Case Study on Big Data Sampling Population Collection Method Errors in Service Business

Ahn, Jinho*, Lee, Jeungsun**

Content

ABSTRACT

As big data become more important socially and economically in recent years, many problems have been derived from the indiscriminate application of big data. Big data are valuable because it can figure out the meaning of informative information hidden within the data. In particular, to predict customer behavior patterns and experiences, structured data that were extracted from Customer Relationship Management (CRM) or unstructured data that were extracted from Social Network Service(SNS) can be defined as a population to interpret the data, during which many errors can occur. However, those errors are usually overlooked. In addition to data analysis techniques, some data, which should be considered in the analysis, are not included in the population and thus do not show any meaningful patterns. Therefore, this study presents the measurement and interpretation of the data generated when the cause of error in the population setting is strong relationship and interaction between people or a person and an object. In other words, it will be shown that if the relationship and interaction are strong, it is important to include data collected from the perspective of user experience and ethnography in the population by comparing various cases of big data application, through which the meaning will be derived and the best direction will be suggested.

Keywords: Big Data, Service Business , thick data, user experience, ethnography

1. Introduction

The value of use in big data is to figure out the meaning of informative information hidden within the data. If the hidden information has a repeating pattern, it is straightforward to figure that out. However, if not, then it becomes problematic.

When feature phones were popular, the pattern of popularization of smart phones such as the iPhone was not apparent in big data analysis. The data analysis of global consulting firms and most feature phone manufacturers could not find any patterns during that time. The pattern was not found because data, which should appear in a meaningful pattern, did not appear in the population, not because there is a problem in the data analysis technique.

As a big data becomes more important socially and economically in recent years, many problems have been derived from the indiscriminate application of big data. In particular, to predict customer behavior patterns and experiences, structured data that were extracted from Customer Relationship Management (CRM) or unstructured data that were extracted from Social Network Service(SNS) can be defined as a population to interpret the data, during which many errors can occur Therefore, this study presents the measurement and interpretation of the data generated when the cause of error in the population setting is strong relationship and interaction between people or a person and an object. In other words, it will be shown that if the relationship and interaction are strong, it is important to include data collected from the perspective of user experience and ethnography in the population by comparing various cases of big data application, through which the meaning will be derived and the best direction will be suggested.

2. Literature Review

2.1 Definition of big data

2.1.1 Data Concept and Research of Big data

The concept of big data has not been exactly defined according to various literature. However, with the development of the Internet and computing, the evolution of mobile devices and sensors, and the development of large social networks, vast amounts of digital data have emerged inside and outside government and corporations, and these data include unstructured data, such as text that is generated from them, and structured data, such as records of large-scale data exchange. Broadly, the technology and architecture to support these data are also called big data.

With the advent of era of big data, effectively utilizing and responding to enormous data have been emerging as a core technology for companies and nations.

People appreciate the potential of big data analytics as a solution to many of our society and business problems. They can use data to make better judgments, to help them live healthier, to plan effective cities, to design products that can sell better, and to create safer societies.

However, because thinking and methodology of

big data cannot be the correct answer for analyzing all data, a proper solution is needed for this. To date, most of the studies have been discussing the problem of using big data in terms of privacy infringement only. However, a discussion seems to be required in terms of the initial data generation and collection processes.

2.1.2 Collection of Big Data

In order to collect data, a collection procedure needs to be designed and sufficient testing should be conducted. Because data collection determines the quality of the service, if a serious problem occurs during the data collection process, it may be necessary to redesign the entire project. To establish a collection plan, the characteristics of the project, in which data collection is involved, should be examined first. Furthermore, the procedures and considerations for data collection should be examined as well.

Considering the importance of the data collection in a big data project, the data collection process itself does not generally determine the quality and success of the project. It is more related to data management rather than on data acquisition. However, in the case of a project that includes data collection procedures, data acquisition often determines the quality and success of the project.

2.1.3 Sample calculation of a Population of Big Data.

A population is the total collection of target subjects from whom information needs to be acquired. Because the population is defined differently depending on what we want to know, it

is very important to define them clearly. A sample population, on the other hand, is a subset of the population and extracted from the population.

The collection of big data can be interpreted as the process of generating samples from a population. In other words, if there is an entire object that needs to be identified through big data analysis, some of the selected (extracted) part in it become a sample that is a category of actual big data collection.

2.2 Failed and Successful case studies of Big data application

2.2.1 Failure of User Experience Analysis in Windows 8

Although more than $100 million were invested to develop Windows 8, this system is one of the failed products. Among various reasons, the biggest reason is that the start button at the bottom left, which is the key feature that Windows-based OS users were familiar with, was eliminated.

Despite innovative attempts including the Metro UI, the system failed because they ignored the user experience, which causes the most interaction in the use of the operation system.

Microsoft has named the next version Windows 10 instead of Windows 9 to show that Windows 10 is completely different from Windows 8. BGR, which is an IT medium, said Microsoft overemphasized touch and gestures, but failed to satisfy users who were traditionally familiar with the keyboard and mouse-based interfaces.

The Operation System is the beginning and end of the relationship between a person and a

computer. In order to use certain IT services, accessing the operating system in a general PC or laptop requires a relationship regardless of the user's choice. Under this relationship, suitable software or hardware is provided by the operating system and used by the user, and interaction is continuously performed to transmit desired messages to the software or hardware. In other words, Windows OS has a strong relationship and interaction with the user, and it is difficult to understand the relationship only by analyzing external data, such as logs of preferred user interfaces, functions, and services.

By collecting and interpreting the data, it was possible to derive more important meanings and to understand how the main users communicate with the Windows OS by watching the actual user environment.

2.2.2 Failure Nokia's market research analysis

There have been cases in Nokia where the procedures and processes of observing these actual offline users and interpreting meaningful data therein were ignored.

Nokia collected extensive data on the ripple effects of smart phones after iPhone was released. They collected data on purchase trends and behavior patterns from more than 100 million customers worldwide and matched them with various indicators of consumption trends, including validated consumption levels. Along with Nokia, the major feature phone companies such as Samsung Electronics collected and analyzed relevant big data. However, the data they analyzed did not provide the meaningful value that

smart phones will dominate the phone market. Because the price of the iPhone was $ 599 when it was launched for the first time, it seemed unlikely that an average person would buy an iPhone over a feature phone, which is a third of the price.

However, in 2009, Tricia Wang, who was a Nokia researcher and an ethnographer, observed 100 Chinese low-income workers and concluded that they can actually buy the iPhone.

She lived with locals for many years and shared everything with them. She conducted an observation study through qualitative data, and concluded that even the poorest people in China want to buy smart phones intensely and the most desired smart phone was the iPhone.[1]

Based on the above process, it can be found that the relationship between a person and a feature phone has a different strength of interaction from the relationship between a person and a smartphone. In other words, the psychological and emotional connection strengths of the smartphone showed a stronger relationship in terms of emotion and attitude.

2.2.3 Comparison between Dramas Productions in Netflix and Amazon

"The history of TV is now divided before and after the House of Cards.", said a TV critic in the US, when Netflix released its first own series drama. 'House of Cards' premiered 13 episodes at once. The term 'binge watching', which was used only among young people, was widely known for this (Jung-ang Ilbo, 2016).[2] House of Cards also received the highest ratings (9.1 out of 10) from critics.

The success of this drama is because Netflix realized the importance of the relationship and interaction between a viewer and a TV object in the perspective of consuming Netflix TV dramas and thus instead of relying on traditional big data analysis, they used an approach that extends the category of factor data to be considered.

For Netflix's services, viewers can unsubscribe at any time. In addition, through the monitor screen, formations of relationship and role for meeting the needs of viewers have an important relationship. There is also a strong interaction of initiatives for finding alternatives to or transforming content, timing, and sequential flow in searching and determining the program that views want to watch. Therefore, in selecting the TV program, trusting only the significant results of big data is not enough.

In this regard, Netflix emphasized that although big data are generally important, it is people who make the final decision with the analyzed data.3

Realizing that there are important factors other than the currently collected big data, Netflix decided to produce 13 episodes of one season at once as soon as the drama production was officially decided.

Amazon, on the other hand, took the traditional approach of collecting, analyzing, and reflecting big data on market reactions by releasing eight pilot programs free of charge. They categorized general information, such as customer's age, gender, place of residence, and viewing environment, and behavioral patterns, such as customer's content purchase list, view time, and taste, based on which they produced dramas. However, Amazon's Alpha House, which was produced based on big data analysis, did not fall under the category of analytical data due to the ambiguous relationship and interaction between the content and the viewer. This is a failure case because quality data from a complex and contextual perspective, which is generally called expert perception, were ignored and they merely relied on typical big data analysis.4

On the other hand, while Netflix thoroughly used big data to find the relationship between the drama and the viewer, they also considered the subjective experience of the relationships and interactions of the experts, which is another factor that is not included in the big data category.

2.2.4 Sucess Case of Big Data Analysis of Google Trends

On the other hand, Google's "Flu Trends" is one of the cases where the use of big data is appropriate. Flu Trends, like Google Trends, collect data on locations where users search for information about "cold symptoms" and viruses. Based on the collected data, the location where the cold is spreading is displayed on a map, and each state in the US is marked with "low", "weak", "medium", and "strong". In addition, a graph that compares the intensity of the common cold virus this year with that in the last year allows predicting in advance how the cold will be spread.

Based on the fact that this approach found the location with the cold infection two weeks earlier than the Centers for Disease Control and Prevention (CDC) did, because the CDC relies on reports from hospitals and state health departments, and that more patients are relying on home remedy rather than going to a hospital,

collecting web search data is more efficient than collecting hospital data.

The validity of such a cold prediction method suggests that determination with data in Google's collection category is possible. In other words, the big data analysis can be accurate enough, because the relationship and interaction between the cold-related data and the user who searches the data are weak. The relationship between Google and the user is also weak, because it is free to create and delete the Google account, and vertical/horizontal relationships do not exist. Furthermore, since the interaction that occurs within them is less likely to control, perform a role, or interact with each other, such data can produce effective results even if the analysis is conducted based on big data.

2.2.5 CRM Marketing Case of Target

The case where Target found out about the pregnancy of a high school student in customer relationship management (CRM) marketing is a typical case of big data application. The pregnancy was predicted based on the statistics of the early-term nutrients, mid-term lotions, and late-term infant supplies that pregnant women purchase, and personalized marketing was used. Even in this case, the pregnancy of a high school girl can be determined by analyzing the purchase data to extract informative information without any other data. Since the relationship and the interaction between a high school girl and a pregnant woman are not deep and strong, for the sampling of such a population, it is possible to analyze the data collected by the company's general data collection method and make a proper

prediction.

2.3. Case Analysis Implications

As a result of analyzing the typical cases of big data application, it was found that some application cases were successful and some were not. The reason for the failure was that the data, which were generated when the interaction between people or a person and an object is strong at the time of data generation in the category, were not included in the population. On the other hand, in the successful cases, significant results could be produced even if they were analyzed by the sample group created by the current data collection method because the relationship and interaction between people or a person and an object in the data population were not strong.

Errors occur in the population estimation method of big data, and in such cases, it is meaningful to collect a small amount of data in a limited way by using an offline-based observation method, which is excluded from the existing digital-based data collection method. However, most of the current big data methodologies estimate the population that excludes such considerations.

Such errors in the population estimation can be interpreted in two perspectives. First, the basic processing of big data finds a pattern of data of events that occur frequently or repeatedly. However, the amount of informative information can be small after processing. On the other hand, the data that have a strong relationship and interaction with unlikely events may have more

informative information than the events that occur frequently. In applied mathematics, this is solved by the concept of 'information theory'.5 Even if the data size becomes larger, the amount of informative information in it can be small.
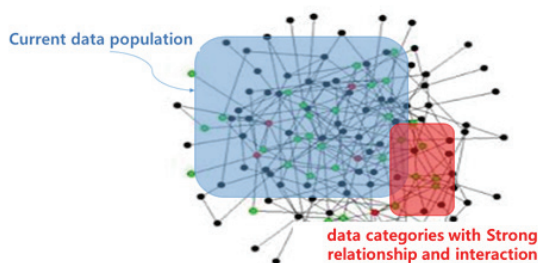

Figure 2-1. Estimation Error Comparison of Big Data Population

Second, there may be an apophenia problem in big data interpretation. A photograph of a "face" on Mars, which was taken by NASA's Mars probe Viking 1 on July 25, 1976, was claimed to be evidence of aliens living on Mars according to some media and public opinion. NASA announced that it was a piece of tomography that was formed by apron-shaped early debris and landslides. The phenomenon that causes such a misinterpretation is called apophenia.

Applying the law of large numbers, which states that the average of random samples from a large population is likely to be close to the average of the entire population, analysis based on apophenia can generate noisy data in big data.

In conclusion, there was a category of data that did not fit into the existing big data population, and those data are generated in a strong relationship and interaction between people or a person and an object at the time of development.

3. Relationship and Interactivity

3.1 Definition of Relationship

Relationship refers to the nature of two or more people, objects, and phenomena that are connected to each other, as opposed to separate, independent, or discontinuous.

Relationships do not just refer to the relationship between me and another person, but the nature of being connected to and related to both living and inanimate, including me to myself, to others, to nature, to the earth, and to the universe.6

In this paper, we limit the scope of relationships to the relationship between people or a person and an object. The purpose of this study is to examine whether or not the characteristics of the data that need to be interpreted can be interpreted only by the digitized big data analysis method according to the strength of the relationship.

The following explains why the strength of the relationship is an important factor in data interpretation. Newborn humans are helpless compare to other baby animals. Therefore, the human is a weak and dependent being who has to meet the basic needs for maintaining life with someone's help from the moment of birth. The only way to survive in the law of the jungle with these unfavorable physical conditions is to have a cooperative lifestyle. The biological condition of a human, such as physical weakness, has made the human a social being who has to live in a relationship with others. Through this process, human beings consciously and unconsciously

learned that the existence of others is essential and that interdependent relationship with others are important values.7

If the relationship hidden in this unconscious nature is strong, there is a limit to the interpretation of meaning in the exposed data only. Human's actions do not directly express the inherent desires. Since the desires change according to the strength of the relationship, the meanings cannot be interpreted properly in the currently collected data currently.

Tab. 3-1 Relationship Types

Types	Details
Primary vs secondary	Choose whether the relationship is granted regardless of your choice, or is a temporary relationship and thus free to join and leave8
Vertical vs horizontal	Choose between an unequal vertical relationship with no equivalence or horizontal relationships with high potential for spontaneous interaction
Existential vs Role	Choose whether the relationship is an existential relationship with sincerity or a role relationship through meetings designed to satisfy needs10
Cooperative vs Competitive	Choose whether they are cooperative or competitive, based on how they feel about their partners

In this study, based on literature research, the relationship refers to the degree of psychological, emotional, and physical connection or relationship between people or a person and an object.

3.2 Definition of Interaction

Interaction is an action that humans take with others in a given environment, or that humans exchange with other humans, objects, or entities, and the object that provides this behavior is called interactive (Lombard, Snyder & Dutch, 2001).11 Interactivity, on the other hand, mainly extends beyond humans to the relationship between humans and objects. If an interaction is used to express human behavior that occurs primarily in interpersonal communication, interactivity is a key element of future technologies and services that can extend beyond human interactions to a range of non-human relationships such as media and devices (Pavlik, 1996).12

Tab. 3-2 Interactivity 3D Structure

Type	Structure
Control (lead)	The strength of control in finding alternatives to or transforming the content, timing, or sequence of interactions
Role exchange	The strength of exchanging roles between message senders and recipients
Interactive conversation	The strength that occurs sequentially depending on B's response to A, when A and B communicate and A's response to B occurs first.

Interactivity is the process of dynamic change in social behavior between people or groups of people who change their behavior or reactions in the group. In this paper, interaction is defined as an action or communication that affects each other and extended to non-human relations.

In this study, the meaning of population estimation of externally generated unstructured data that were alienated from previous big data research was studied in terms of relationship and interactivity. When the data are analyzed in terms of the relationship and interactivity, they need to

be examined thickly rather than deeply from the contextual perspective. Therefore, such data are defined as thick data.

4. Concept and Definition of Thick Data

4.1 Why Thick data?

Let's take an example of the data analysis. There is an old man who lives for 300,000 won a month. In this situation, we would like to predict if the old man can purchase a smart phone that is one million won.

The data that can be collected from the perspective of general big data include government assistance such as basic pensions for the elderly, housing information, cash receipts for purchase transactions, details of credit card use, and public transportation. Based on these data categories, the analysis will not show any data that would conclude that the old man can purchase a 1 million won smart phone. Therefore, the possibility of purchasing the smart phone will be very small.

However, if we expand the population to the data with a strong relationship and interactivity, we may be able to conclude that he can purchase the smart phone.

First of all, he has the only son who is married and has a relationship with his eight-year-old grandson. The son and the couple are ordinary middle-class people, which means that they can't afford to buy a smart phone for the eight-year-old son. The grandson is in elementary school and he has a best friend, whose parents are doctors. The parents of his friend bought a million won smart phone for the friend and constantly shows off his smart phone to the grandson.

The old man found that the grandson wants to buy the million won and his son and the wife are not buying it for him. The old man can save 200,000 won for 5 months and eventually buy the smart phone for his grandson. The relationship between the old man and his grandson is strong. It is almost impossible to find out the desire to buy the smart phone due to the love of the old man toward the grandson, which cannot be caught by the general big data population.

The below image illustrates the relationship and interactivity in terms of data collection categories. The scores on the charts are generated by creating a checklist based on the relationship type and the three-dimensional structure of the interaction.
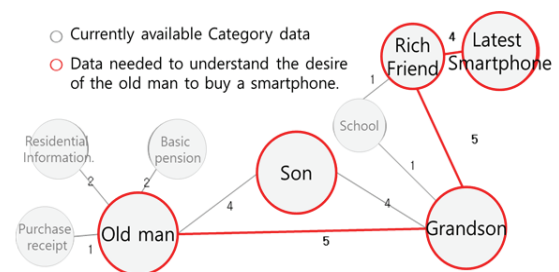

Fig. 4-1. Comparison of data collection categories through a case

Likewise, thick data can be defined as external data with the unstructured and contextual nature that can be generated or collected in a state where there are a strong relationship and interaction between people or a person and an

object that are difficult to be sampled by the existing collection method because they are excluded from the existing big data population.

4.2 Why Thick, not deep?

Why is it called thick data, not deep data, if the data are examined in-depth? There are unique traits that each person has but rarely expresses, but these traits do not appear frequently and thus are difficult to be observed. Furthermore, such traits only appear when the person has a relationship and interaction with another person. As a result, it rarely occurs. The characteristic of the cognitive science approach is to look deeply at one person, but when informative data is extracted, there is a significant amount of information in the thick relationship between people.
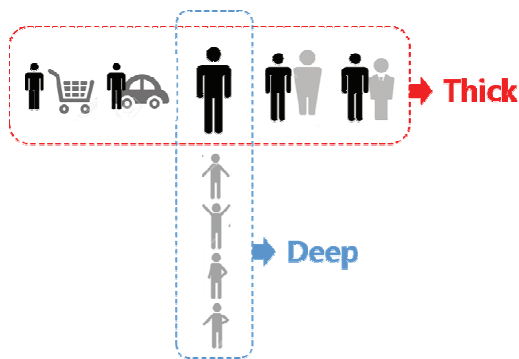

Fig. 4-2. Perspective Differences of Thick and Deep

Deep is a singular comparison inside a person, but thick is a plural comparison from a contextual perspective. The below chart was revised by the authors based on Tracia Wang's table published in Medium.

Tab. 4-1. Big data vs Thick data

Item	Big data	Thick data
Perspective	Historical data (Interpretation of generated data)	Current data (interpreting the generating data)
Method	Patterns found within processed (quantified) data categories of collected unspecified data	Meaning found in categories of data that are generated in social contexts (relationships and interactions)
Feature	Variables separated to identify patterns	Irreducible complexity accepted
Collection	Acquisition based on various sensors that convert physical signals into electrical signals	Acquisition by observation method based on User eXperience or ethnography
Strength	Common features found in data	Hidden implications found in data
Weakness	Limitations on analyzing Data Contextually	Limited (small) data scale

Thick data are the data generated in the interaction and interactivity between people or a person and an object in a social context. The optimized methods of collecting and processing data for the thick data interpretation include ethnography, which is a method used by anthropologists who live with locals in other cultures to observe and record them in detail, and User eXeprience, which focuses on the overall emotions and memories of perceptions, reactions, and behaviors that users experience and feel while directly or indirectly using a system, product or service. These two methods collect and interpret data by observing various contexts of social, cultural and economic viewpoints based on the observed people.

The below chart illustrates how the collection of thick data in an observation analysis is an example of judging data incrementally by classifying them according to the level of relationship and interactivity strength.

Tab. 4-2. Five Steps of Observation Analysis for Thick Data Collection

Step	Concept	Example
Step 1	Simple Cognition (RI: level 1)	A man uses a smartphone.
Step 2	A detailed description of the situation (RI: level 2)	A man in his mid-40s is holding a smartphone with his right hand.
Step 3	A detailed Description in the interactivity perspective (RI: Level 3)	A man takes out a smartphone from the back pocket of his pants and calls with his right hand.
Step 4	A detailed Description in Contextual Perspective (RI: Level 4)	A man in his mid-40s, who looked worried on the side of the road, suddenly takes out his smartphone and calls with his right hand.
Step 5	Observation in step 4 + rational reasoning (RI: level 5)	A right-handed professional man in his mid-40s, who was looked worried on the side of the road, calls using his Samsung Galaxy Phone. His frowning face indicates that there is an issue in his family.

 * R:Relationship, I:Interactivity
** RI strength:1 Very weak, 2Weak, 3neutral, 4strong, very strong

5. Conclusion

Big data will certainly be an important core technology that will change the future of mankind. This study does not deny this prospect. However, this study is intended to present a problem when the data collection of all phenomena is performed using the current method only.

In this study, some cases where errors can occur when the analysis is conducted using big data are presented. Furthermore, based on the analysis, it was suggested that the errors in the population estimation of data can occur when the relationship and interaction are strong during data generation. This study also suggested that collecting and interpreting data from the perspective of thick data is effective for those data that are usually excluded from this population. The study also suggested that user experience and Ethnography are the most optimal collection methods for including thick data with strong relationship and interactivity in the population of big data.

In the end, it is important to recognize these two methods as important methodologies for unstructured and qualitative research from the perspective of data science.

In big data, not only a large amount of diverse information is used, but also that information is collected and processed quickly. Furthermore, it recognizes the uncertainty of data and finds business value through analysis. In other words, the process of finding business value from diverse and crowded data is called big data analysis. In other words, although data are analyzed through difficult analysis methods, if the analysis results do not have business value, then they are not considered as big data. Therefore, measuring the strength of relationship and interactivity between people in data during the big data analysis will be an effective method to reinforce the business
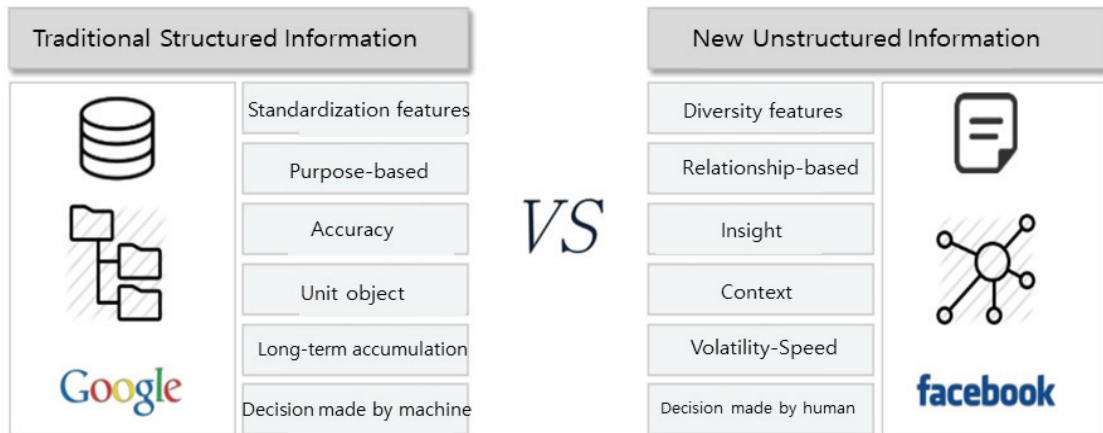
Fig. 5-1. Comparison of traditional structured information and new unstructured

value of big data.

If the thick data theory-based research in the data science perspective becomes active, the following effects are expected.

First, there will be an opportunity to systematically approach the establishment of qualitative data-based strategies that are highly unstructured and difficult to be collected in the digital infrastructure.

Second, economic value can be determined by standardization and quantification of procedures and methodologies for collecting, processing, and utilizing data generated from relationships and interactions between people or a person and an object. Criteria, such as size and scope, can be created for the area.

The importance is all recognized, but they can create criteria such as size and scope for areas where the importance is undeniable but the criteria for the measurement and evaluation are ambiguous.

References

[1] Ahn, Jinho(2018.9.26.), 'The correlation between fat finger and user experience (UX)', Electronic Press. Retrieved from https://www.etnews.com/20180926000120?m=1 (안진호, 2018.9.26. 팻핑거와 사용자 경험과의 상관성, 전자신문)

[2] Cheil Magazine&Blog, Understand consumers with data.(2017.10.13.) Retrieved from https://blog.cheil.com/magazine/26429 (제일매거진&블로그, 데이터로 고객을 이해하라. 2017. 10.13)

[3] Choi, Jimin(2014), A Study on the Moral Education Meaning of Relationship, Korea National University of Education Graduate School.

[4] Fromm.E.(1963). Theartofloving, NewYork: Bantam Book.

[5] Kang, Sun-Bo, Park Eui-Soo, Kim Gui-Sung, Song Sun-Jae, Chung Yun-Kyoung, Kim Young-Rae, Ko Mi-Suk(2008), A Foundational Study for the Vision of Education for the Human

Nature for 21st Century, Korea University Institute of Educational Research, Vol. 30, p.16.

[6] Park, Sanghyun(2016.2.2.) Netflix knew and Amazon did not know… Don't just trust big data ChungAngIlbo Press, Retrived from https://news.joins.com/article/19513378 (박상현, 2016.2.2. 넷플릭스는 알고 아마존은 몰랐다.. 빅데이터만 믿지 마, 중앙일보)

[7] Pavlik, J. V. (1997). New media technology: Cultural and commercial perspectives. Allyn & Bacon.

[8] TED Talks,(2017). Tricia Wang: The human insights missing from big data[Video]. Retrieved from https://www.ted.com/talks/tricia_wang_the_human_insights_missing_from_big_data/reading-list

[9] Tricia wang.(2013, May 13). "Why Big Data Needs Thick Data, Ethnography matters, Retrieved fromhttp://ethnographymatters.net/blog/2013/05/13/big-data-needs-thick-data/

[10] Lee, Jaechang(1998), 『Human Relationship』, Mooneumsa: Seoul, pp.26-27.(이재창, 인간관계론, 문음사, 서울: 26-27).

[11] Lee, Jeungsun, Ahn Jinho(2019), A Study on the Effect of Relationship and Interaction at the Time of Data Generation on Big Data Analysis, IT service Science seminar.(이정선, 안진호(2019), 빅데이터 분석에서 관계성과 상호작용성의 효과성에 대한 연구, IT 서비스학회 추계학술대회).

[12] Lee, Seongtae(2006), 『Human relationship』, Yangseowon: Seoul, pp.14-15.(이성태(2006), 인간관계론, 양서원, 서울: 14-15).

[13] Lombard, M. S. D., & Snyder-Dutch, J. J.(2001): "Interactive Advertising and Presence: A Framework". Journal of Interactive Advertising, 1(2), 581-592.

Ahn, Jinho (pibuchi@gmail.com)

Ahn, Jinho is the president of idinno lab, co. ltd(www.idinnolab.co.kr), a business consulting firm in Korea. he obtained a Bachelor's in Industrial Design, a Master's degree and his Ph.D. degree in Business Adminstration at the University of Kookmin. His research interest focused on Service Design and Service System.

Lee, JeungSun (jslee@eulji.ac.kr)

Lee, Jeungsun is an assistant professor of Mortuary Science Department at Eulji University in Seongnam. She received her Ph.D. in Service Management at Kookmin University in Seoul, Korea. She holds a Master's and Bachelor's degree in Business Administration from Texas A&M University at San Antonio, Texas. Her research interests include areas in Service Philosophy, Service Innovation, as well as Funeral Service and Mortuary Science.

서비스 비즈니스의 빅데이터 모집단 산정방식 오류에 관한 사례연구

안 진 호*, 이 정 선 **

요 약

최근 빅데이터가 사회, 경제적으로 중요해지면서, 무분별한 적용으로 인해 많은 문제가 발생하고 있다. 빅데이터의 활용가치는 데이터 안에 숨어있는 가치 있는 정보의 의미를 알아내는 것이 중요하다. 특히 고객의 행동패턴이나 경험 등을 예측하기 위하여, 기업내의 CRM (Customer Relationship Management)에서 추출한 정형데이터나 SNS(Social Network Service) 등에서 추출한 비정형데이터를 모집단으로 규정하고, 해석하는 과정 등에서 수많은 오류가 발생할 수 있지만, 대부분 이를 간과하고 있는 것이 현실이다. 이는, 데이터 분석기술 외에 고려해야 할 사항이 유의미한 패턴으로 나타나야 할 데이터들이 모집단 안에 빠져있음을 나타내는 것이기도 한다.

이에 본 연구에서는 그 모집단 설정의 오류 발생의 원인이 사람과 사람, 사람과 사물간의 관계와 상호작용이 강한 경우에 발생하는 데이터의 측정과 해석은 그 강도가 강한 경우에는 User eXerience(UX)와 ethnography(민속지학) 관점으로 수집된 데이터를 모집단에 포함하는 것이 중요하다는 것을 Big data 적용의 다양한 사례를 비교하여 도출하였다. 그리고 도출된 의미를 파악하여 최적의 방향을 제안하였다.

표제어: 빅데이터, 서비스 비즈니스, thick data, 사용자경험, 민속지학

* ㈜아이디이노랩 대표, pibuchi@gmail.com
** 교신저자, 을지대학교 바이오융합대학 장례지도학과 교수, jslee@eulji.ac.kr