ISSN 2234-3075 (Print)
ISSN 2288-7806 (Online)
Http://dx.doi.org/10.14771/AIM.5.2.3

The Current Situation of the Big Data Utilization in the Agricultural Food Area and its Future Direction[1)]

Daniel Byungho Chung · Jongpyo Cho · Junghoon Moon

ABSTRACT The purpose of this study is to prove that new values for the agricultural food area can be created by combining various big data collected in the agricultural food area and analyzing them in an appropriate analysis method. For this, the analysis techniques generally used were studied, and the use of the big data in the various areas of the current society was explored through practical application instances. In addition, by the current status and analysis instances of the big data use in the agricultural food area, this study was conducted to verify how the new values found were being used.

Key words Big data, Jeongbu3.0, Combination among the big data

D. B. Chung · J. P. Cho · J. Moon
Regional Information
Seoul National University, South Korea

D. B. Chung
e-mail: jncoboyss@gmail.com

J. P. Cho
e-mail: jongpyocho@snu.ac.kr

J. Moon
e-mail: moonj@snu.ac.kr

1 Introduction

With innovative developments of Internet technology and the generalization of smart devices, various social networks such as Twitter, Facebook, etc., have settled down as a new trend in society. As a result, the boundary between the consumers' and producers' source of information has disappeared, and tremendous amounts of data has been produced every second. According to the studies of the McKinsey Global Institute (McKinsey & Company, 2011), worldwide enterprises produced a total 7 exabytes of data and individuals produced 6 exabytes of data in 2010. If 1 exabyte is about 4,000 times the amount of data from the library of the US congress, tremendous amounts of data is being produced and distributed every year.

However, the question is whether the 'big data' is being efficiently used. There has been an effort to extract valuable information in business administrations and economics for a long time, and the achievements are starting to appear. As one of the examples of the achievement, the late night bus route was newly established by combining the bus route administration data of the city of Seoul and floating population analysis data of KT in the current transportation area (SisaJeju, 2013). With the practical achievement in the transportation area, the major government organizations and enterprises in Korea are trying to use the big data in the various ways. The agricultural food area is also starting to make an effort for this. Although there are various forms of the agricultural food related database and the 'big data' across a whole procedure including production, distribution, processing of agricultural food is collected through various information businesses, there are only few instances that effectively use this and studies that analyze organically from using the big data (Junghoon Moon, 2013).

The purpose of this study is to verity that new values that practically help the agricultural food area can be

1 This study was supported by the research project (project number: PJ008729) of Rural Development Administration

created by combining the various types of the big data collected in the agricultural food area and analyzing them in an appropriate analysis method. By introducing the three instances of the 'big data' analysis related to the agricultural food area, the study will show how the extracted data could be utilized.

2 Big data

2.1 Definition of Big data

According to the report of McKinsey & Company published in 2011, the 'big data' is defined as an ultra-mass storage data created by condensing tremendous amounts of data at once. However, this report judged that the most important factor of the 'big data' is not its capability to capacitate tremendous amounts of data, but its possibility to process the existing data through a processing method (McKinsey& Company, 2011).

It can be assumed, from the definition mentioned earlier, the biggest advantage of the 'big data' is its potential to collect data from almost every area and its significantly high reliability of the analysis even from its tremendous size. Above all, it has an unlimited potential because it can record and store all the world's phenomena which increases its importance (Changkeun Yoon, 2012).

However, it is not only important to compile or collect 'big data', but it is also vital to extract necessary information from the 'big data' and fuse them together. A variety of public information is showcased under Jeongbu3.0 to civilians. Through this, the mass storage data is opened and easily shared, but there is difficulty to effectively utilize it. First of all, a linking ability from collecting the necessary data among numerous amounts of data is required, and it is essential to have the technology to process the big data collected by this way. However, if these conditions are satisfied, anyone can produce profit by developing a distinguished service which could provide a convenient service that customers need from 'big data' that are open to the public. For this reason, whether possessing the technology that can utilize the big data in the future or not will be one of the most important factor that decides a national or corporate competitiveness.

2.2 Analytical techniques of the 'big data'

Typical analytic techniques of the big data consist of Text Mining, Opinion Mining, Social Network Analytics, and Cluster Analysis (Sungwoo Choi, Hoyoung Kim, & Youngkook Kim, 2012).

2.2.1 Text Mining

Text mining means extracting the necessary information or knowledge from the data created by text. As 80% of the information that companies possess is stored in the form of a text document, texts are the most common means to store information. However, in order to extract any necessary information from texts created by atypical types of data, various text analytic techniques should be applied. Different techniques such as information retrieval, text analysis, information extraction, clustering, categorization, etc., are necessary depending on the situation. A standard of judgment for data categorization is created in this procedure so that the necessary information and knowledge can be obtained (Tan, 1999).

2.2.2 Opinion Mining

Mining technology means the entire technology used to extract the core information that is valuable to an individual or a company from a tremendous quantity of information existing online like 'big data'. Opinion mining is also called a 'Sentiment Analysis' for it is a kind of analytic technique that grafts the mining technology to a technique that process natural language (Esuli & Sebastiani, 2006). Through the opinion mining technique, it is possible to judge corresponding incidents and preferences about products from texts arbitrarily created by the public such as various social media, news comments, or product reviews. This way, the public preference is divided into a positive, a negative, or a neutral state, and it becomes possible from constructing a vocabulary that will become the standard of categorization (Pang & Lee, 2004; Turney, 2002). Like this, the main purpose of using opinion mining is to read the responses of the public by categorizing the public preference into specific incidents or products.

2.2.3 Social Network Analytics

Social network analysis refers to the analysis of individual reputation and impact based on the connection structure of social network. In recent years, the major research trend has changed from focusing information from individuals to studies that focus on the connection network itself which is structured by relationships and interactions between people (Sangwon Park et al., 2009). The revolutionary development of smart devices and rapid distribution of social network allow individuals to communicate with an unspecific crowd with their experience or opinion. Thanks to this, individual users can exert their own reputation and influence through social network based on "participation, communication, and sharing."

By objectifying these, the influences, concerns, inclinations and action patterns of individuals or groups can be revealed in social network. The companies are already using this for their business by analyzing big data through SNS data analysis technique. However, it is necessary to pay attention to this analysis technique because it has the possibility to intrude personal privacy. In order to minimize the side effects, companies should take social responsibilities and practice methods such as self-monitoring or risk relief programs and emphasize the importance of these techniques to the public (Jeongsook Kim, 2012).

2.2.4 Cluster Analysis

The purpose of cluster analysis is ultimately to figure out the groups with similar characteristics by combining individuals with similar characteristics. For example, it is possible to categorize groups with common interests or hobbies through the cluster analysis (Manmo Kang, Sangrak Kim, & Sangmu Park, 2012).

2.3 Examples of the application of 'big data'

As the key example of the big data use, there is the service, Google Flu Trends. This service provided by Google warns of flu by predicting the flu occurrence in the U.S. This service started from the hypothesis in which people would more often search the flu related words when they have flu and conducted by using the big data, Internet search word log data of Google. By analyzing billions of Internet search word logs, the pattern among the flu related search words was extracted and the area where flu occurs was predicted; this data was compared and analyzed with the flu monitoring data of US Centers for Disease Control and Prevention, the result showed almost complete coincidence (Nature, 2008).

One of the most common 'big data', public data means the data possessed by the government or a public institute which is accumulated during the middle of public institute business procedures. Because this is directly connected to the public service of the government, this data has a higher relative value than other data. Advanced countries such as USA or England are already using these data to improve the quality of the public service (Manjae Lee, 2012).

By beginning with the slogan of 'open, share, communication, and cooperation,' Jeongbu3.0 makes it easier to access public data. Also, if these data are combined with a data that a company owns such as social data, image data, location information, and portal data, it can be enabled to become utilized in various areas. The city of Seoul has a plan to create new late night bus routes by promoting a demonstration project to draw optimum late night bus routes with KT. To achieve this, the route administration data for the city of Seoul and the floating population analysis data of KT were combined in order to for it to be used (SisaJeju, 2013).

One of Korea's well known navigation service, T-map, is a service provided by SK Telecom from combining their customer information with the public data from the Korean National Police Agency, the Korean National Oil Corporation, and each local government. The simple navigation functions as well as a variety of convenience services are provided from obtaining the main street CCTV information from the Korean National Police Agency, the gas price of national gas stations from the Korean National Oil Corporation, and the real time arrival information of the buses from the local governments.

Besides the public area, there are continuous efforts for 'big data' utilization in various areas. The Ministry of Future Creation and Sciences plans to select and perform the 2013 'big data' business consulting project in employment, youth welfare, small business aid, and health and medical areas. Prediction of jobs that will be on supply and demand, early warning systems for youths at risk, free market space for small and medium distributors, and personalized health information offers are four projects that were selected. Each project will be conducted mainly by the Ministry of Employment and Labor, the Ministry of Gender Equality and Family, the Korea Chamber of Commerce and Industry, and the Health Insurance Review and Assessment Service (Ministry of Science, ICT and Future Planning, 2013).

3 The big data analysis of agricultural food area

3.1 Current status of the 'big data' used in the agricultural food area

Various kinds of 'big data' are already accumulated in the all processes of agricultural food area such as production, processing, distribution, etc. While there are data with low utilization, some data can be immediately utilized through a simple analysis. In reality, information obtained through techniques, such as the RFID sensor, in the production step is not highly utilized compared to its utilization. A form that analyzes and exchanges information among materials without any human manipulation or effort from using the food history tracking management system and IoT technique will be a very important issue in the agricultural food area. Also, 'big

data', such as price information, obtained in the distribution step will create huge benefit depend on the usage in the future.

Previously, there were fewer efforts to utilize the big data even though these big data had been collected. However, with the increasing concerns from the government about 'big data', a variety of 'big data' demonstration service businesses are being conducted. Although the agricultural food area was not included in the big data demonstration service businesses selected by the government on June, 2013, the Ministry of Food Agriculture, Forestry and Fisheries as well as the agricultural administration is preparing the ICT convergence technique that utilizes by combining various big data including agricultural food statistic data.

3.2 The big data analysis instances in the agricultural food area

3.2.1 The study on the regional difference of food consumption pattern of urban families in China

The main purpose of the study (Siyoung Lee & Myungsook Park, 2012) is to increase the accuracy in the study about regional food consumption pattern in China and extract the data usefully utilized for China related policy makers, food traders, and etc.

The study used the big data, Per Capita Annual Food Consumption Expenditure of Urban Households by Region from 2004 to 2009 published in the statistical year book of China. The regional food consumption pattern of urban families in China was found by analyzing that. For the analysis methods, cluster analysis and time series regression analysis of the regional food consumption panel data were used.

Table 1 shows the regions in China and food classification items by listing. China is divided into 31 regions in total, and food is classified into 19 items. However, food is classified into five items according to the level of importance and interconnectivity because food has an interdependent relationship.

Table 2 Table of food categorization according to the importance and interconnectivity (Siyoung Lee &Myungsook Park, 2012)

Category	Sector
Grain & beans	grain, starches and tubers, beans and bean products
Animal food	milk, meat, poultry, eggs and processed products
Aquatic food	aquatic food
Vegetables	vegetables
Fruits	dried and fresh melons and fruits

Table 2 shows five items grouped by their interconnectivity of food consumed in urban households in China. The items are 'Grain and Beans,' 'Animal Food,' 'Aquatic Food,' 'Vegetables,' and 'Fruits,' and detailed items are also listed.

By using the panel data about the regional food consumption expenditure for six years of 31 regions in China, cluster analysis was performed by setting the adequate number of cluster complexly considering the change of pseudo-F and CCC. Additionally, the regression model was inferred as a dummy variable when it included the cluster that was created to distinguish the difference of consumption pattern of each category depending on groups. As a last step, the regression model was reassumed as an error component model in which the hypothesis was that the model interception was shaped by the regional and time effect of the panel data

The result of this study can be classified into three major categories. First, the food consumption pattern in China was classified into five clusters. Additionally, the most important factor deciding the food consumption expenditure in the entire category was the real disposable income per capita, and the entire expenditure increased together when the real disposable income per capita increased.

Also, there are differences in the income elasticity depending on clusters. Because the real consumption ex-

Table 1 Regions in China and food classifications (Siyoung Lee &Myungsook Park, 2012)

Region	Food
Beijing, Guangxi, Ghongqing, Sichuan, Tibet, Tianjin, Liaoning, Jinagsu, Anhui, Jiangxi, Shandong, Hubei, Hunan, Guizhou, Zhejiang, Fujian, Hainan, Hebei, Shanxi, Inner Mongolia, Jilin, Heilongjiang, Henan, Yunnan, Shaanx, Gansu, Qinghai, Ningxia, Xinjiang, Shanghai, Guangdong	grain, starches and tubers, beans and bean products, oil and fats, milk and processed products, meat, poultry and processed products, eggs, aquatic products, vegetables, dried and fresh melons and fruits, dining out, tobacco, liquor and beverages, condiments, sugar, cake, food processing service fees, etc.
Total: 31	Total: 19

penditure change ratio of most real disposable income per capita stays the level of 0.89~4.12%, food consumption of clusters depending on the category becomes an income inelasticity. However, in specific clusters, income elastic consumption expenditure change sometimes occurs.

Lastly, it can be drawn that food consumption pattern of the clusters depending on the category is distinguishable from the significance of the coefficient of dummy variables presented in clusters. Here, it is shown that the factors that decidethe food consumption pattern depending on the category are not monolithic, but the different regional factors depending on each region somewhat exist.

Fig. 1 visualizes the results drawn by the cluster analysis of Per Capita Annual Food Consumption Expenditure by Region in 2009 in the map of China. Through the distribution types of each cluster, it is possible to infer that there are influences by regional characteristics. The regional characteristics include geographical characteristics such as climate, soil fertility, etc. as well as culture, history, industry structure, etc. Cluster II is mainly located in the northern area, and Cluster III is located in southern area in China. On the other hand, Cluster IV and Cluster V are located near the sea, and the per capita income of these regions is higher than that of other regions.

Fig. 1 Cluster analysis result of the panel data in 2009 (Siyoung Lee & Myungsook Park, 2012)

Table 3 Average income elasticity by region between 2004 and 2009 (Siyoung Lee & Myungsook Park, 2012)

Region	Cluster by Year						e^1_i	e^2_i	e^3_i	e^4_i	e^5_i
	04	05	06	07	08	09					
Hebei	1	1	1	1	2	2	0.4030	0.6542	0.7638	0.5312	0.7132
Shanxi	1	1	1	1	1	2	0.2984	0.7414	2.0619	0.6185	0.8223
Inner Mongolia	1	1	1	1	2	2	0.3691	0.7091	1.6657	0.7911	0.7590
Jilin	1	1	1	1	2	2	0.3262	0.6450	0.7092	0.4982	0.5148
Heilongjiang	1	1	1	1	2	2	0.2955	0.6313	0.7484	0.6229	0.5739
Henan	1	1	1	1	2	2	0.3450	0.6732	1.7558	0.6031	0.8076
Guizhou	1	1	1	2	2	2	0.3652	0.4821	1.5119	0.5043	0.6467
Gansu	1	1	1	1	2	2	0.3348	0.6640	1.5283	0.5240	0.5784
Qinghai	1	1	1	1	2	2	0.3179	0.5730	1.3201	0.5875	0.6602
Ningxia	1	1	1	1	2	2	0.3337	0.6598	1.6253	0.6221	0.5835
Xinjiang	1	1	1	1	2	2	0.3207	0.5538	1.3157	0.6028	0.5400
Liaoning	2	2	2	2	3	3	0.3384	0.5593	0.3777	0.4690	0.4918
Jiangsu	2	2	2	3	3	3	0.4930	0.6320	0.4315	0.6603	0.8711
Shandong	1	2	2	2	3	3	0.3485	0.4801	0.6176	0.5052	0.6948
Hubei	1	1	1	2	2	3	0.3321	0.5498	0.5399	0.4276	0.7787
Hunan	1	2	2	2	3	3	0.3514	0.5438	0.6467	0.5374	0.6543
Guangxi	2	2	2	2	3	3	0.3938	0.3998	0.4513	0.5896	0.7293
Hainan	2	2	2	2	3	3	0.4591	0.3888	0.1915	0.4873	0.7630
Chongqing	2	2	2	3	3	3	0.4164	0.4567	0.7515	0.5096	0.7656
Sichuan	1	2	2	2	3	3	0.3474	0.4093	0.9188	0.4395	0.6837
Yunnan	2	2	2	2	3	3	0.3664	0.5953	0.9861	0.4414	0.7102
Tibet	3	3	2	3	3	3	0.2776	0.3575	1.6167	0.3700	0.7359
Shaanxi	1	1	1	2	2	3	0.3197	0.7039	1.4062	0.5836	0.5352
Beijing	3	3	4	4	4	4	0.5944	0.7628	0.9333	0.9712	0.6334
Tianjin	2	3	3	3	4	4	0.5191	0.7071	0.4350	0.7142	0.6299
Zhejiang	3	3	4	4	4	4	0.6118	0.9085	0.3042	0.7862	0.7781
Fujian	3	3	3	3	4	4	0.3871	0.5289	0.1795	0.6147	0.7577
Shanghai	4	4	4	5	5	5	0.5821	0.7598	0.3095	0.7896	0.7429
Guangdong	3	3	4	4	4	5	0.4875	0.4951	0.3239	0.6519	0.8227

The result of the time series regression analysis of the regional food consumption panel data is as follows:

Except Cluster I, there was no significant difference among clusters in 'Grain & Beans.' When considering that Cluster I was not shown in all regions in China in 2009, it might be expected that 'Grain & Beans' was similarly consumed in all regions in China. It also can be thought that this category was used for all Chinese dishes as the fundamental material. 'Animal Food' was least used in Cluster I and II. 'Aquatic Food' was most used in Cluster IV. Additionally, in Cluster III and IV, much more 'Vegetables' and 'Fruits' were consumed than in Cluster I and II.

Lastly, it was possible to obtain the average income elasticity by region of each category between 2004 and 2009.

Table 3 shows the average income elasticity by region for six years between 2004 and 2009. The analysis of this table shows that in the categories of 'Grain & Beans' and 'Vegetables,' Cluster IV and V showed significantly higher income elasticity compared to other Clusters. Additionally, the regions included in Cluster II have a significantly high income elasticity about 'Aquatic Food,' and the reason for this could be explained as the difficulty of 'Aquatic Food' delivery due to the low accessibility to the sea.

The researchers of this study are expecting that these findings can be usefully applied for the future policy makers related to China and food traders.

3.2.2 Analytic study on the Psychological recovery of a consumer after the carcinogen incident

This study (RDA, 2013) showed the influence on the consuming reduction due to the crisis related to the health issue about food safety and the recovering pattern and period through the 'big data' analysis and also examined the instance of ramen benzopyrene content that occurred in 2012. The Twit big data collected by the web crawling technique and the agricultural food purchasing history big data collected as a part of consumer panel investigation were used, and the time for recovering the consumer repurchasing psychology and influence were analyzed.

The N company carcinogen incident became known when the TV solely reported the N company benzopyrene incident on October 2012. The main content of the TV report was that 0~4.7ppb benzopyrene per 1kg of seasoning was found in six products when Korean Food and Drug Administration tested the benzopyrene content from the N company's udon noodles and ramen products that were randomly collected. In addition, the problem of this incident was escalated because it was also shown that the newspaper did not take any action for four months even though they had already known the fact of the detection. When the public criticism about that incident was getting stronger, Korean Food and Drug Administration explained that the level of benzopyrene content found was harmless. However, the public criticism was not easily subsided.

This study was conducted focusing on the procedures of the time until the consumers recovered their consumption psychology about the N company's products after the series of the incident. The following figure shows the analysis procedure.

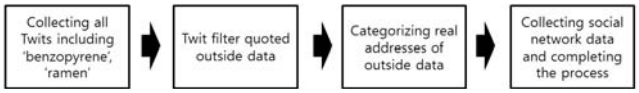

Fig. 2 The analytic study procedure direction of the consumption psychology recovery after the N company carcinogen incident (RDA, 2013)

First of all, the total 4,625 Twits including keywords, 'ramen' and 'benzopyren,' were collected by using the social matrix from October, 2012 when the benzopyrene issue occurred to December, 2012. The social matrix is a tool that helps analyzing blogs and Twitter documents based on the natural language processing technique and text mining technique. This service is provided by Daumsoft which is a Korean text mining specialized company. This tool also provides both positive and negative emotional information about each keyword as well as the most spread Twit messages depending on the date. It analyzes more than 100 million Twits generated per year from 2400 thousand Twitter Korean accounts and more than 100 million blog posts and searches conspired meanings and reputations (ETRI, 2013).

Among these collected data, Twits quoting outside data such as news, blogs, etc. are selected. In this procedure, each Twit is classified depending on the essential addresses of real addresses. In addition, the sources of Twit quoting outside data related to benzopyrene as a final output and the number of times quoting them were structured as a database depending on the date.

The following graphs are drawn through a series of procedures.

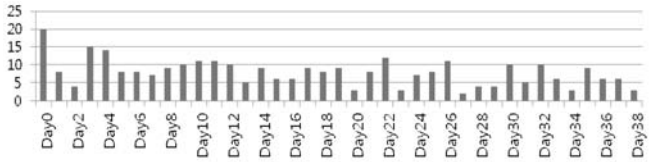

Fig. 3 Time it took until consuming was restarted(RDA, 2013)

The above Fig. 3 visualizes the time it took for consumers to repurchase N company's ramen in the entire sample. In the standard date (Day 0) when benzopyrene incident was reported in October, the time of repurchasing over time is presented as the graph; the analysis result shows that 237 people among the entire panel 543 people did not purchase the N company's products until after December.

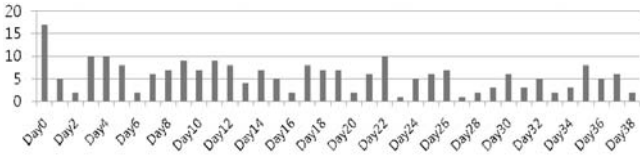

Fig. 4 Time it took until the consumption restarted (the main media, "Besides Internet," group) (RDA, 2013)

The above graph of Fig. 4 visualizes the time it took until 398 people whose main media was not the Internet, repurchased N company's ramen. In the analysis result, 175 people, almost half of the people, did not purchase the N company's ramen until December.

By using a discrete time hazard model with time-varying covariates, the result was analyzed concerning the time for recovering their psychology to repurchase after the incident and its influence. The following graph of Fig. 5 was created by simulating the 200 days after the incident through the Hazard model analysis.

The graph shows that the consumers started to repurchase the N company's products after the 80 days of the N company benzopyrene incident. In addition, the types of media providing information influenced the consumers' repurchasing. Among Blogs or newspaper media, the period of repurchasing the N company's products of the customers who obtained information from Chosun Ilbo and Kyunghyang Shinmun was relatively short. On the other hand, the broadcast media report tended to decrease the intention to repurchase the N company's products, and this tendency was shown more particularly in SBS.

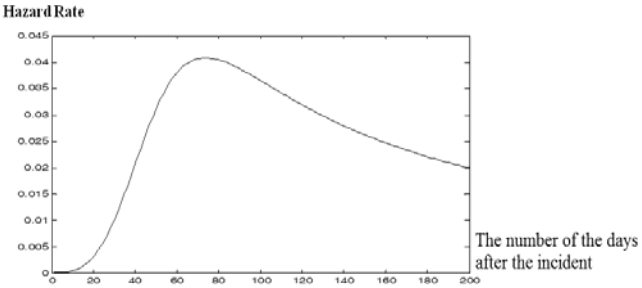

Fig. 5 Simulation results using the Hazard Function (RDA, 2013)

The noticeable point is that the demographic characteristics did not show a major influence on the repurchasing period. There were no significant differences in the non-bold font parts of Table 4 including monthly income, age, the number of family member, etc.; lastly, the repurchasing period was longer when there were more children, and full-time homemakers tended to repurchase earlier.

In this study, it was verified that the consumers controlled their food purchasing behaviors influenced by various types of media, and they tended to be influenced

Table 4 Hazard model table and analysis result (bold font means a significant difference) (RDA, 2013)

Variable	Est.	S.E.	Variable	Est.	S.E.
Visiting supermarket	0.0046	0.0066	More than 3,500,000KRW monthly income	0.010	0.0060
Portal site report	**0.0063**	**0.0002**	Age	-0.001	0.0004
ChosunIlbo news	**0.2871**	**0.0106**	The number of children	**-0.018**	**0.0054**
JoongangIlbo news	**-0.0150**	**0.0072**	The number of family members	-0.001	0.0041
Donga Ilbo news	**-0.0427**	**0.0031**	Full-time homemaker	**0.014**	**0.0059**
Hankyoreh Shinmun news	**-0.0338**	**0.0026**	College/University degree and higher	0.001	0.0067
Kyunghyang Shinmun news	**0.7326**	**0.0381**	Frequently reading health articles	0.005	0.0065
MBC report	**-0.0939**	**0.0030**	Using Twitter	0.000	0.0371
SBS report	**-2.3165**	**0.1211**	Major media: TV	**0.016**	**0.0069**
KBS report	**-0.0142**	**0.0017**	Major media: newspaper	0.015	0.0114
Blog information	**0.1010**	**0.0050**	Major media: Internet	0.010	0.0082
			Government trust	-0.013	0.0191

by trusting corresponding media particularly when they obtained the information from TV media. It shows the importance of the reliability of the media itself and the information conveyed by the media in consumer behaviors.

3.2.3 Korean National Food Cluster: The analysis report of food market in Korea and Northeast Asia (Korean National Food Cluster, 2013)

The purpose of the study (Korean National Food Cluster, 2013) was to promote the development potentials to overseas food companies. The analysis report written by this study can be shown at the resources part of the Korean national food cluster website; it is expected to additionally analyze the bakery chain market, beer market, chocolate market, coffee bean market, and meat processing food market in the future.

An infrastructure construction business, the Korean national food cluster that is creating a mutual synergy effect by directly connecting company/research institutes with other related institutes, is performed as a part of the business to develop Korea as the center of the Northeast Asia food market that is rapidly growing.

In this study, the food related household expenditure panel data provided by RDA was analyzed and used. The corresponding data were collected from 700 households of the capital region of Korea between 2010 and 2012. By analyzing the corresponding data through the opinion mining, the fermented milk "Culture Code" and the purchasing trend information according to consumers' age, the number of children, income, and quarter of the year were extracted. The analysis data extracted by this can be used to establish marketing strategy when overseas companies enter the Korean market.

Table 5 Fermented milk 'Culture Code' analysis by using the opinion mining technique (Korean National Food Cluster, 2013)

Related keywords	Major brands	Emotional Keywords
ingredients taste food breakfast meal lactobacillus fruits constipation solution plain losing weight	Vulgaris Namyang Pororo Pure Pasteur	help effective fresh like regular white well-being encourage happy soft

Table 5 shows the analysis contents of the consumers' "Culture Code" about the fermented milk extracted through the opinion mining technique by using the social matrix tool. When analyzing this, consumers showed an overall positive feeling including 'help,' 'fresh,' 'well-being,' 'happy,' etc. They also showed a high interest in the functionality of fermented milk such as 'ingredients,' 'lactobacillus,' 'constipation solution,' 'losing weight,' etc. Based on this result, it was shown that consumers had a positive feeling overall and they recognized fermented milk both simply as a drink and as one of the functional foods. In addition, there is the possibility that the consumption of fermented milk can be expanded as an alternative meal since the search words including 'food,' 'morning,' 'meal,' etc. were shown.

Additionally, when categorizing the opinion mining analysis results depending on the brands accordingly to the product forms for liquid type fermented milk, the interest for additional information such as the related company was higher than the product quality, and for functional food, the interest for the product quality was high, and the emotional attitude about the taste was noticeable in the type of semi-liquid fermented milk.

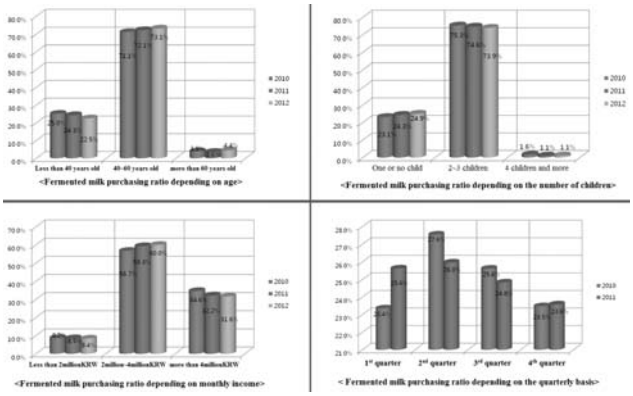

Fig. 6 Korean fermented milk consumer behaviors analysis (Korean National Food Cluster, 2013)

Fig. 6 shows the visualization by analyzing the consumers' fermented milk purchasing inclination information in the Korean fermented milk market according to their age, number of children, income, and quarter of the year.

In the consumers' fermented milk purchasing behavior depending on the characteristics, the fermented milk purchasing ratio of more than 40 years old and less than 60 years old consumers, households with 2~3 children, and consumers whose monthly income was 2million ~ 4million KRW was highest. Among them, the fermented milk purchasing ratio of more than 40 years old and less than 60 years old consumers and consumers whose monthly income was 2million ~ 4million KRW was continuously increasing, and this shows who the major consumers of the fermented milk market in Korea

were. While the purchasing frequency depending on the quarterly basis of the year was highest in the 2ndquarter and 3rdquarter, the 1stquarter was also rapidly increasing.

Based on the report using this type of 'big data', it seems that Korean consumers fundamentally have a positive recognition about fermented milk. In addition, it is shown that fermented milk was simply regarded as a simple snack in the past but the recognition about it was interchanged by regarding it as an alternative meal or a functional food. Here, it is expected that Korean fermented milk market size will keep increasing. Through the purchasing ration depending on age, the number of children, and income, it is possible to target a consumer class when entering to the Korean fermented milk market (more than 40 years old, 2million ~ 4million KRW monthly income, one or no child)

4 Conclusion

As this study shows, the 'big data' utilization in the agricultural food area is still in an early stage. While the 'big data' utilization in economics, business administrations, and other industries is apparently increasing, the application in the agricultural food area is still marginal. However, as analyzed in this study, valuable results can be drawn through a combined analysis of the 'big data' about the fermented milk market, different products, and customer panel data. Many more meaningful results can be expected if the 'big data' is actively used in the future agricultural food area.

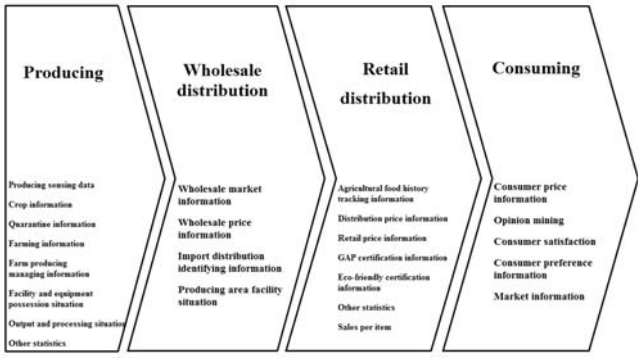

Fig. 7 Big data in the agricultural food area

Fig. 7 above is a concise value chain showing the 'big data' in the agricultural food area that can be obtained through various techniques and public data. The related 'big data' information obtained across the overall agricultural food area including production, wholesale, retail, consumption, etc., can be an opportunity to produce valuable data with the stream of Jeongbu3.0 starting with the slogan of 'opening, sharing, communicating, cooperating.'

In addition, as the concept that is rapidly increasing currently and deeply connected to the big data in the agricultural food area, the concept of IoT (Internet of Things) means that all physical things can become virtually the computer connected to Internet (ITU, 2005). This also means that all things can play the role as the small computer connected to Internet rather than being a physical computer (Fleisch, 2010). Particularly, with the occurrence and utilizing technique development of the 'big data' such as the current RFID technique and IoT technique, the data to utilize are continuously produced in the agricultural food area, including producing history tracking by utilizing RFID, automatic information input in the producing stage by using the IoT technique, automatic distributing information in the distribution stage, price information in the processing stage, panel data in the consuming stage, etc. How to utilize this is a very important question in the future of the agricultural food area; if appropriately utilized, it can become an opportunity to create new additional values in the agricultural food area.

Additionally, through Jeongbu3.0 promoted by the government, it has tried to change the paradigm toward the direction in which the public information should be actively opened to the public. Unlike the past in which a two-way communication was pursued by regarding consumers as "prosumers" and providing those who were targets of service with the only manufactured information, through opening information on an open platform that is in a cooperative space, consumers are regarded as "cresumers" of active resources who create new values in the framework where the government suggests values and the consumers decide on values. With this, there are many types of the big data that have high application potentials in the agricultural food area. While the data can be obtained from the producing and contributing stage, there are tremendous types of big data about producing, processing, contributing, selling, etc. possessed by the system of institutes. However, the level of utilization of these data is relatively low compared to that of the accumulated data, and the big data has not been combined or synthetically used. In this situation, if the information of organizations can be shared and the big data in the agricultural food area that is scattered everywhere can be effectively used through the Jeongbu3.0, the development in the agricultural food area and new values can be created. There will be significant contribution in analyzing the continuously generated 'big data' in the agricultural food area from the producing stage, all the way to the consuming stage.

REFERENCES

Manmo Kang, Sangrak Kim, & Sangmu Park. (2012). Analysis and Utilization of Big Data, Journal of Computing Science and Engineering, 30(6), 25-32.

Korean National Food Cluster (2013). Korean National Food Cluster: The analysis report of food market in Korea and Northeast Asia.
(http://www.foodpolis.kr/dnb/resource/main/index.php)

Jeongsook Kim. (2012). Big data Utilization and related Technique and Technology Analysis. Journal of Korea Contents Association, 10(1), 34-40.

Rural Development Administration. (2013). 2013 Agricultural food consuming trend conference data.

Junghoon Moon. (2013). The current and future status of Jeongbu3.0 and ICT convergence big data in the agricultural food area.2013 International Symposium on ICT Convergence in the Food and Agriculture Industry: the Presentation Material about the Use of Smart Value Chain and Big Data.

Ministry of Science, ICT and Future Planning (2013). A press release of Ministry of Science, ICT and Future Planning.
(http://www.msip.go.kr/www/brd/m_211/view-.do?seq=615)

Sangwon Park, Eunjung Choi, Minsoo Park, Jungkyu Kim, Eunseok Suh, &Youngtaek Park. (2009). An Expert Recommendation System using Ontology-based Social Network Analysis. Journal of Computing Science and Engineering, 15(5), 390-394.

Sisajeju. (2013) Utilization of Big Data, Change the life. (http://www.sisajeju.com/news/articleView.html?idxno=201807)

Changgeun Yoon. (2012). Into the world: Worldwide trend: America: Grafting artificial intelligence technology and big data and the possibility to use them, looked through IBM Watson. Journal of Korean Association Regional Information, 74(Single volume), 68-91.

Manjae Lee. (2012). Big Data Analytics and Utilization of Public Data, Journal of Computing Science and Engineering, 30(6), 33-39.

Siyoung Lee, & Myungsook Park. (2012). An Analysis on Regional Differences in the Food Consumption Patterns of Chinese Urban Households. 30(1), 87-111.

Sungwoo Choi, Hoyoung Kim, & Youngkook Kim. (2012).Research Trend of Bigdata Technology and Analysis Technique, 39(2A), 194-196.

Esuli, Andrea, & Sebastiani, Fabrizio. (2006). Sentiwordnet: A publicly available lexical resource for opinion mining. Paper presented at the Proceedings of LREC.

ETRI. (2013). Social big data issue detection and predictive analysis.

Fleisch. (2010). What is the Internet of Things? Auto-ID Labs White Paper.

ITU. (2005). “The Internet of Things”. ITU Report, Genf

McKinsey & Company. (2011). Big Data: the Next Frontier for Innovation, Competition and Productivity

Nature. (2008) Web data predict flu.
(http://www.nature.com/news/2008/081119/full/456287a.html)

Pang, Bo, & Lee, Lillian. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Paper presented at the Proceedings of the 42nd annual meeting on Association for Computational Linguistics.

Tan, Ah-Hwee. (1999). Text mining: The state of the art and the challenges. Paper presented at the Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases.

Turney, Peter D. (2002). Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Paper presented at the Proceedings of the 40th annual meeting on association for computational linguistics.