International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.12 (2016), pp.237-246
http://dx.doi.org/10.14257/ijmue.2016.11.12.22

Correlation Analysis: Game Professional Score and User Score on Steam

Hyemin Park[1] and *Haewon Byun[2]

1Department of Statistics, Sungshin Women's University, Korea
2School of Information Technology, Sungshin Women's University, Korea
[1]srg700@naver.com, [2]hyewon@sungshin.ac.kr
**Corresponding Author*

Abstract

Game ratings have been one of the most important factors in the online game purchase. The game ratings are numerical scores and its two types are professional score and user score. This paper presents the experiments focusing on the relationship between professional score and user score across 2,069 games distributed by the Steam platform. We also investigated a various game genres for the analysis of the relationship between two rating scores. A better understanding of the relationships of the two scores would be of great value when game users want to choose their favorite games. Our experiments found that indie game and casual game has no differences between two scores. Notably, most major games which are developed by some big game companies have differences between professional scores and user scores. Another more factors can affect the game ratings for the big major games, such as advertising and promotion.

Keywords*: Game rating, Game review, Steam, Metacritic, Metascore, User score, Correlation analysis, Indie game, Casual game, Game genre*

1. Introduction

To purchase a game, game users have investigated many factors such as game trailer, screen shot, developer, distributor, review and game rating. Game is a sort of experience product, so before buying the product, they need some opinion of another people to play the game previously. Especially, game users focus on review and game rating and they visit to a website, Metacritic which is a world largest website that aggregates game reviews [1]. In Metacritic, they provide two types of game ratings which are numeric scores [2]. One is the professional score which is evaluated with reviews provided by game trailers and professional reviews on game webzines. The other is the user score evaluated by the average of many user scores in the website. Usually, game users decide if they buy a game by considering both professional score and user score. However, the scores are often different each other: one score is good and the other is average or bad. In this case, it is important that game users know what score is more reliant.

In movie area, there have been many researches of relationship between grading marks and sales number [3]. This paper is to measure the impact of user reviews on box office performance of movies. In the case of movie which is also an experimental good such as game, there are differences between professional movie score and user movie score by movie genre [4]. This means that different movie audiences by genre may affect user grading [5]. In other research, movie release year, major film distributor and user score have impact on movie box office [6].

Game is different from movies with many characteristics. For example, movie release time depends on summer vacation or winter vacation season because it is important if movie customers can visit cinema. In contrast to movie, game release date depends on game trend at that time. There are also some differences between indie movie and indie

ISSN: 1975-0080 IJMUE

game. Indie movie is minor and we can watch an indie movie in a small theater. In the case of indie game, even though it is a small indie game, many people often focus on the indie game with high grading score. With the different characteristics of movie and game, it is required to measure the impact of game user reviews on game sales [7].

In this paper, we present experiments performed on 2,069 games via the Steam platform [8]. Data was collected covering professional scores and user scores, as well as information on the genre and release year. The results provide insights into the relationship between professional score and user score and what differences are between game genre and release year. Our experimental results help game users to choose games to purchase and play.

The structure of this paper is as follows: Section 2 presents related work in ratings. In Section 3, we introduce the two websites, Metacritic and Steam, and in Section 4, we deal with data collection and filtering and experiment design. In Section 5, experimental results and analysis are presented. Section 6 offers brief concluding remarks and present future works.

2. Related Work

There are some researches on the analysis of Steam users' behavior. Sifa *et al.* presented relationship between Metacritic's professional score and the game sales [9]. In game ratings study, to predict Amazon score of the video game review, Myana and Tumkur used the features of review length, review timestamp and length of review summary [12]. Brodzki said that people tend to buy or try things after confirming critics rave about them [13]. The author presented a slight positive correlation between popularity and Metacritic professional score in a top 100 popular games. On the other hand, in the studies of movie rating, several studies predicted IMDB movie ratings. In order to predict the ratings, they used social media like tweets from Twitter and comments from YouTube [10]. The other study used Google Trends and Google AdWords Statistics [11]. Jacobsen presented that expert review influenced consumer ratings in a brewing industry [14].

3. Metacritic and Steam

Metacritic is a website that aggregates reviews of music albums, games, movies, TV shows, DVDs, and formerly, books. For each product, a numerical score each review is obtained and the total is averaged [2]. Especially, a rating of each professional is weighted score by Metacritic [15]. This site has two ratings that is metascore as professional rating and userscore. (See Figure 1)

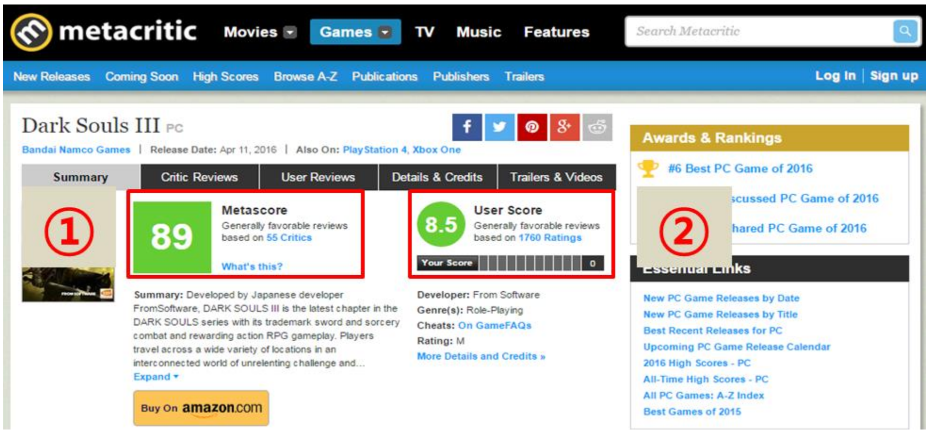

Figure 1. Each Game Page [16] 1: Metascore 2: Userscore

The metascore has a value from 0 to 100 points, and userscore has a value from 0 to 10. A green score which has a value from 75 to 100 points means "good", an yellow score which has a value from 50 to 75 points means "average", and a red score which has a value from 0 to 49 points means "bad". (See Figure 2)

General Meaning of Score	Movies, TV & Music	Games
Universal Acclaim	81 - 100	90 - 100
Generally Favorable Reviews	61 - 80	75 - 89
Mixed or Average Reviews	40 - 60	50 - 74
Generally Unfavorable Reviews	20 - 39	20 - 49
Overwhelming Dislike	0 - 19	0 - 19

Figure 2. Range of Ratings and Color in Metacritic [15]

Steam is a digital distribution platform (such as Origin, Uplay) developed by Valve Corporation offering digital rights management (DRM), multiplayer gaming and social networking services [17-18]. The reason why we target of steam in this paper is that Steam has over 125 million active users [19]. Compared to other platforms, Steam provides the user with convenient system such as installation and automatic updating of games on multiple computers, and community features such as friends list and groups, and *etc*. Furthermore, it additionally provides the user with system that can search games by genres. (See Figure 4)

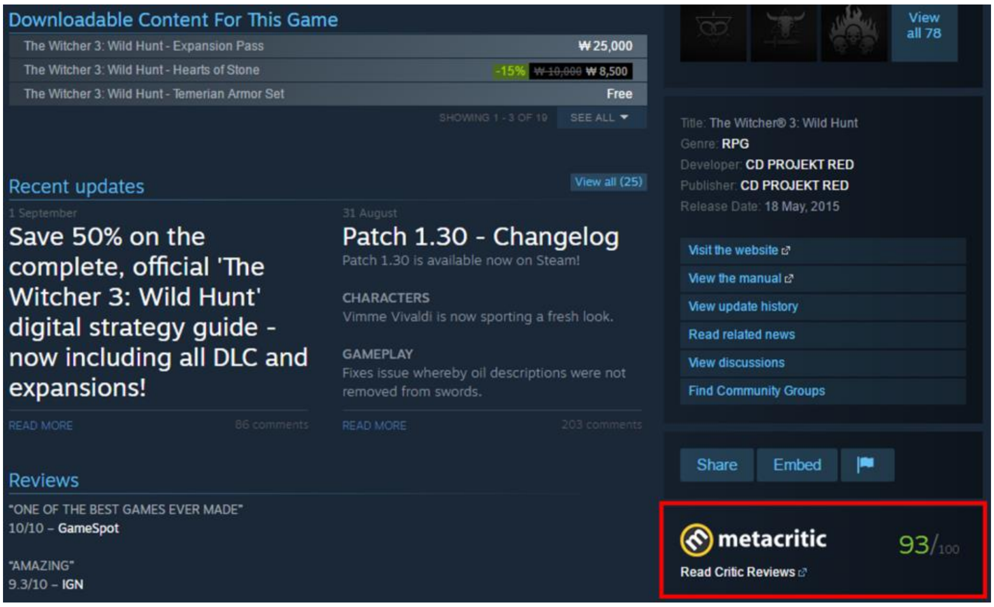

Figure 3. Metascore in Steam [8]

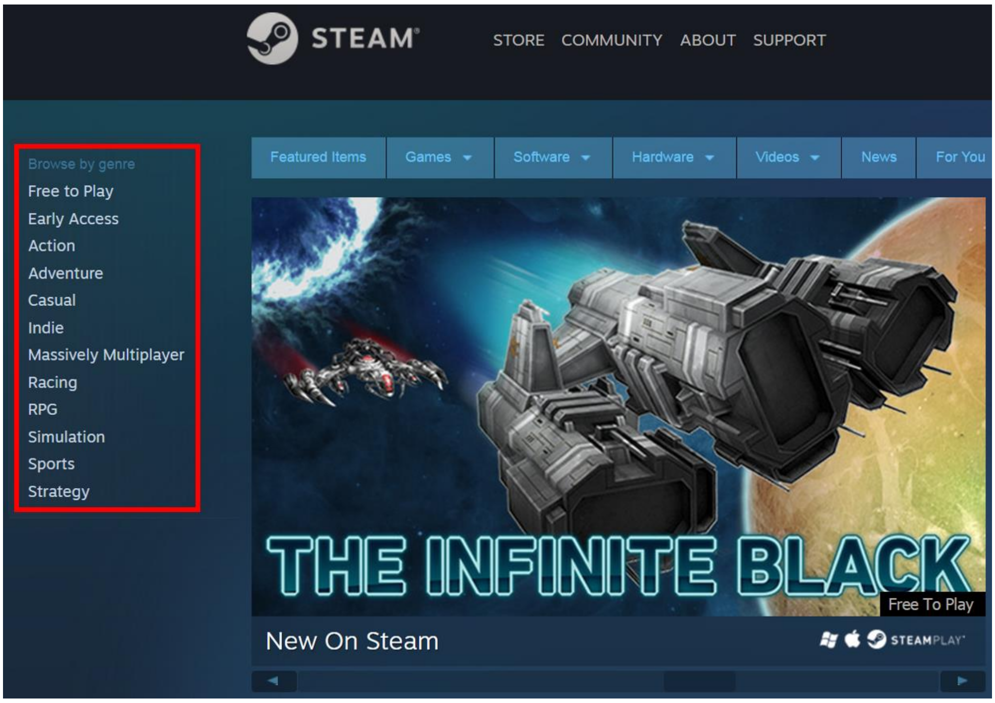

Figure 4. Genres Search in Steam [8]

4. Experimental Setup

4.1. Dataset and Data-Preprocessing

The dataset consists of game title, metascore, userscore, released year, genre variable. Game title, metascore, userscore, released year variables were collected from Metacritic, and genre variables were collected from Steam. Table 1 presents the description of variable about dataset.

```
...
doc = Jsoup.connect("http://www.metacritic.com/browse/games/"
        + "score/metascore/year/pc/filtered?sort=desc&page="
        + site[i][j]
        + "&year_selected="
        + year[i]).userAgent("Mozilla").get();

product_rows = doc.select(".product_row.game");
System.out.println(product_rows.size());

for(x=0;x<product_rows.size();x++){
        product_row = product_rows.get(x);
        metascore = product_row.select(".product_item.product_score").text();
        game = product_row.select(".product_item.product_title a").text();
        userscore = product_row.select(".product_item.product_userscore_txt").text();
        date = product_row.select(".product_item.product_date").text();
...
```

Figure 5. A Part of Jsoup HTML Parser Coding

In order to collect Metacritic's data, we use a jsoup HTML Parser [20] – one of HTML Parser Open Source – to obtain metascore and userscore, which allowed us to get PC game ratings (including Steam) for games released in the interval from 2011 to 2015.

Table 1. Variable of Dataset

Source	Variable	Description
Metacritic	Name	Game Title
	Year	Released Date
	Metascore	Professional Score
	Userscore	User Score
	userX10	Normalized User Score
Steam	Free to Play	Genre – Free Game
	Action	Genre – Action
	Strategy	Genre – Strategy
	Casual	Genre – Casual
	MM	Genre – Massively Multiplayer
	Sports	Genre – Sports
	Indie	Genre – Indie
	Adventure	Genre – Adventure
	Simulation	Genre – Simulation
	Racing	Genre – Racing
	RPG	Genre – RPG
	Early	Genre – Early Access

Jsoup Parser is Java HTML Parser which provides a very convenient API for extracting and manipulating data.

We observed 2,069 games from Metacritic. However, we know that the userscore field can have the value "tbd", which means "No user score yes. Awaiting 3 more ratings". Some games are not registered in Steam. For object of experiment, these games are eliminated. Furthermore, we find that rating's range is difference. So, we decide to normalize userscore. In order to normalize it, we calculate userscore * 10, userX10. (See Table 2)

Table 2. A Part of Metacritic Data

Name	Year	Metascore	Userscore	userX10
Game A	2014	61	6.3	63
Game B	2015	60	8.2	82
Game C	2015	80	8.2	82

In Table 1, Steam has 11 Genres, which is Free to Play, Action, Strategy, Casual, Massively Multiplayer, Sports, Indie, Adventure, Simulation, Racing, RPG, Early Access. However, we exclude that Early Access game because the number of that game is too small and not official launch game. And in this paper, we don't deal with the Free game. Therefore, we use 10 genres in this paper. Table 3 shows a part of dummy variables like 0/1 about genres. A value '1' of Action variable means that this game is Action genre. Since almost games have one or more genres, we made the variable as shown above. In Table 4, sum of game is gradually increasing with the passage of time. Especially, Indie game is highly increasing with the passage of time.

Table 3. Example about Genre Variable from Steam

Name	Action	Strategy	Casual	…	RPG
Game A	1	0	0	…	0
Game B	1	0	0	…	0
Game C	0	0	1	…	0

Table 4. The Number of Game by Genre and Year

Genre	2011	2012	2013	2014	2015	Sum
Action	61	78	102	98	106	445
Strategy	27	36	42	42	53	200
Casual	14	13	16	17	11	71
MM	4	4	6	8	0	22
Sports	3	3	8	10	9	33
Indie	45	67	102	126	137	477
Adventure	29	51	74	82	95	331
Simulation	12	13	23	23	29	100
Racing	6	4	8	8	9	35
RPG	24	20	27	46	50	167
Early Access	0	2	0	0	0	2

4.2. Experimental Method

We tried three types of experiments. For each experiment, we analyzed linear correlation analysis between metascore and userscore, and performed the simple correspondence analysis to check out the difference between the two scores using SAS 9.4.

The experiments are following list.

Case1) We analyzed the two scores by genre.

Case2) We analyzed the two scores by year.

Case3) We analyzed that the two scores compared with between Indie game and Major game.

In correspondence analysis, our study is that two scores are different, which would give us the following null (denoted H0) and alternative (denoted H1) hypotheses:

$$H_0 : metascore = userscore$$

$$H_1 : not\ H_0$$

5. Result

Before we start experiments, we investigate two score's correlation and difference in all dataset. The result is that correlation coefficient is positive (0.5903) and two scores difference is 2.1708, and P-value is <.0001. So, we reject H0 in all dataset.

In the first experiment, we analyzed the two scores by genre. Result of correlation analysis is that all correlation coefficient is positive. In Table 5, Casual game has the lowest coefficient and Simulation game has the highest coefficient. Result of simple correspondence analysis is that only Indie and Casual genre rejected H0 and the others did not reject H0. This result shows that the two scores of the two genres are not different. Also, all average of difference between the two scores is positive. This means metascore is bigger than userscore.

Our second experiment analyzed the two scores by year. The linear correlation of this experiment gradually had decreased with time (See second columns in Table 6) In particular, the correlation coefficient is more than 0.7 in 2011.

Table 5. Result of Experiment by Genre. D_Means is Average of Difference between the Two Scores

Genres	Corr	t-Value	p-value	D_Mean
Action	0.618	3.71	**0.0002***	2.0697
Strategy	0.595	2.44	**0.0154***	1.8250
Casual	0.339	0.36	0.7183	0.4648
MM	0.687	2.79	**0.0111***	7.0455
Sports	0.476	5.53	**<.0001***	12.1515
Indie	0.565	0.58	0.5600	0.2746
Adventure	0.588	2.18	**0.0298***	1.3202
Simulation	0.725	3.32	**0.0013***	3.6600
Racing	0.574	3.36	**0.0019***	5.9429
RPG	0.635	3.19	**0.0017***	2.8503

(p-value < 0.05: *)
Corr : Correlation coefficient

Our second experiment analyzed the two scores by year. The linear correlation of this experiment gradually had decreased with time (See second columns in Table 6) In particular, the correlation coefficient is more than 0.7 in 2011.

Table 6. Result of Experiment by Year

Year	Corr	t-Value	Pr > |t|	D_Mean
2011	0.734	3.67	**0.0004***	3.5982
2012	0.623	1.28	0.2046	1.2031
2013	0.552	0.40	0.6914	0.3520
2014	0.540	2.21	**0.0284***	1.6737
2015	0.557	4.49	**<.0001***	3.9662

(p-value < 0.05: *)
Corr : Correlation coefficient

For the reason of the first experiment's results that Indie game's the two scores is same, this shows the following results. Figure 6 shows that the average of difference between the two scores reduced from 2011 to 2013. However, average of difference between the two scores increased from 2013 to 2015. This results are the opposite of what growth rate of Indie game. (See Figure 7)

The last experiment is that compared with between Indie genre and Major game. (See Table 7 and Table 8) In correlation analysis, all correlation coefficient is positive. In particular, the correlation coefficient is more than 0.7 in 2011.

In totally, the Indie game shows the difference between the two scores and the Major game doesn't show that. In Indie game, userscore is higher than metascore in all years except for 2015. 2015 p-value is almost significance level 5%, but we conclude that it is different. In Major game, metascore is higher than userscore in all years except for 2013.

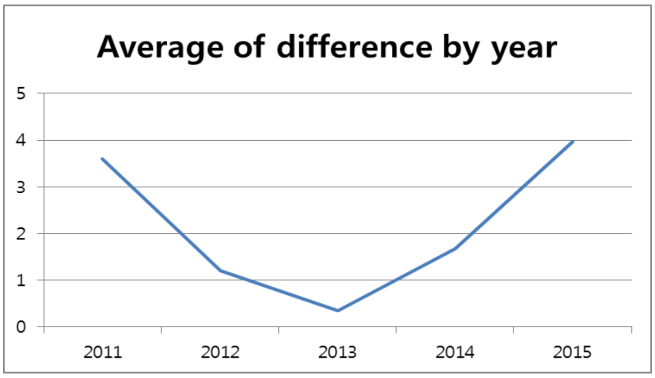

Figure 6. Average of Difference (= D_Mean) by Year

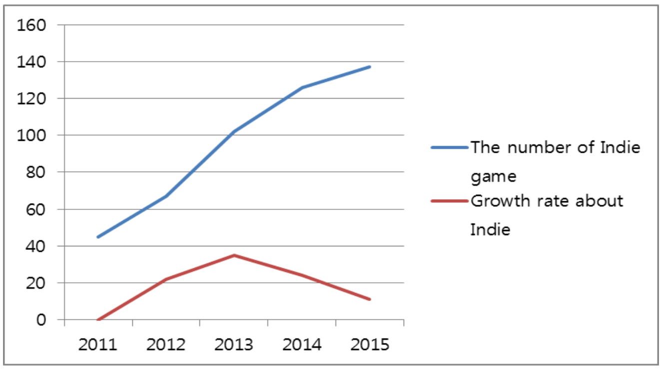

Figure 7. The Number of Indie and Growth Rate about the Game

Table 7. Indie Game Experiment Results

	Indie game			
Year	**Corr**	**t-Value**	**Pr > |t|**	**D_Mean**
2011	0.787	-0.35	0.7287	-0.4
2012	0.65	-0.98	0.3316	-1.3433
2013	0.55	-0.26	0.7943	-0.2549
2014	0.559	-0.16	0.8711	-0.127
2015	0.461	2.01	**0.046***	2.0511
Total	0.565	0.58	0.56	0.2746

(p-value < 0.05: *)
Corr : Correlation coefficient

Table 8. Major Game Experiment Result

	Major game			
Year	**Corr**	**t-Value**	**Pr > |t|**	**D_Mean**
2011	0.732	4.63	**<.0001***	6.2836
2012	0.63	3.35	**0.0014***	4.0000
2013	0.552	0.72	0.4738	1.1558
2014	0.569	3.35	**0.0014***	5.2188
2015	0.672	4.8	**<.0001***	7.7143
Total	0.63	7.09	<.0001	4.8024

(p-value < 0.05: *)

Corr : Correlation coefficient

In totally, the Indie game shows the difference between the two scores and the Major game doesn't show that. In Indie game, userscore is higher than metascore in all years except for 2015. 2015 p-value is almost significance level 5%, but we conclude that it is different. In Major game, metascore is higher than userscore in all years except for 2013.

6. Conclusion and Future Work

Here analyses have been presented focusing on the relationship between professional score and user score of the digital games on game distribution platform Steam, covering 2,069 games. Results reveal that there is no correlation between the two scores in most of the games. Additionally, we have an experiment on each genre of games with the two scores. Results indicate that most games have differences between professional score and user score except casual game and indie game.

The analysis also shows that the professional scores and user scores of the most games have correlation in the year of 2012 and 2013 because the number of indie games increased in 2012 and 2013. In the case of major games, there is no correlation between two scores. When game users purchase digital games on online game platform, they are better to consider that the professional score of indie game has correlation with the user score. However, the professional score of major game has not correlation with the user score.

Future work will focus on more detailed analyses of the relationship between review score and sales number in order to predict how many games will be sold in the early release time. Addition to review scores, we will also cover game properties such as developers, distributors, game buyers, duration after release. Multiple genre of a game is also important issue because one game can have one more genre.

Acknowledgments

This work was supported by the Sungshin Women's University Research Grant of 2016.

References

[1] http://www.metacritic.com/.
[2] http://en.wikipedia.org/wiki/Metacritic.
[3] P. K. Chintagunta, S. Gopinath and S. Venkataraman, "The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets", Marketing Science. vol. 29, no. 5, **(2010)**, pp. 944-957.
[4] I. Kim, K. Chun and H. Lee, "The Effect of Professional Critics' Reviews on Online User Reviews and Box Office: US Motion Picture Industry", Korean Academy Of Management., vol. 20, no. 3, **(2012)**, pp. 1-27.
[5] G. Hsu, "Jacks of all trades and masters of none: Audiences' reaction to spanning genres in feature film production", Administrative Science Quarterly., vol. 51, no. 3, **(2006)**, pp. 420-450.
[6] S. Park and W. Jung, "The Determinants of Motion Picture Box Office Performance: Evidence from Movies Released in Korea, 2006-2008", Journal of Communication Science., vol. 9, no.4, **(2009)**, pp. 243-276.
[7] https://goo.gl/BjRszV.
[8] http://store.steampowered.com/.
[9] R. Sifa, A. Drachen and C. Bauckhage, "Large-Scale Cross-Game Player Behavior Analysis on steam", Proceedings of the Eleventh Artificial Intelligence and Interactive Digital Entertainment International Conference (aiide 2015), **(2015)**, pp. 198-204.
[10] A. Oghina, M. Breuss, M. Tsagkias and M. de Rijke, "Predicting IMDB Movie Ratings using Social Media", Proceedings of the 34th European Conference on Advances in Information Retrieval, Barcelona, Spain, **(2012)**, pp. 503-507.
[11] D. Demir, O. Kapralova and H. Lai, "Predicting IMDB Movie Ratings using Google Trends", Dept. Elect. Eng., Stanford Univ., California, **(2012)**.
[12] P. K. Myana and K. Tumkur, "Score Prediction of Amazon Video Game Reviews through Collaborative Filtering Techniques", **(2015)**.

[13] E. Brodzki, "Statistical Analysis of Game Behavior", Statistical Analysis of Gamer Behavior, **(2010)**.
[14] G. D. Jacobsen, "Consumers, Experts, and Online Product Evaluations: Evidence from the Brewing Industry", Journal of Public Economics, vol. 126, **(2015)**, pp. 114-123.
[15] http://www.metacritic.com/about-metascores.
[16] http://www.metacritic.com/game/pc/dark-souls-iii.
[17] http://en.wikipedia.org/wiki/Steam_(software).
[18] http://en.wikipedia.org/wiki/Digital_distribution.
[19] http://goo.gl/CQ27TW.
[20] http://jsoup.org.

Authors

Hyemin Park, is a senior in Department of Statistics of Sungshin Women's University. Her research interests include game ratings, game reviews, machine learning, data-mining, big data analysis, text-mining and sentiment analysis.

Haewon Byun, is a professor in School of Information of Sungshin Women's University. Her research interests include computer graphics, virtual reality and big data analysis.