Analysis of User’s Behaviors and Growth Factors of Shopping Mall using Bigdata

Dong Woo Kim¹, Tae Gu Kang², Guozhong Li³ and Seong Taek Park²*

¹The Korea Association of Software Manpower, Korea; dongwoo1108@naver.com
²Department of Management Information Systems, Chungbuk National University, Heungdeok-gu, Cheongju-si, Chungcheongbuk-do, Korea; snookerk@hanmail.net, solpherd@cbnu.ac.kr
³Department of Management Information Science and Information System, Kunming University of Science of Technology, Yunnan, Kunming, Wuhua, China; misgukchung@nate.com

Abstract
As recognition and value on big data has become significant, companies, public institutions, intelligence agencies have started to have interest on big data. Of course, there were analytic techniques based on data in past as well, but data was limited and analysis also limited. However nowadays, interest on structured data due to social media and unstructured data has started to increase that interest on big data analysis is rising. This paper systematically analyzes behavior of internet shopping mall users through big data analysis and proposes a strategic operation plan using this analysis.

Keywords: Association Rules, Big Data, Growth Factors, Shopping Mall, SOHO

1. Introduction
With the development of the Internet, foundation of small ventures tends to increase day by day. Following this, the core of shopping moved from offline to online and is now moving to mobile nowadays due to the rapid development of smart devices (mobile devices). Also, the concept of social commerce has emerged due to the appearance of social network services. For instance, existing commerce that took place in offline stores have moved online for Amazon and Alibaba to exist.

Comparing the outcome of representative e-commerce companies Amazon and Alibaba, GMV (Gross Merchandise Volume) was $115 Bil standard to December 2013 accounting standards and Alibaba was $270 Bil to be 2.3 times larger than Amazon. For sales, Amazon recorded $74.5 Bil and Alibaba recorded $8.4 Bil as Amazon was 8.9 times higher. For net profit, Amazon was $0.27 Bil and Alibaba was $3.8 Bil as Alibaba was about 14 times higher¹.

<table>
<thead>
<tr>
<th></th>
<th>GMV</th>
<th>Revenue</th>
<th>Net Profit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amazon</td>
<td>$115 B</td>
<td>$74.5 B</td>
<td>$0.27 B</td>
</tr>
<tr>
<td>Alibaba</td>
<td>$270 B</td>
<td>$8.4 B</td>
<td>$3.8 B</td>
</tr>
</tbody>
</table>

(Source: Securities And Exchange Commission, Amazon Annual Report)

However as the e-commerce platform has been activated compared to the past, much change has occurred in shopping malls. Large shopping malls such as Amazon and Alibaba were main markets in the past, but small shopping malls are appearing nowadays. The appearance of small shopping malls has taken place not only because of e-commerce platform, but also has influence of consumption patterns of consumers.

Nowadays, personality of consumers is being regarded significant and products that distinguish others are being found. Small shopping malls that can put emphasis on personality are appearing nowadays due to this tendency of consumption.

*Author for correspondence*
Small Office Home Office (SOHO) refers to small business in which an individual does business in a small office or at home. Recently, there has been a prediction that with the development of mobile technologies, shopping malls using mobile devices increase and specialized SOHO online shopping malls are rapidly going mobile. In addition to the trend of spreading mobile shopping, as solutions for establishing shopping malls strengthen mobile support, the weight of mobile transactions on SOHO shopping malls in the first half of 2014 increased to 30%, three times of that in the same period of the previous year and the turnover is estimated at 500 billion won.

Like this, as online purchases are vitalized, academia as well as related industries have a lot of interest in research. Recently, as the uses of various statistical tools such as data mining and big data have begun, studies to understand the characteristics and tendencies of consumers and analyze their behavior using them are conducted.

2. Theoretical Background

Whether customers use an online shopping mall or an offline shopping mall differ depending on the value they consider. Unlike offline shopping malls, online malls do not operate face-to-face, but they have a more advantageous environment of data collection, so the analysis and utilization of these secured data can be used for service and marketing.

Lee et al. drew a fuzzy cognitive map through log data collected from certain shopping malls and analyzed patterns of users’ purchases using them.

Choi et al. proposed a personalized referral system using RFM model and k-means method through customer information, item information and purchase history information. Kim and Kim proposed a genetic algorithm-based product recommendation system and verified the proposed method through building up a virtual shopping mall.

Park et al. analyzed Web traffic data of 166 domestic shopping malls in order to understand strategic factors affecting the growth of Internet shopping malls.

Byeon and Kim constructed the density and intensity of the network with click stream data of Website visitors, though which they analyzed customers’ purchasing behaviors.

Kim investigated internet fashion shopping mall visits subject to female university students by exposing shopping mall names and product information to main visiting routes. By this, an effective and suiting promotion method was analyzed and proposed to increase visiting customers in small fashion shopping malls operated by little capital.

Ji analyzed the influence of interaction of product property and consumers on purchase satisfaction and repurchase intention subject to small shopping malls.

To propose the most appropriate search advertisement strategy for small internet shopping malls, Choi analyzed types of search advertisements focusing on statistic data gained from data given by each advertisement company and shopping mall self-built analysis system and proposed a successful search strategy plan.

The preceding research referred until now were mainly subject on large shopping malls on plans for shopping malls and mostly used data mining methods for analysis. Also, most preceding researches subject to small shopping malls surveyed consumers of certain group in which there is limitation in generalization.

Until now, there are no references of analyzing or using big data subject to small shopping malls. For this, big data is used in this study and a new approach method on shopping malls is searchingly proposed.

3. Research Model and Design

Understanding and reflecting customers’ responses in real time is very helpful for shopping malls to enhance their revenue. Currently, large shopping malls and social commerce networks are developing and providing big data-based customized services to increase their market share. They analyze past purchase history, visit categories, product views count, actual purchasing state and final payment price recommending related products through providing customers with customized services.

The data used in this study were secured at www.gonsen.co.kr over the past three years. Using the collected data, this study attempts to analyze customer identification, demographic variables and customer segmentation through lifestyle.

This study tries to analyze the secured data using a variety of techniques of analysis such as RFM and Mining. “G” to be analyzed in this paper was established in 2008, which is a company that is operating an online shopping mall for women’s clothing, shoes and accessories.

In this study, the correlation between shopping mall’s data and shopping mall’s goods was analyzed. After
definition, analysis methods to use information data for searching alternatives to solve this problem were planned. Open source tool R was then used to realize analysis and visualization and the procedure to interpret this was performed.

3.1 Definition of Problem
SOHO mall As competition increases through accurate Big data analysis rather than traditional rule of thumb formula midst of a lively and made a focus on any marketing activity that unfolds.

3.2 Information Required for Problem
Receiving data from the last few years, the gender ratio, sales, product information and data necessary for such purchase price.

3.3 Data Required to Derive Information and Analysis Method
Data is first collected and the data was organized through preprocessing procedure for analysis. Only data required for analysis was gathered to make one new data group subject to analysis. Lastly, analysis tool “R” was used to calculate desired information. Statistical program R is a free software based on open source which allows process of big data of various forms, statistics, and realization of visualization in which analysis was conducted using R in this study.

4. Research of Study
We gathered data generated from “G” SOHO shopping mall. Also, we applied to associations rules, classification analysis, and RFM (Recency, Frequency, Monetary value) analysis in order to analyze collected data. In fact, there are many analysis methods. We thought that these methods are most suitable for “G” SOHO shopping mall’s data and purpose of our study. As a result, association rules and classification analysis only led to significant results. Thus, in this study, we present results using association rules and classification analysis.

4.1 Association Rules
Association rules are proposed Agrawal et al\textsuperscript{11}. This method is very important data mining model studied extensively by related academic community. This is first used in market basket analysis to find how items purchased by consumers are interrelated\textsuperscript{10}. This assumes that all data are categorical, not numerical.

For example, supermarket transaction data collected by POS (Point Of Sales) system is very useful. Generally, supermarket has a set of items (I):

\[ I = \{ i_1, i_2, \ldots, i_m \} \]

Transaction \( t \) are composed of a set of items (I):

\[ t \subseteq I \]

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{supermarket_receipt.png}
\caption{A supermarket receipt.}
\end{figure}

In this example, we can express \( t \) as follows:

\[ t = \{ \text{organic courgettes, tri trompetti, organic carrots, walnut bread, organic cheese, walnuts, olive oil, organic soup, organic mushrooms, truffle torte, 100 fairy light} \} \]

Finally, database \( T \) is a set of transactions as below:

\[ T = \{ t_1, t_2, \ldots, t_n \} \]

Association rules find easily patterns how each items are interrelated. The following figure shows process of association rules.
Analysis of User’s Behaviors and Growth Factors of Shopping Mall using Bigdata

Return to “G” SOHO shopping mall’s case, we found the following conclusions by using association rules analysis (See Figure 3). People who purchase many bottoms tend to purchase more outer clothes and people who purchase tops are more likely to purchase accessories and bottoms.

4.2 Classification Analysis
Classification analysis is also data mining analysis method such as association rules. It predicts categorical class labels and classifies data based on the values in a classifying attribute and uses it in new set. The process of classification analysis works as the following figure.

G has provided marketing for woman customer. But, the results of analysis show that the great majority of their customers is men. Also, product preference was significantly different from gender. If we had more auxiliary variables such as age, income level and region, we would be able to investigate more closely. But it is important to remember that the idea of clustering is just to see how groups are formed.

The next figure shows classification results for products men buy. The most purchased products by man are neat (7) and cardigan (7), followed by coat (5), jeans (5) and bikini (5). But, from a common-sense point of view, these results are incomprehensible. The best possible explanation is that the man would have purchased one-piece and bikini and gave it to the woman.

Tops the man purchased most are neat (7) and coat (5), followed by cardigan, (3), jacket, (2), shirt (2) and T-shirt (1). Mostly, the best selling man’s tops in the shopping mall are shirt and T-shirt. But, it doesn’t suit this result at all. Neat and coat are more expensive than shirt and T-shirt. It means that marketing for men will have some effectiveness in increasing sales.

Next, male customers who visit this shopping mall purchase one-piece the most (7). It seems that they are not interested in bottoms this shopping mall sells. But, as above, they buy tops for gift. At least, it’s fortunate that jeans (5) this shopping mall sells are quite popular products. Thus, this shopping mall should bring many kinds of jeans and new product line with an eye-catching design.
Figure 5. Purchase of man.

Figure 6. Purchase of Top (Man).
5. Conclusion

With the rapid development of ICT, with the merit that they can make purchases anywhere with the Internet connected disregarding the time and place, many consumers use online shopping malls. However, individual consumers want different things and the reason for visiting shopping malls, also, vary, so a method of analysis different from that in the past came to be necessary.

In other words, analyzing customers’ behaviors using various methodologies of analysis with the secured data and providing them what they want, timely, through that would allow efficient operation of shopping malls.

The results of this study are as follow:

- Implication of this study is that analysis was conducted using practical data of small shopping malls. Due to this, a strategic plan suiting small shopping malls could be derived.
- Operators of other small shopping malls who did not consider data important used the result of this study that working groups can use this as guidance when establishing strategic plans for their shopping mall.
- However, this study has the following limitations despite these implications.
  - There is limit in systematic analysis by only using data of certain years.
  - These are study results subject to certain shopping malls that there is difficulty in generalization.
  - Recognition on data storage and storing methods was lacked by G company that the desired high quality data could not be given for this study. VOC text mining was planned to be conducted through web crawling, but VOC could only be seen by customers and the manager that limitation exists for study of VOC.
  - Considering that most studies focus on large shopping malls, this study seems to have sufficient significance. In the future, it will be necessary to carry out studies to secure and analyze non-stereotyped data as well as stereotyped data and analyze them from various perspectives.

6. References


