A Study of Evaluating of Game Review Data Using Multiple Regression

Lee-Sac Lee, Sang-Hyun Lee, and Hyun-Seok Hwang

***Abstract*—As the use of the Internet increase, online reviews also become popular. Usefulness of the reviews can help people to decide product purchase. Many previous researches show that if the quality of reviews was high, the sales volume of that product would increase. Therefore, we investigated that what factors in the reviews can affect the reviewer's evaluation results.**

We perform a case study of popular online game reviews, STEAM community review, to identify the influencing variables in review data set using multiple regression method. We also conduct sentiment analysis of review comments to mashup sentiment results to original data set. We will provide analysis results and interpretation of the results with further research directions.

***Index Terms*—Game data analysis, multiple linear regression model, online review.**

I. Introduction

As the network developed through the popular use of the internet and computers, the world-wide accessibility to networks enables people to communicate without any limitation. Also, their desire to express themselves on the internet increased actively [1]. The online game can be one of the most typical utilization of the internet network using the computer. The online game began to gain great popularity in 2001 and the people started to enjoy the console or CD games by downloading and playing on the web [2]. The communication became more and more active in this game network and the online game communities have been built in cyber space [3].

The representative communication within this online game community can be the online game review. Recently it is possible to contact the online reviews very easily. Consumers help other consumers make a better choice on the basis of their experience or provide their experience as the review for the improvement of the quality or service. People can gain the response of real users' experiences from online game reviews. Furthermore the highly useful reviews give much information to potential buyers and make real consumers by activating purchase intention. On the contrary some negative reviews can diminish peoples' intention to purchase an item and disseminate negative reputation of the item [4].

Therefore, in this study, the reviews of 11 games with a variety of genres including Role-Playing-Game, First-Person-Shooting, and Sports provided by "STEAM", a popular game platform, is analyzed in order to find the fundamental factors affecting the usability of the game review.

In Section II, we review the previous works related to online review and analysis of the review. Section III covers a research framework and a research procedure in this paper. In Section IV, we conduct a case study and provide the results of this research. We also discuss the results and concluding remarks with limitations and further research questions in Section V.

Manuscript received August 29, 2016; revised November 15, 2016. This research was supported by the MSIP(Ministry of Science, ICT and Future Planning), Korea, under the CPRC(Communication Policy Research Center) support program (IITP-2016-R0880161008000100 2) supervised by the IITP(Institute for Information & communications Technology Promotion)"

Lee-Sac Lee and Sang-Hyun Lee are with the Graduate School of Interaction Design, Hallym University, Chun-Cheon, Republic of Korea (e-mail: isaac6143@naver.com, s038987@naver.com).

Hyun-Seok Hwang is with the Business Administration Department, Hallym University, Chun-Cheon, Republic of Korea (e-mail: hshwang@hallym.ac.kr).

II. Related Works

A. Online Review

Consumers post their opinions online about their experience of buying goods in person. Other people's review has a feature that consumers make and distribute by themselves. Therefore, it is more reliable than different industrial information provided by the companies or online vendors [5].

If the reviews provided online are credible and enough information is provided for the users' decision making, the review can have a positive effect on perceiving the usefulness on that online market itself, so it can facilitate the constant visit. The online review not only lets consumers perceive the usefulness of the online market but also can have a positive effect on the purchase intention of the product itself by letting consumers accept the online reviews. It indicates that people who read the online reviews are likely to buy that target on the basis of those reviews. The evaluation of the reviews can also lead to the purchase. It was researched that if the number of reviews became bigger and the average review assessment heightened more and more, the sale of goods heightened as well [6].

Thus it is confirmed the online reviews are playing an important role in users' decision making and the acceptance of the online reviews affect the behavior intention and influence even the possibility of the substantial behavior.

B. Online Review Analysis

There is a study that obtains the opinions or information of the people who left the reviews to analyze. Also, there is a study that conducts the statistical verification between the sale

doi: 10.18178/ijimt.2017.8.2.710

data on the product or another data such as the users' nature and the online review. The study on the online review is largely divided into two things. The data in which the users' responses more frank and direct than other study methods are reflected can be obtained by extracting and analyzing the information from the opinions of the people who left the review. In Byun Dae-Ho(2015)'s study, a study of apprehending if the pertinent theme was positive or negative by text-mining the reviews on the use experience of the smart TV was conducted. Since the users' frank emotion is revealed in the review, it was intended to grasp the users' emotion on the product under the theme of review [7].

The method of verifying the sale data on the product, user nature, and online review statistically has strength in terms of the statistical usability. Since many people voluntarily participate in the online review, it is possible to gain plenty of samples although not collecting the samples. Also, there are abundant data to analyze in association with the level of satisfaction through the buyer feature or score [8].

Therefore, in this study, the relationship between features related to this online review analysis and the review usability will be examined through the sentiment analysis on the review itself. Also, the effect of the review data and user's nature on the review usability will be apprehended by grasping the relationship between the review data and user nature.

C. Online Game

It is early 2000s that the online game heat was prevalent and it is also early 2000s that the study on the online game was conducted [2]. At the beginning of the study, the explorative research was mainly carried out and the interactional natures only the online game had began to be mentioned. The behavior study on the online game users was prevalent in mid-2000s and subsequent studies examined the users' cognitive and emotional response or suggested the flow experience of the game use [9].

Since the mobile market developed, the game market has also leaned much to the mobile sector not the online game so the study on the game was also expanded to the mobile market. As the mobile market grew at an alarming rate, the online game market declined sharply but it still accounts for the biggest proportion in the game market. Recently the sales of the online game market are increasing little by little [2] .

As the former studies mentioned before, the studies related to the game use behavior such as the user's game use motive, response, and flow have been carried out consistently since the initial market of the online game. However, there are few studies on the review within the game community and the factor of game users' review acceptance. Therefore, in this study, the variables influencing the usability of the review would be apprehended through the community review analysis within the online game.

D. Game Data Analysis

Basically the online game data is classified into different things according to the kind of game; it can be largely divided into two things: the user character information and the game system information[10]. The user character information is comprised of the basic information about the game user profile and the information that the game user character has. On the contrary, the game system information means to manage the information like the elements in the game such as the region (map), NPC, quest, and item [11].

Two kinds of game data and other data above can be divided into the dynamic data that connects to the database in real time during the game to influence and the static data that has a small effect on playing the game. Since many users connect to the online game intensively, it is necessary to process the mass game data promptly in order to carry out the game effectively in response to it. Therefore, there is a game that divides several game data according to the type and manages in a different way [11].

In this study, the data on the basic information about the user profile is inadequate. However, an attempt to grasp the usability of the review by using the information the game users had online and the information the game user character had was made.

III. Research Framework

To identify the factors influencing review evaluation, we suggest a research framework shown in Fig. 1. We classify the potential factors affecting review evaluation into three categories - Review characteristics, Reviewer characteristics, and Sentiment polarity.

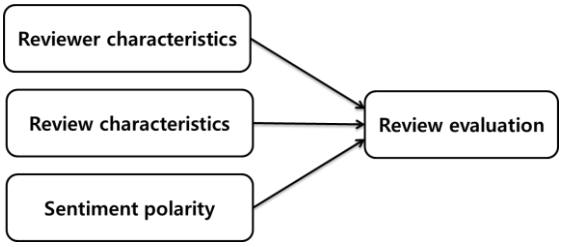

Fig. 1. Research framework.

We assume that game community members will determine whether a review article is helpful or not based on the above three facets. If a reviewer has provided enough qualified reviews, then this reviewer characteristic will affect other gamers and review will be evaluated as "Helpful". Review characteristic like the number of review exposure to others is expected to affect review evaluation. The last factor, sentiment polarity, is assumed to influence review evaluation. The sentimental polarity of review text is a critical to determine the evaluation results.

Fig. 2 shows a suggested research procedure in this research.

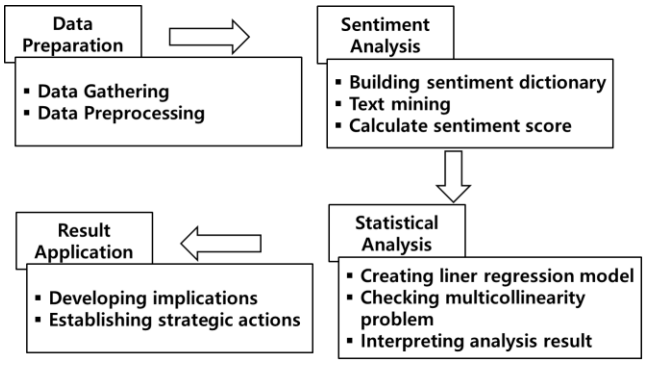

Fig. 2. Research procedure.

In step 1, we prepare research data. After gathering a data set, we perform data preprocessing - handling missing values, eliminating outliers, transforming values. We conduct sentiment analysis using review text in the second step. Each review text contains review sentences and sentiment polarities. We assume that positive or negative review words will affect review evaluation - positive review sentences can influence more/less than negative review on review evaluation results.

To perform sentiment analysis we are required to build a sentiment dictionary composed of positive and negative words. After building the dictionary we compare sentiment words in review text with the words in the dictionary using text mining. We define 'sentiment score' is the number of positive words minus the number of negative words found in each review.

IV. A Case Study

To verify the feasibility of the suggested framework, we conduct a case study of online game review articles posted on STEAM community site. STEAM is the one of the most a popular game platform in the world. We chose review articles of 11 steam games.

Fig. 3 shows a statistical model to analyze the research framework.

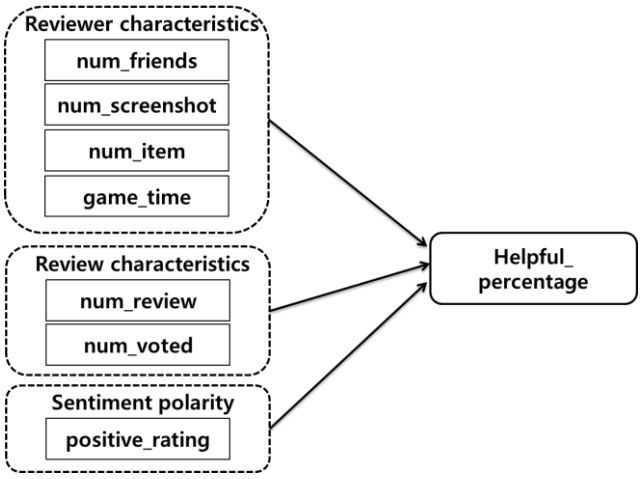

Fig. 3. Analysis model.

As shown in Fig. 3, we chose 'num_friends', 'num_screenshot', 'num_item', 'game_time' as reviewer characteristics, 'num_review', 'num_voted' as review characteristics, 'positive_rating' as sentiment polarity.

A. Research Variables

In this research we extract 8 variables - 7 variables from 31 variables in STEAM community posts and 1 variable from sentiment analysis as shown in Table I. The usability of the review, 'helpful_percentage', was designated as a dependent variable and others, dependent. Table I shows 8 research variables and their meanings.

In this study, the analysis model of a case study was verified through the regression analysis. Linear Regression is an approach for modeling the relationship between one or multiple scalar independent variable(s) and one scalar dependent variable using linear weighted combination of independent variables.

TABLE I: Research Variables

Variable	Description	Role
Helpful_percentage	Helpful percentage of review	Dependent
Num_review	Number of review	Independent
Num_voted	Number of voted helpfulness	Independent
Positive_rating	(Positive word percentage - negative word percentage) of each review	Independent
Num_friends	Number of reviewer's friends	Independent
Num_screenshot	Number of screenshot	Independent
Num_item	Number of item	Independent
Game_time	User game time for last 2 weeks	Independent

B. Descriptive Statistics

Table II shows fundamental descriptive statistics including maximum, minimum, mean and standard deviation.

TABLE II: Descriptive Statistics of Variables

Variables	Min	Max	Mean	S.D
Helpful_percentage	0.00	1.00	.7785	.30597
Num_friends	0	2222	89.27	134.700
Num_review	0	910	7.03	18.092
Num_screenshot	0	10808	155.41	517.803
Num_voted	1	22026	220.72	1410.450
Num_item	0	13333	114.15	386.678
Game_time	0	332	10.65	22.865
Positive_rating	-20	17	.83	1.624

Table III shows a correlation value among 8 research variables. All seven dependent variables have positive coefficients with dependent variable and these coefficient values are statistically significant ($p < .05$).

TABLE III: Correlation Analysis Result

Variables	Helpful percentage	Num_ friends	Num_ review	Num_ screen shot
Helpful_ percentage	1	.190**	.071**	.078**
Variables	**Num_ voted**	**Num_ item**	**Game time**	**Positive_ rating**
Helpful_ percentage	.050**	.140**	.087**	.018**

Table IV shows us all independent variables are significant in predicting the dependent variable since p-values are less than 0.05. Therefore seven dependent variables - the number of reviews, the number of votes for helpful, positive words count - negative words count of each review post, the number of reviewer's friends, the number of screenshots, the number

of items, user game time for last 2 weeks - have significant influences on the dependent variable, helpful_percentage.

Since the usability of the review was calculated by the ratio of the number of votes for helpful useful and total number of votes, the usability of review became smaller as the number of voters increase.

The number of friend (t=33.912, p<0.001) and number of item (t=20.227, p<0.001) had the greatest influence. That is, it was confirmed the reviews of people who had many friends and many items were evaluated to be more useful.

No multicollinearity problem is found since VIF (Variance Inflation Factor) values are lower than recommended level, 10 in Table IV.

Consequently, it was verified that not only the reviewer characteristics influence the usability of review, but also review characteristics. Review article's sentimental polarity can affect gamer's usability decision.

TABLE IV: MULTIPLE LINEAR REGRESSION ANALYSIS

	UNSTANDARDIZED COEFFICIENTS		STANDARDIZED COEFFICIENTS	T	SIG	VIF
	B	S.D	BETA			
NUM_ FRIENDS	.000	.000	.180	33.912	.000	1.201
NUM_ REVIEW	.001	.000	.048	9.048	.000	1.179
NUM_ SCREEN SHOT	1.485E-05	.000	.025	4.850	.000	1.176
NUM_ VOTED	-6.582E-06	.000	-.031	-6.031	.000	1.149
NUM_ ITEM	7.985E-05	.000	.103	20.227	.000	1.103
GAME TIME	.001	.000	.042	8.271	.000	1.075
POSITIVE_ RATING	.003	.001	.015	3.128	.002	1.003

MODEL ADJUSTED-R SQUARE : 0.059, F : 358.60514175735, P-VALUE : 0.000

TABLE V: R-SQUARE CHANGE IN HIERARCHICAL REGRESSION

ADDED VARIABLE	R	BETA	R^2	ΔR^2	ΔF
NUM_ FRIENDS	.201	.727	.040	.040	1690.236***
NUM_ ITEM	.229	.000	.053	.012	515.579***
NUM_ REVIEW	.236	7.985E-05	.055	.003	124.286***
GAME_ TIME	.239	.001	.057	.002	74.809***
NUM_ VOTED	.240	.001	.058	.001	21.417***
NUM_ SCREENSHOT	.242	1.485E-05	.059	.001	37.290***
POSITIVE_ RATING	.243	.003	.059	.000	9.786**

<0.05, *<0.01

To consider relative importance of the independent variables, we build a hierarchical causality model and analyzed model using a hierarchical regression method.

The results of the hierarchical regression are shown in Table V.

1st column denotes added the variables of hierarchical models and R-square changes. As shown in Table V, all hierarchical models have significant R-square changes.

V. CONCLUSION

In this study we verified the influence of three aspects on review usability using a multiple regression method. 7 variables extracted from review data set and 1 data was mashed up with reviewers' using sentiment analysis. All independent variables are significant in predicting the dependent variable.

It was confirmed through the study that the reviews of users having many online friends and the reviews of users having many items had the greatest influence on the usability of the game review. Although the positive coefficient, a variable drawn by the sentiment analysis, was expected to have a great effect on the usability of the game review, it turned out to have the smallest influence among seven dependent variables.

However, this study has some limitations: i) We need to divide reviewers considering reviewers demographic information to gain more sophisticated analysis. ii) Gamers, the readers of review posts, are needed to be considered since the same review posts may have different meaning to individual gamers.

Since only natures the game served by STEAM were considered, most games were limited to PC games. It is expected that it will be difficult to apply this study to the games of other fields such as the mobile game or VR game of which market is becoming bigger recently.

In the future we expect this research will be expanded to supplementary works on sentiment analysis using game-specific sentiment dictionary.

REFERENCES

[1] S. –B. Park and N.-H. Chung, "A factors affecting self presentation desire and participation intention in online game," *The e-Business Studies*, vol. 11, no. 5, pp. 291-308, 2010

[2] *White Paper on Korean Games*, 2015.

[3] S. -H. Kim and C.-H. Kim, "Antecedents and consequences of addiction in online game," *The e-Business Studies*, vol. 7, no. 5, pp. 55-81, 2006.

[4] J. Kwon and K.-Y. Lee, *A Study on the Determining Factors of Online Review Helpfulness*, 2012, pp. 205-211.

[5] K.-Y. Lee, "What make users trust the words of online reviews?" *DAEHAN Association of Business Administration, Korea*, pp. 598-611, 2016.

[6] J.-Y. Kim, E.-K. Suh, and K.-S. Suh, "Effects of perceived similarity between consumer and product reviews on consumer behaviors," *Korean Journal of Business Administration*, vol. 18, no. 3, pp. 68-90 , 2008.

[7] D.-H. Byun, "Evaluating user experience of smart television using emotional representation language," *Journal of the Korea Contents Association*, vol. 15, no. 5, pp. 12-11, 2015.

[8] H. J. Hwang, H. R. Shim, and J. Choi, "Exploration of user experience research method with big data analysis: Focusing on the online review analysis of echo," *Journal of the Korea Contents Association*, vol. 16, no. 8, pp. 517-528, 2016.

[9] J.-H. Lee and C. Park, "A Comprehensive review of online game behavior researches," *The e-Business Studies*, vol. 13, no. 1, pp. 293-322, 2012.

[10] U. Tae-Sun, *Game Planning & Design*, Infopub, 2003.

[11] S.-J. Yoon and T.-S. Yoon, "Analysis of Game data element and Management type in Online Game," *Journal of the Korea Computer Association*, no. 9, pp. 71-79, 2006.

Lee-Sac Lee is a master's candidate in Interaction Design School at Hallym University. He majored in business administration at Hallym University. He is a research associate in CPRC (Communication Policy Research Center). His research interests include data mining, interaction design and big data analytics.

Sang-Hyun Lee is a master's candidate in Interaction Design School at Hallym University. He majored in business administration at Hallym University. She is a research associate in CPRC (Communication Policy Research Center). His research interests cover social network analysis, interaction design and big data analytics.

Hyun-Seok Hwang is a professor of business administration and a research fellow of Hallym Business Research Institute at Hallym University, Chuncheon, South Korea. He received his PhD in industrial engineering from the Pohang University of Science and Technology (POSTECH), South Korea. His current research focuses on big data analytics, opinion mining, interaction design.