitive edge to the business [24]. Indeed, some researchers showed the competitive intelligence via comparing the rivals in various industries such as the high-end smart-phone market [23] and the pizza chain industry [12].

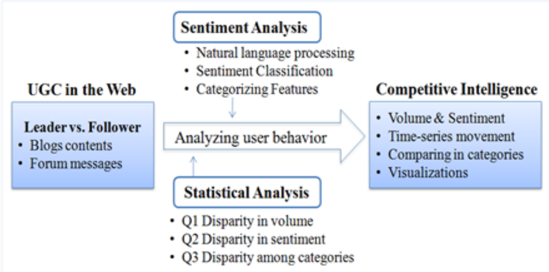
Meanwhile, opinion mining and sentiment analysis is gradually deployed to extract, classify, understand, and assess the opinions implicit in text contents [21][25][26]. Opinion mining and sentiment analysis enable marketers and researchers to extract consumers’ opinions, sentiments, and interests from enormous amount of spontaneous, unsolicited comments generated by the entire target consumer population. Therefore, many researchers have used the method to examine the influence of consumer word-of-mouth on marketing and sales, and showed that online media content provides more credible information sources than traditional advertisements in forming their opinion regarding a firm’s reputation.

In the ways to analyze consumer sentiment in the context, two approaches are generally employed: the sentiment lexicon-based approach and the machine learning method. The sentiment lexicon based approach applies the linguistic resource with sentiment polarity and the machine learning approach uses the classification algorithm to determine sentiment of the context. Chen [1] used SentiWordNet for a market intelligence framework to analyze stakeholder opinions of Wal-Mart. He gathered board messages from multiple web forums such as yahoo, employee communities and Wall Street Journal, and drove the opinion mining results such as a message traffic graph, a moving average of sentiment scores, and a top five active authors. On the other hand, a study gathered tweets on 63 movies from June 2009 to February 2010 and assigned them to four types: intention, positive, negative, or neutral using support vector machine and naïve Bayesian classifier [6]. They stated that the valence of movie tweets has significant influence on consumers' willingness to watch a movie. In addition, Ortigosa et al. [27] tried to combine and apply lexicon-based classification technique and machine learning algorithm in sentiment analysis of Facebook users for e-learning, and then revealed emotional movement of users.

Our brief literature review showed that the UGC of social media has significant influence on marketing and sales in various industries such as movies, music album, restaurants, and e-commerce, however most studies analyzing the context of them have focused on just a single case or a specific phenomenon, based on simple polarity classification with only positive or negative sentiment. In addition, Korean word-of-mouth data has seldom been examined in sentiment analysis because Korean is an agglutinative language, which makes it difficult to analyze morphemes within corpora [15]. In this paper, our contribution is two-fold: (1) we compare two rival products by using sentiment analysis and statistics for finding competitive intelligence, and (2) we develop a ramen domain-specific sentiment dictionary for the Korean instant noodle market for opinion mining.

3. Proposed Approach

In business world, competitive analytics comparing rivals is very typical analysis and still in challenging area, because the comparison with rival companies can help a firm to discover a rich, in-depth competitive strategy. Fan & Gordon [22] explained how social media analysis can provide competitive intelligence to help organizations to understand their suppliers, competitors, environments, and overall business trends. For the comparison of consumer behavior toward the competing products, this study employs a multiple case study approach and suggests a hybrid opinion mining method combining sentiment analysis and statistics. Sentiment analysis presents various dimensions of something that is inherently consumers' sentiment and statistical analysis invalidates the differences of sentiment analysis results between the products by quantitative measurements. We follow the guidelines for conducting multiple case studies suggested by Yin [16] and thus select two popular competing instant noodles in the same industry: a market leader and a market follower in the Korean instant noodle industry.



(Figure 1) Proposed Approach

To compare consumer emotion in blogs online, we seek to investigate the following three research questions in mining consumers' opinion toward rival firms competing in the same industry:

Q1. Given rival firms with different market shares, how would the market share gap is manifested by UGC volume in the social media?

Q2. Given rival firms with different market shares, how would the market share gap is manifested by consumer sentiment in the context of UGC?

Underneath the "hard" objective reality in market share, consumers' "soft" opinions toward the rivaling brands drive their purchasing decisions which ultimately lead to difference in market shares. If leaders of the rivaling firms are able to gather real-time intelligence from these opinions and gauge differences in volume and sentiment toward the competing brands, more timely actions may be taken to influence their opinions and improve their brands' market shares.

Q3. Are there differences in consumer sentiment toward various product features such as soup, recipe, and ingredients, as well as marketing features like price, distribution, and promotion?

These features for product and marketing underlie the most crucial decisions related to a firm's competitive strategy. By examining and comparing consumers' sentiments toward these features, we may gain invaluable insights which may help long-term as well as on-going marketing decision makings. Since product taste such as soup, noodle and ingredient seems to be a critical factor for its market success, in food business, the result examining the differences between consumers' sentiments toward the taste features can provide significant insights to understand why they have established a market leader and a follower in Ramen business.

3.1 Data

To drive a clear and comprehensive comparison, we targeted two popular and competing instant noodles, Ns-ramen and Sy-ramen, in the Korean ramen market. According to the Nielsen report [28], market size of “Ramen” business in Korea was over $2 billion in the year 2012, and the two products have been rivals for several decades. Sy-ramen has been produced as the first ramen in Korea since 1963. Whereas, Ns-ramen has kept No.1 market position for recent 10 years in the ramen business. Based on market share data in April, 2012 [28], Ns-ramen is a market leader (15% market share) and Sy-ramen is a follower (5.1% market share).

To collect data, we have used a crawler program to scrap the user-generated contents mentioning the ramen names in online communities. The crawler program gathered 19,386 pieces of UGC between January and September in the year 2012 from portal web sites: Naver.com, Daum.com, and other related sites in South Korea (see Table 1). This web-search program used the two ramen product names as unique keywords: Sy-ramen and Ns-ramen.

〈Table 1〉 Amount of Collected Data

Ramen	2012.1	2012.2	2012.3	2012.4	2012.5
Ns	1,061	1,112	1,282	1,140	1,152
Sy	813	672	853	619	609
Sum	1,874	1,784	2,135	1,759	1,761
	2012.6	2012.7	2012.8	2012.9	sum
Ns	1,045	1,184	1,303	1,215	10,494
Sy	575	701	745	637	6,224
Sum	1,620	1,885	2,048	1,852	16,718

Not surprisingly, the volume of UGC data shows a large gap between the two firms reflecting a distinct difference in their respective market shares. Comparison of the collected data amount shows that Ns-ramen is almost two times that of Sy-ramen.

4. Sentiment Analysis Results

4.1 Korean Sentiment Dictionary for Ramen

After gathering the UGC data, we developed a Korean sentiment dictionary for the lexicon based sentiment analysis. For the works, we considered the process proposed in the Opinion-Mining Methodology for social media analytics [29] and conducted three steps; 1) choose the sample data to make a domain specific dictionary, 2) extract a domain specific terms, and 3) test a generated preliminary dictionary.

As the first process, the sample data which has sentiment polarity about ramen has built for the ramen specific lexicon. We randomly sampled 4000 blogs from the collected data and manually checked whether each blog has sentiment as well as which polarity, positive or negative. Through the manual check and labeling of them, finally, we selected 1000 blogs consisting negative 500 and positive 500. Next, we performed natural language processing on the blog content. We eliminated useless letters (e.g., kkk-ㅋㅋㅋ and hhh-ㅎㅎㅎ), and characters such as emoticons (e.g., -.-, ^^, L), numbers, and punctuation. Then, we parsed the data to break down the content into a term level-unigram. The terms were sorted by frequency and high frequency words over five times were remained as the preliminary sentiment terms. In this processing for term extraction, we used “KoNLP” (the package for Korean natural language processing) and “tm” (the package for text mining) in the R project.

As the result, the sentiment dictionary is made of 233 positive words and 242 negative words, and it should be a domain specific sentiment dictionary based on instant noodle business. Table 2 shows a few sample words of the sentiment lexicon. Terms expressing positive emotions included “spicy”, “delicious”, and “savory”. Meanwhile, negative words such as “trash”, “terrible”, and “penalty”, clearly disclosed the authors’ unpleasant feelings towards the movies. In special, negative terms included some words unrelated with taste such as “rigging”, “penalty”, and “Fair Trade Commission”. We deduced the terms came from the negative event which Trade Commission had imposed a huge fine to instant noodle companies due to unfair price-fixing in March, 2012.

(Table 2) Sample of Sentiment Dictionary

Negative	Positive
Tasteless	Spicy
Terrible	Savory
Trash	Differentiation
FTC (Fair Trade Commission)	Delicious
Penalty	Chewy
Rigging	Like

After sentiment words were defined, we regarded the classified terms as a preliminary sentiment dictionary to apply into sentiment analysis of consumer reviews about ramens. We tested the preliminary sentiment dictionary through sentiment lexicon-based classification of the sample data set. The sentiment of a ramen blog was evaluated by the distances between the number of appearances of the positive and negative terms. We made a formula to classify the sentiment of each blog. The appearance of sentiment words in each blog was calculated as a positive and negative point, and the gap of both sides judges a sentiment of a blog as a below rule:

IF positive point UCG(i) > negative point UCG(i) THEN

UCG(i)Sentiment = Positive;

ELSEIF positive pointUCG (i) <= negative point UCG(i) THEN

UCG(i)Sentiment = Negative;

ELSEIF positive point UCG(i) = Null && negative point UCG(i) = Null THEN

UCG(i)Sentiment = Neutral;

END IF;

We employed the measurements such as accuracy, precision and recall to investigate the performance of the lexicon based sentiment classification [6][15]. The accuracy is defined as the percentage of sentiments correctly predicted of the total instances and thus the sentiment dictionary reached 75.6% accuracy of the generated sentiment dictionary (Please refer to Table 3).

(Table 3) Classification Performance by Sentiment Dictionary

Sentiment	Accuracy	Precision	Recall
Positive	75.6%	71.1%	86.2%
Negative		82.5%	65.0%

Comparing to the previous research, which showed 81.4% accuracy using 2,633 term features [5], we deemed the performance of the sentiment lexicon is not low to apply it to sentiment analysis of all user-generated content despite of the accuracy under 80%.

4.2 Sentiment ANalysis by the Lexicon

Using the generated sentiment dictionary, we conducted sentiment analysis with the data, which amounted to 17,348 texts of two rival ramens. Table 4 shows the summary of sentiment analysis and thus we can know intuitively that Ns-ramen has more positive sentiment than Sy-ramen. Sy-ramen have a 6.9% negative and a 21.8% positive data, while Ns-ramen showed a ratio of negative 4.2% and positive 26.8% in the data.

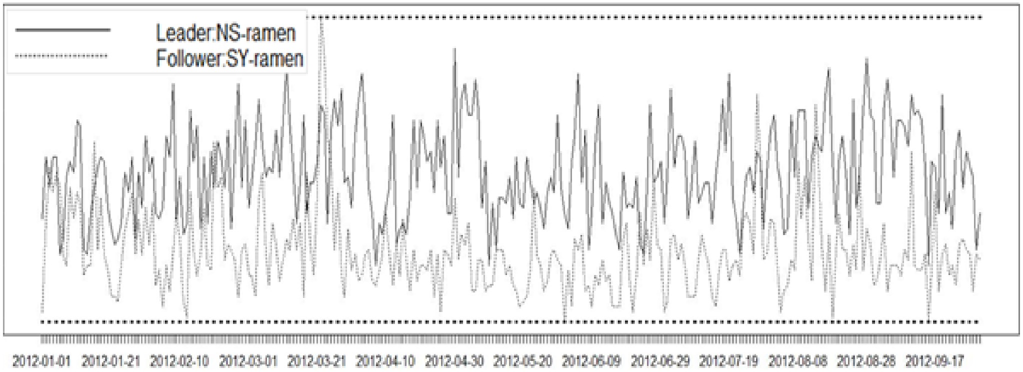
(Table 4) Classification Result by Sentiment Dictionary

Ramen	Negative	Neutral	Positive	Sum
Ns	470 (4.2%)	7,707 (69.3%)	2,947 (26.5%)	11,124
Sy	429 (6.9%)	4,439 (71.3%)	1,356 (21.8%)	6,224
Sum	899 (5.2%)	12,146 (70.0%)	4,303 (24.8%)	17,348

4.3 Feature Categorization of the Content

In addition, we categorized the collected data to examine more intimately in various aspects of the ramen. Such categorization was successfully attempted by precious research such as hospitality features [8] and cellular phone [13]. In this study, the categories include the widely accepted 4P's of marketing (product, price, promotion, and place), the environment, and management. To make the taxonomy, we chose high frequency terms from the nouns used in making sentiment dictionary, and selected reasonable features manually. For example, we defined a soup taste feature as the words meaning the texture of food and employed hot, spicy, silky, rich, light, and so on. Another instance, we defined top management features and set the related words such as CEO, CFO, Chairman, and also the name of CEO.

Feature classification results reveal that the most common type of on-line comments center around product features, with “soup taste” mentioned the most (20-25 %). For



(Figure 2) Time-series Volume Comparison

comments on competitor, there are some differences (Sy-ramen 26.9% vs. Ns-ramen 15.9%) though the total counts are similar (Sy-ramen 1,676 vs. Ns-ramen 1,672).

(Table 5) Categorization Result of Product Features

Feature		Sy	N=6,224	Ns	N=10,494
Product	Noodle	410	6.6%	575	5.5%
	Soup	1,297	20.8%	2,700	25.7%
	Ingredient	514	8.3%	762	7.3%
	Recipe	1,008	16.2%	2,176	20.7%
	Design	98	1.6%	193	1.8%
Price		280	4.5%	332	3.2%
Promotion		490	7.9%	666	6.3%
Place		246	4.0%	357	3.4%
Environment		272	4.4%	150	1.4%
Mgmt	Top Mgmt	231	3.7%	132	1.3%
	Competitor	1,676	26.9%	1,672	15.9%
	Distribution	176	2.8%	127	1.2%

5. Statistical Analysis Results

In this section, we compared the statistical difference of the customer sentiment between the two ramen competitors.

5.1 Comparison of UGC Volume

First we performed T-test on the daily volume of both competitors' content to examine to check whether the volume gap of user-generated content exists between competitors in consonant with their market shares. The result reported that the daily volume of the data is significantly different between Ns-ramen and Sy-ramen (see Table 6). The mean of Ns-ramen's daily content (38.30) is nearly doubled than Sy-ramen's (22.72).

(Table 6) T-Test Result of Volume

Daily	Mean	SD	t	p
Ns-ramen	38.30	9.11	20.85	0.00**
Sy-ramen	22.72	8.40		

** denotes significance levels at .01, respectively

To provide comprehensive understand the gap between two rival products, we depicted time-based graphs of both ramens and merged them in one plot to compare the gap (see Figure 2).

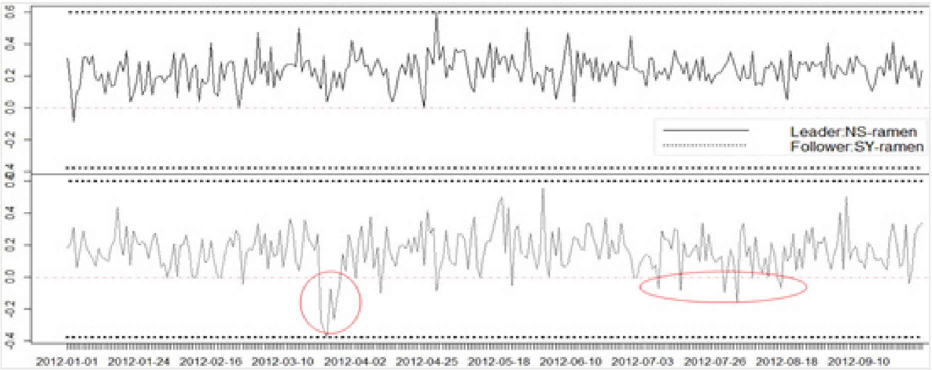
5.2 Comparison of UGC Sentiment

Then, we compared the daily sentiment score of both competitors to examine whether the gap of consumer sentiment existed between competitors. The range for sentiment scores is between +1 (extremely positive) and -1 (extremely negative). According to the T-test result as shown Table 7, Ns-ramen (0.23) is significantly positioned in more positive area than Sy-ramen (0.16).

(Table 7) T-Test Result of Sentiment

Daily	Mean	SD	t	p
Ns-ramen	0.23	0.09	7.69	0.00**
Sy-ramen	0.16	0.13		

** denotes significance levels at .01, respectively



(Figure 3) Time-series Sentiment Comparison

The visualization of the sentiment graph, as shown below in Figure 3, presents a comparison of sentiment toward the two competing firms. As can be seen, though both graphs showed a similar wave staying mostly positive, the sentiment wave of Sy-ramen went down to zero or below more often. For instance, in March, 2012, Fair Trade Commission (FTC) had imposed a huge fine to instant noodle companies due to unfair price-fixing, and the all companies took a negative effect. However, the graph reveals that Ns-ramen had overcome the negative event with positive blogs from the fan, but Sy-ramen had dramatically dived in negative sentiment.

The results in Figure 4 clearly portray the dynamic market conditions in the instant noodle business, where the volume of Ns-ramen as the market leader is almost always higher than Sy-ramen. Moreover, sentiment toward Ns-ramen is also almost always more positive than SY-rumen. These statistics and opinion mining results provide substantive bases for explaining the reasons why Ns-ramen has higher market share.

5.3 Comparison of Sentiment among Features

Taste features of instant noodle, especially soup, are critical factors for market success for Ramen products. In both ramens, taste features generate more positive sentiment from customers than other features. We further explore consumers' sentiments toward Ramen taste in Q3 to see if there are significant differences in sentiments between the two companies for each individual taste feature. As expected, results of T-test, as shown in Table 8, indicate the significant disparity between the market leader Ns-ramen and the follower Sy-ramen in soup taste. In addition, ingredient has also the significant gap between rivals except noodle feature. In fact, the soup taste of Ns-ramen, emphasizing hot and spicy, is very popular and loved in the market because most Korean prefers hot and spicy taste in food. For these three features, we do see that the market leader has higher sentiment score.

(Table 8) T-Test Result of Product Features' Sentiment

Features		Sy-ramen		Ns-ramen		t	P
		Mean	SD	Mean	SD		
Pro-duct	Noodle	0.49	0.51	0.51	0.54	-0.44	0.66
	Soup	0.36	0.56	0.44	0.54	-4.46	0.00**
Ingredient		0.56	0.63	0.68	0.52	-3.77	0.00**
	Recipe	0.28	0.55	0.31	0.54	-1.50	0.13
	Design	0.27	0.79	0.48	0.78	-2.16	0.03*
Price		0.13	0.53	0.21	0.53	-1.92	0.06
Promotion		0.13	0.68	0.25	0.61	-3.03	0.00**
Place		0.20	0.67	0.46	0.64	-4.80	0.00**
Environment		-0.29	0.70	0.10	0.68	-5.61	0.00**
Mgt	Top Mgt	0.03	0.72	0.10	0.71	-1.71	0.09
Competitor		0.16	0.58	0.25	0.56	-4.44	0.00**
Distribution		-0.04	0.60	0.02	0.63	-0.77	0.44

* and ** denote significance levels at .05 and .01, respectively

We have presented analysis results in the forms of traditional tables and graphs, showing either volume or sentiment. Both volume and sentiment simultaneously with advanced visualization modes for presentation. Tree map, one of the most comprehensive and holistic modes for

(Figure 4) Valence Tree Maps: March vs. June, 2012

visualization, is shown in Figure 4 below for a targeted time period: March and June, 2012. This map is aimed to help analysts and decision makers to understand "big picture" business situation with a hierarchical structure. A swift glance can quickly detect areas that are weak, strong, positive, negative, quiet or loud.

As illustrated in Figure 4, Sy-ramen in March faces a very negative sentiment in red-color by "Penalty", "Unfair", and "Fine" in management and environment categories, and its reputation is adversely affected. Fortunately, the statues for product categories remain relatively positive with green-color though their volume is decreased in comparison with other maps. On the other hand, NS food's tree maps for March and June are somewhat similar in volume and sentiment. Although NS food also had many negative buzz in environment and management categories during the March crisis, there were very positive contents in the product area as well. Several months later, in June, when crisis was over, the online behavior of consumer became calm, whereas there are still some areas with dark color in Sy-ramen.

6. Discussion and Implications

If a business can analyze the market reputation about its products and services, the information would help the firm to manage sales, marketing, and execute business strategy more effectively and efficiently [4]. Recently, the social media including blogospheres, social networking sites, and online communities, has become the critical source to understand the consumer. In this regard, many studies have tried to analyze consumer behaviors in online communications, and thus opinion mining and sentiment analysis has been employed in various industries including entertainment business, hospitality, and retail business. In addition, comparison study is tried to investigate the market competition and provide competitive intelligence. Our multiple case study supports definite evidence as showing obvious difference in consumer opinion toward rival instant noodles.

The results reveal that the two noodles are significantly different in the volume of blogs and the sentiment melting in the content. According to the analysis result, the market leader, Ns-ramen, has almost twice volume of daily blogs than the follower, Sy-ramen. The sentiment analysis result showed that Ns-ramen is more positive than Sy-ramen as well. It means that quantity and quality of UGC is preferred to Ns-ramen, at the same time the result explains why Ns-ramen is the market leader. Furthermore, we conducted further sentiment analysis by categorization of ramen features to clearly compare the disparity between two competitors. Since taste features of ramen are critical factors for market success, we had regarded the gap in taste features such as soup, ingredient, and noodle exist between the market leader Ns-ramen and the follower Sy-ramen. As expected, the analysis results prove that they have the significant gap in soup and ingredient as the market position gap.

As findings for our study, we show that such differences in market success can be explained and predicted by several categories of analytics on consumer opinions and sentiment toward the two rival ramens. First, we found significant differences in volume and sentiment between the market leader and the follower (Q1). Next, we revealed that feature categories in the instant noodle business can be significantly distinguished and ordered for the two rivals in sentiment levels (Q2). Last, expectedly, for taste features like soup and ingredient, which are technically critical for market success, showed significant gaps between competitors (Q3). Therefore, the dominance in general reputation enjoyed by the market leader is clearly evident. In addition, overall sentiment and reputation of online customer behavior reflect not only the product features but also every respects of business like management and external environment, and the market leader is generally ahead of the follower in this regard. And also, this study shows the feasibility of the Koran sentiment dictionary, which is purpose to compare business rival products in sentiment analysis. The demonstrated lexicon achieved 75.6 % classification accuracy despite the small number of linguistic features, 233 positive words and 242 negative words. This result should support a more efficient and effective means of conducting sentiment analysis on big data and provide a good reference for potential users of sentiment analysis as well.

Consequently, the findings propose the firm to gain possible competitive advantage by analyzing social media data. A company can constantly monitor its own reputation in the market as well competitors with public social media data. Indeed, the company can build its own business taxonomy and sentiment dictionary according to the purpose of the competitive analysis system. The results from the analysis system can provide the qualitative, quantitative, and visualized information, and the firm can use them to make business decisions.

7. Conclusion and Future Research

As communication in the social media has become a common channel among consumers, it is important to extract the valuable information in the social media data. Especially, firms want to know how to find the consumers' opinions and what the competitive advantage is in the competing market. In order to accomplish the comparison study for rival products and companies, we attempted to do competitive analysis and opinion mining in social media through a multiple case study. Since there were few studies Korean alphabet "Hangeul" in sentiment analysis, also developed the domain specific sentiment dictionary of Korean texts. We conducted a case study and selected two popular and competing instant noodles in the Korean instant noodle industry: a market leader and a market follower. We gathered 19,386 pieces of user-generated blogs between January and September, 2012 year from portal web sites, developed the Korean sentiment dictionary, and defined the taxonomy for categorization. In the context of our study, we employed opinion mining to present consumers' sentiment and statistical analysis to demonstrate the differences between the competitors. Our results show that, indeed, the sentiment portrayed by the opinion mining clearly differentiate the two rival noodles and convincingly confirm that one is a market leader and the other is a follower. In this regard, we expect this comparison suggests that, for business users, the online behavior disparity between rivals can provide timely and effective intelligence for social media strategy and general marketing strategies. For researchers, it provides compelling support for the validity and predictive power of the opinion mining analytics for competitive market intelligence as well.

Future studies may proceed in a number of directions. First, the current study may be extended to include two or more rivals in the same market. Since the competition in food industry, indeed, is very complicated and intense, future research should involve more competitors to reveal competitive intelligence in the real business world. On the other hand, future researchers have to consider the communication channel in the Web such as social networking sites (e.g., Twitter and Facebook), Wikipedia, and also mass media service. Since each channel has its own character in communicating among users, analysis result might derive different implications depending on channels. For instance, Twitter allowing 140 letters in a post is limited in expressing an author's opinion but is strength on very speedy diffusion. Consumer opinions in crowd tweet buzz can be helpful to quickly catch the hottest topic and identify more dynamic market trends. Finally, industries other than instant noodle can be studied. It is possible that different industries may entail a different set of concerns and priorities. For example, health care service, entertainment business, financial industry, and educational institutions can be studied and further advance the method and techniques/tools.

References

[1] Chen, H., "Business and Market Intelligence 2.0," Part 2. IEEE Intelligent Systems, 25(2), 2-5, 2010. https://doi.org/10.1109/MIS.2010.43

[2] Chen, H., & Zimbra, D., "AI and Opinion Mining," IEEE Intelligent Systems, 25(3), 74-76, 2010. https://doi.org/10.1109/MIS.2010.75

[3] Mangold, W. G., & Faulds, D. J., "Social media: The new hybrid element of the promotion mix," Business Horizons, 52(4), 357-365, 2009. https://doi.org/10.1016/j.bushor.2009.03.002

[4] Chau, M., & Xu, J., "Business Intelligence in Blogs: Understanding Consumer Interactions and Communities," MIS Quarterly, 36(4), 1189-1216, 2012. https://www.misq.org/business-ingelligence-in-blogs-understanding-consumer-interactions-and-communities.html

[5] Pang, B., Lee, L., & Vaithyanathan, S., "Thumbs up?: sentiment classification using machine learning techniques," Proceedings of the Conference on Empirical Methods in Natural Language Processing, 79-86, 2002. https://doi.org/10.3115/1118693.1118704

[6] Rui, H., Liu, Y., & Whinston, A., "Whose and what chatter matters? The effect of tweets on movie sales," Decision Support Systems, 55(4), 863-870, 2013. https://doi.org/10.1016/j.dss.2012.12.022

[7] Dhar, V., & Chang, E. a., "Does Chatter Matter? The Impact of User-Generated Content on Music Sales," Journal of Interactive Marketing, 23(4), 300–307, 2009. https://doi.org/10.1016/j.intmar.2009.07.004

[8] Wu, Y., Wei, F., Liu, S., Au, N., Cui, W., Zhou, H., & Qu, H., "OpinionSeer: interactive visualization of hotel customer feedback," IEEE Transactions on Visualization and Computer Graphics, 16(6), 1109-18, 2010. https://doi.org/10.1109/TVCG.2010.183

[9] Ye, Q., Law, R., Gu, B., & Chen, W., "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, 27(2), 634-639, 2011. https://doi.org/10.1016/j.chb.2010.04.014

[10] Lu, X., Ba, S., Huang, L., & Feng, Y., "Promotional Marketing or Word-of-Mouth? Evidence from Online Restaurant Reviews," Information Systems Research, 24(3), 596-612, 2013. http://pubsonline.informs.org/doi/abs/10.1287/isre.1120.0454

[11] Zhang, Z., Ye, Q., Law, R., & Li, Y., "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, 29(4), 694-700, 2010. https://doi.org/10.1016/j.ijhm.2010.02.002

[12] He, W., Zha, S., & Li, L., "Social media competitive analysis and text mining: A case study in the pizza industry," International Journal of Information Management, 33, 464-472, 2013. https://doi.org/10.1016/j.ijinfomgt.2013.01.001

[13] Liu, B., "Sentiment Analysis: A Mutlifaceted Problem," IEEE Intelligent Systems, 25(3), 76-80, 2010. https://doi.org/10.1109/MIS.2010.75

[14] McCarthy, J., Rowley, J., Jane Ashworth, C., & Pioch, E., "Managing brand presence through social

media: the case of UK football clubs," Internet Research, 24(2), 181--204, 2014. https://doi.org/10.1108/IntR-08-2012-0154

[15] Lim, J., & Kim, J., "An Empirical Comparison of Machine Learning Models for Classifying Emotions in Korean Twitter," Journal of Korea Multimedia Society, 17(2), 232-239, 2014. https://doi.org/10.9717/kmms.2014.17.2.232

[16] Yin, R. K., "Case study research: Design and methods," Applied Social Research Methods Series, 18(2), 54-56, 2013. https://doi.org/10.1097/00001610-199503000-00004

[17] Liu, Y., Yu, X., An, A., & Huang, X., "Riding the tide of sentiment change: Sentiment analysis with evolving online reviews," World Wide Web, 16(4), 477-496, 2013. https://doi.org/10.1007/s11280-012-0179-z

[18] Lusch, R. F., Liu, Y., & Chen, Y., "The Phase Transition of Markets and Organizations: The New Intelligence and Entrenreneurial Frontier," IEEE Intelligent Systems, 25(1), 5-8, 2010. https://doi.org/10.1109/MIS.2010.75

[19] Wang, Y., Aguirre-urreta, M., & Song, J., "Investigating the Value of Information in Mobile Commerce : A Text Mining Approach," Asia Pacific Journal of Information Systems, 26(4), 577-592, 2016.

[20] Kietzmann, J. H., Hermkens, K., McCarthy, I. P., & Silvestre, B. S., "Social media? Get serious! Understanding the functional building blocks of social media," Business Horizons, 54(3), 241-251, 2011. https://doi.org/10.1016/j.bushor.2011.01.005

[21] Chen, H., Chiang, R. H. L., & Storey, V. C., "Business Intelligence and Analytics: From Big Data To Big Impact," MIS Quarterly, 36(4), 1165–1188, 2012. https://misq.org/skin/frontend/default/misq/pdf/V36I4/SI_ChenIntroduction.pdf

[22] Fan, W., & Gordon, M. D., "The power of social media analytics," Communications of the ACM, 57(6), 74-81, 2014. https://doi.org/10.1145/2602574

[23] Kim, Y., Dwivedi, R., Zhang, J., & Jeong, S. R., "Competitive intelligence in social media Twitter: iPhone 6 vs. Galaxy S5," Online Information Review, 40(1), 42-61, 2016. https://doi.org/http://dx.doi.org/10.1108/OIR-03-2015-0068

[24] Kahaner, L., "Competitive Intelligence : How to Gather, Analyze, and Use Information to Move Your Business to the Top," Touchstone, 1998. http://www.amazon.com/Competitive-Intelligence-Analyze-Information-Business/dp/0684844044

[25] Kim, Y., Kim, N., & Jeong, S. R., "Stock-index Invest Model Using News Big Data Opinion Mining," Journal of Intelligence and Inforamtion Systems, 18(2), 143-156, 2012.

[26] Pang, B., & Lee, L., "Opinion Mining and Sentiment Analysis," Foundations and Trends in information Retrieval, 2008. https://doi.org/10.1561/1500000011

[27] Ortigosa-Hernandez, J., Rodriguez, J.D., Alzate, L., Lucania, M., Inza, I., Lozano, J.A., "Approaching Sentiment Analysis by Using Semi-Supervised Learning of Multi-dimensional Classifiers," Neurocomputing 92, 98-115, 2012. https://doi.org/10.1016/j.neucom.2012.01.030

[28] AC Nielsen Korea, The Revenue of Instant Noodle Products in April 2012.

[29] Kim, Y., & Jeong, S. R., "Opinion-Mining Methodology for Social Media Analytics," KSII Transactions on Internet and Information Systems, 9(1), 391-406, 2015. https://doi.org/10.3837/tiis.2015.01.024

◐ Authors ◑

Yoosin Kim
2000 B.A. in MIS, Kookmin University
2009 M.A. in MIS, Graduate School of Business IT, Kookmin University
2013 Ph.D. in MIS, Graduate School of Business IT, Kookmin University
2016～Present: Research Professor, Dept. of Big Data Analytics, University of Seoul,
Research Interests: Text Mining, AI Big Data Analytics and Utilization, System Implementation etc
E-mail : yoosin25@uos.ac.kr

Seung Ryul Jeong
1985 B.A. in Economics, Sogang Univ., Seoul, Korea
1989 M.S. in MIS, Univ. of Wisconsin, WI, U.S.A.
1995 Ph.D. in MIS, Univ. of South Carolina, SC, U.S.A.
1997～Present: Professor, Graduate School of Business IT, Kookmin Univ., Korea
Research Interests: Text Mining, Opinion Mining, Business Analytics, System Implementation, Process Innovation, Project Management, etc.
E-mail : srjeong@kookmin.ac.kr