

International Journal of Applied Business and Economic Research

ISSN : 0972-7302

available at <http://www.serialsjournal.com>

© Serials Publications Pvt. Ltd.

Volume 15 • Number 20 • 2017

A Study of Analyzing STEAM Game Review Data using Text Mining

Ha-Na Kang¹, Hye-Ryeon Yong² and Hyun-Seok Hwang^{3*}

^{1&2}Graduate School of Interaction Design, Hallym University.

³Dept. of Business Administration of Business, Hallym University.

*Corresponding Author. E-mail: hshwang@hallym.ac.kr

Abstract: Due to the construction of infrastructure of wired and wireless networks and rapid development of the speed, digital based data, which is no longer manageable with general technology, is increasing explosively and its form and quantity are tremendous. In accordance with this, the application plan of previously unused data and the area of value creation through this are gradually widened. Especially, the importance of text data analysis, which represents the public's opinion such as social network services (social media) and online product reviews, is magnified. Like this, the development of comprehension and prediction solutions of customer needs that utilize the reviews is estimated to optimize the values of all the future industries and technologies and is expected to become the base of upcoming economic effect creation.

In other words, seeing the hidden value through data can suggest pending issues that enhance the competitiveness of the company. Especially, as extensive amount of data is created in games, it is deemed as a promising business that can expect high growth through future data analysis.

However, the research on the opinion analysis based on the importance of game reviews and texts is unsatisfactory.

Therefore, in this research, we would like to look into utilization of review data within a game and examine text data analysis technique and application plan based on the existing preceding research. In addition, through opinion mining, we have tried to investigate what are some major keywords per topic within a game and compare what are the characteristics of each game that can be inferred through comparison of two analytical techniques. Moreover, we wanted to suggest practical implications for creating economic value such as game sales and system improvement through utilization of review data in the game industry in the future.

Keywords: Game Review, Opinion Mining, Word2Vec, Topic Modeling, Big Data.

1. INTRODUCTION

Popular use of smart devices generates voluminous amounts of structured or unstructured data that organizations can potentially mine and analyze for business gains. Especially, the increase of text data in its speed of creation and quantity are so significant that it is almost unmanageable.

In compliance with this trend, the researches utilizing text data are actively progressed. One of the most popular text analysis is mining the public's opinions in social network services, blog, or text reviews.

If the vast amount of text data can be refined and made into information, it can not only have value beyond the form of vast amount of data but also helpful in industry decision making.

In the research of Lee *et al.*, (2015), through mining the customers' review of domestic and international clothing products sold online, they can uncover the problems in purchase process, the differences in product features and customer needs (Lee and Lee, 2015).

In addition, Hwang (2016) tried to identify the customer's perception of automobile brands and brand cluster through opinion mining of social media posts (Hwang, 2016).

The value of information mined from text data can also be applied to the gaming industries. The user evaluation of a game can be analyzed to discover user's emotional polarities like positive/negative opinions on the games, comparison with other games with the same genre.

STEAM, the largest game platform, is no exception; through various formats of social network services, such as Friends, Game Hubs, Discussion, Workshop and Greenlight, users can actively shares their opinions.

Although text reviews on games are increasing, little research on mining the game reviews has been conducted enough.

This research, therefore, focuses on understanding game user's opinion mined from review texts. In this research, we analyze game users' review posts on STEAM, and tries to mine texts using 'Word2Vec' and 'Topic Modeling' techniques.

The composition of this paper is as follows. In chapter II, the research related to opinion mining and analyzing techniques will be examined and in chapter III, the framework of this research will be suggested. We provides the results of a case study in chapter IV. Chapter V will present the discussion and limitation of this research with future research directions.

2. RELATED WORKS

2.1. Opinion Mining

Opinion Mining, one of the most common application areas of text mining, is a technique of finding patterns in opinions and emotions appeared on text documents like websites and social media(Lee *et al.*, 2014).

Opinion Mining can be applicable to discover meaningful information hidden in product evaluations, brand reviews, and newspaper articles on specific topics Kim and Lim, 2011). This opinion mining can

grasp the reasons as to why people like or hate certain products and services. It can also check how general opinion or public interest change regarding certain issue (Liu and Zhang, 2012).

Online review is a user's written opinion online on goods and services such as clothing, furniture, home electronics, books, and music, etc. Therefore, a review can be a very important factor that affects purchase decision of other users who wants to use the same products and the formation of corporate brand images (Chen and Xie, 2008)

In the research of Kim and Jeong (2013), VOC (Voice of Customer) analysis system was proposed, which categorizes various customer opinions into positive/negative evaluations through opinion mining (Kim and Jeong, 2013). Godes and Mayzlin (2004) stated that online reviews of audience have meaningful influence on the success of a film (Godes and Mayzlin, 2004).

In addition, according to the research of Kim *et al.*(2011), positive and negative opinions on news contents influence the fluctuations of stock prices, and this opinion mining can better explain the flow of the price index of stocks (Kim *et al.*, 2012).

Just like the online reviews that are written after purchasing products, it has been discovered that the opinion sharing in the community when playing games and the user's game purchase review significantly affects other users.

In the research of Kang *et al.* (2017), it is found that the more accumulated evaluations from many users regarding the game, the more useful the reviews are. User reviews enable users to obtain not only information on the game, but decision making to use the game (Kang *et al.*, 2017).

2.2. Word2Vec

In order to process natural language, words should be converted into numerical value. One approach of processing a natural language is representing words in vector space (Kwon, 2016). Word2Vec, one of vector space models, extracts corpus and converts the given corpus into vectors. Representing a word again into vector in the light of the word's meaning and context is called word embedding (Mikolov, 2013).

On the other hand, the distributed representation presents words as dense forms within the vector space, and presents similar meaning words to be placed near each other (Seok *et al.*, 2015).

Word2Vec originally used for the artificial neural network research with the premise that words in the same context have close meanings. By progressing with learning through text, other words that appear in the same sentence as one word are perceived as related words.

Through this learning process, because the related words have high probability of reentering in close places within the text, the words that repeatedly appear together during the learning process are put into the cluster (O *et al.*, 2016).

Kim and Park (2017) found similarity among vocabularies by applying Word2Vec based on the reviews posted by learners who took a course and clustered the words using clustering algorithm. Through this research, the morpheme of before/after taking the courses were presented and compared (Kim and Park, 2017). However, the meaning shall be the same, the spelling of the word may change like past-present-future form, may be perceived as a different word based on the form of the word although the meaning is similar, and abstract verbs or adjectives can produce senseless results to the learning results.

Much attention should be paid to the preprocessing of the vocabulary when text documents is mined with Word2Vec,

2.3. Topic Modeling

Due to active opinion shared online, tremendous amount of typical and atypical data are generated in one day. Especially, the Topic Modeling technique is a proposed method to find hidden topic within the enormous amount of text data. Topic Modeling is a research technique analyzing various trend and major issues by finding dormant topics within the document based on the keywords constituting a text.

Topic Modeling utilizes statistical inference technique, and it is based on the algorithm of Blei *et al.* (2003), LDA (Latent Dirichlet Allocation) (Blei *et al.*, 2003).

LDA is a modeling technique that extracts topics on general text documents. This is a modeling algorithm that reclassifies texts after extracting latent topic information from massive amount of texts (Kim, 2017).

Kim *et al.*, (2012) it proposed a model that automatically recommends TV programs based on the LDA algorithm and user preferences (Blei, 2012). Grimmer (2010) examined in what way they promote their work to the voters on analysis of press release of U.S. Senates by using Topic Modeling (Grimmer, 2010). Bae *et al.*, (2014) tried to confirm the usefulness of Twitter through using Twitter data, and to trace the change in topics over time Bae *et al.*, 2014).

Kang *et al.*,(2013) analyzed the topics in newspaper articles reported on presidential candidates and examined how those topics are connected and how public opinion changed over time (Kang *et al.*, 2013).

Topic Modeling has been used mainly for analyzing newspaper articles or political posts, and public opinions on social network services (SNS) recently.

Little research has been performed to examine major topics and issues of video games and to understand the similarities / differences between Topic Modeling and Word2Vec.

3. RESEARCH FRAMEWORK

The framework of this research is shown in Figure 3.1. In order to analyze clusters and topics of written text documents, the following research framework was proposed.

In the 1st step, we gather an appropriate text data for opinion mining. The possible source of data may be newspaper articles, brand posts, microblogs, shared datasets, and online shopping reviews.

The second step is extracting stem words. Preprocessing the raw data is essential in this step. Data with the duplicate value, missing value, punctuation, stop-word, and white space and any meaningless words like finite/infinite articles are removed.

In step 3 we adopt two text mining techniques, Word2Vec and Topic Modeling, to model stem words extracted in the second step. We map out the stem words in 100~300 dimensional vector space using Word2Vec and find n word-clusters using a clustering algorithm. We also identify n topics in stem words using Topic Modeling.

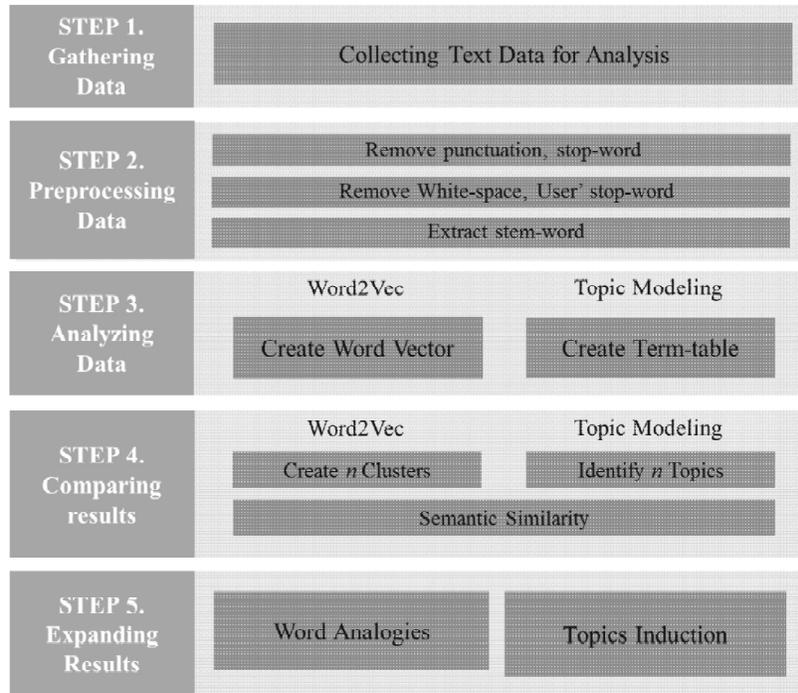


Figure 1: Research Framework

To verify the similarity of mined results in the modeling techniques, we compare two results whether the clusters in Word2Vec model and topics of Topic Modeling model have common theme or not in step 4.

Step 5 provides us an opportunity to expand the results by making up for shortcomings of other technique. One of the most useful advantages in Word2Vec is word analogies calculated by simple vector operations as shown in equation (1). We can infer a counter word of King is Queen based on the relationship of Man and Woman.

$$\text{Man} : \text{Woman} = \text{King} : (\text{Queen}) \quad (1)$$

4. CASE STUDY

In this research, a case study was performed based on the framework proposed in chapter 3. We choose STEAM, a global game platform, developed by Valve Corporation. STEAM supports various community activities, opinion sharing compared to other game platforms.

Through opinion mining of STEAM's review data, we expect to understand STEAM game users' mind and issues of video games.

4.1. Gathering Data

Among 800 games supported by STEAM, we choose 11 popular games containing 79,437 reviews. The data includes game name, text review contents written by game users. Eleven game names are listed in Table 1.

Table 1
Data descriptions

<i>Game Name</i>	<i># reviews</i>
Counter Strike	79,437
Counter Strike Global Offensive	
Dota2	
Arma3	
Sid Meier's Civilization cl	
GTA	
Football Manager	
Garry's Mod	
Team Fortress	
Warframe	
The Elder Scroll	

4. 2 Preprocessing Data

Before analyzing the data, we execute preprocessing. First, the missing value and overlapping values are removed. Punctuations, stop-words (article, preposition, postposition, conjunction, etc.), and white-spaces are also removed.

In addition, meaningless user's stop-word are defined and removed. Sample user's stop-words are listed in Table 2. Finally, 57,575 stem words are extracted after all .

Table 2
User's stop words

can, say, one, way, use, also, howev, tell, will, much, need, take, tend, even, like, particular, rather, said, get, well, make, ask, come, end, first, two, help, often, may, might, see, someth, thing, point, post, look, right, now, think, 've, 're , anoth, put, set, new, good, want, sure, kind, larg, yes, day, etc quit, sinc, attempt, lack, seen, awar, littl, ever, moreov, though, abl, enough, far, earli, away, achiev, draw, last, never, brief, bit, entir, brief, great, lot, game, STEAM, mod, play

4.3. Analyzing Data

Normally Word2Vec creates 100~300 dimensional space to represent word vectors. We use 100 dimensional space to avoid sparse vectors. R, well-known open software for data science, is used to build a Word2Vec model and a Topic Modeling model.

4.3.1 Topic Modeling

In this research, LDA (Latent Dirichlet Allocation) based Topic Modeling algorithm and R-package LDAvis, which visualizes the algorithm, were used to extract keywords LDA, the most representative method of Topic Modeling technique, is useful in scaling down the dimensions of the data, and can extract topics that are contextually meaningful and consistent (Mimno and McCallum, 2008). The number of repetition of sampling and the number of topics that will be extracted should be set to execute modeling in LDA. In this

research, 5,000 times of sampling were repeated and total of 10 topics were extracted. The number of topics is determined after several tries, based on ease of interpretation. Figure 2 and Figure 3 represent visualization of the Topic Modeling result through LDAvis package.

Figure 2 presents the position of 10 extracted topics. This figure is a 2 dimensional map depicting 10 topics and their boundary using Principal Component Analysis.

Therefore, on the screen, the topics may seem to be overlapped but in fact, the locations of the topics are all different. Figure 3 shows 30 words of a topic and their relative frequency. We can found that some words-gta, online and rockstar—are appeared intensively in one topic.

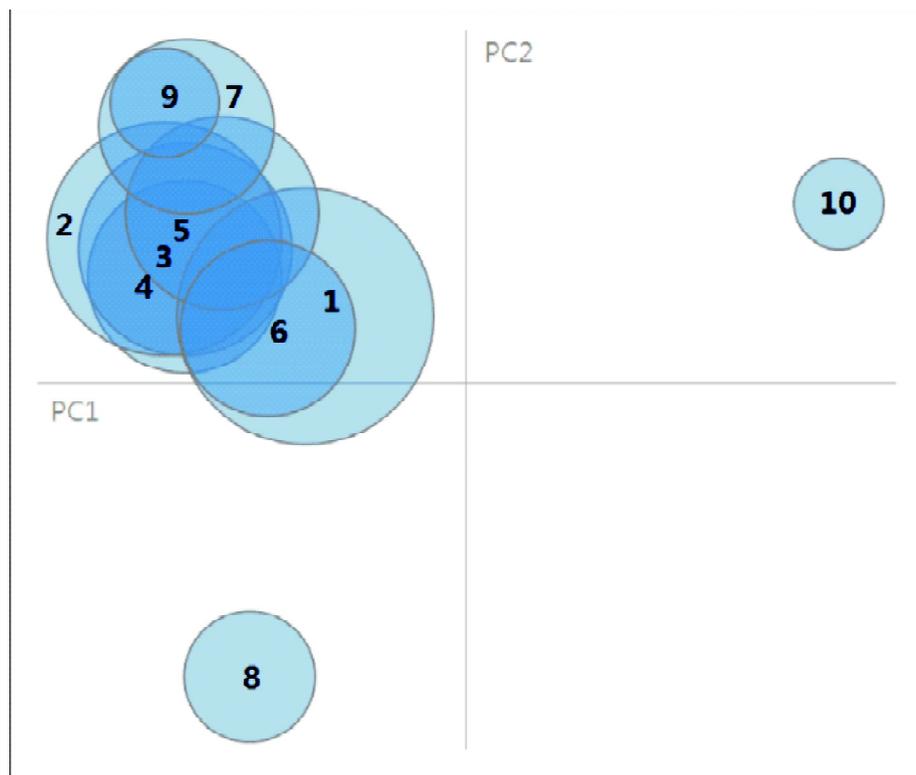


Figure 2: Visualized results of Topic Modeling

4.3.2 Word2Vec

Word2Vec method is one of deep learning methods that vectorizes the words included in the text. Since vector is represented as not just a simple numerical concept but a semantic notion, it has the function of containing a meaning (Heo and Ohn, 2017). Therefore, this makes it possible to not only express complex concepts but also infer and deduct other words (Lee *et al.*, 2017).

The number of clusters was set to 10, just like it was in topic modeling, and 30 words that were close to the center of the 10 clusters were extracted and listed.

This research uses R-package *rword2vec* to vectorize stem words and clusters generated through Word2Vec. We tested an analogy function of Word2Vec to confirm the accuracy of the vector model.

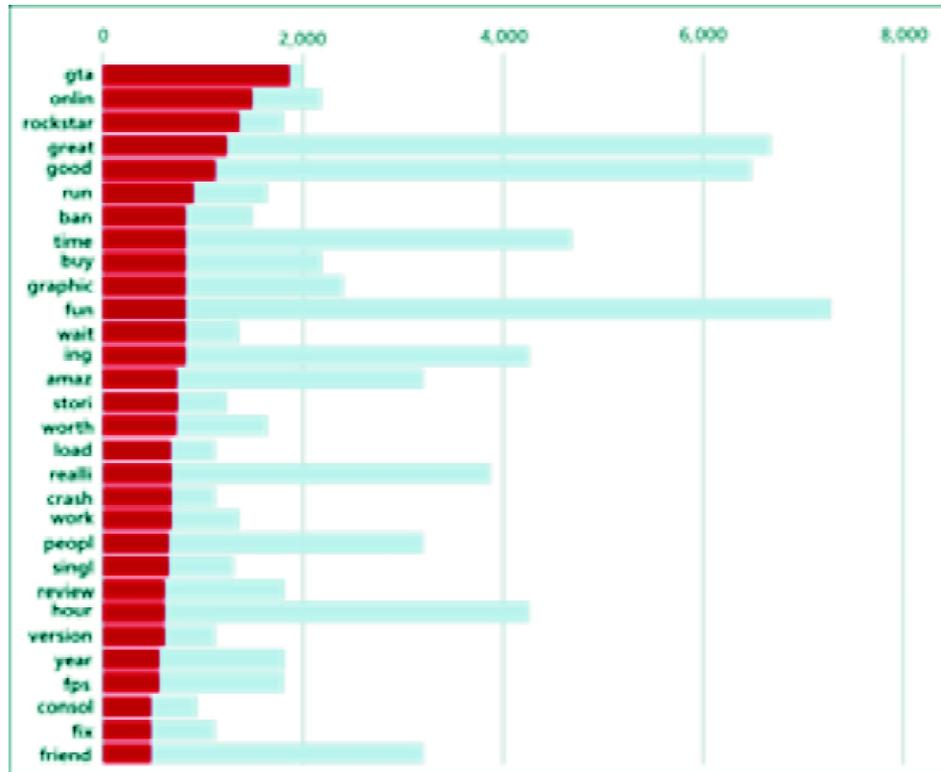


Figure 3: Frequencies of Words in a Topic

In the next chapter we will present word analogies results inferred from the Word2Vec model.

Figure 4 shows implementation codes for the Word2Vec model.

```
1 review.doc <- as.character(review.doc)
2 writeLines(review.doc, con="review_merge_stem.txt")
3 # Remove 1-2 letter words
4 review.doc <- gsub(" *\\b[[:alpha:]]{1,2}\\b *",
5   " ", review.doc)
6
7 # Remove excessive spacing
8 review.doc <- gsub("^ +| +$|( ) +", "\\1",
9   review.doc)
10 review.doc <- gsub("\\\"", "\"", review.doc)
11 review.doc <- gsub(",", ", ", review.doc)
12 writeLines(review.doc,
13   con="review_merge_stem2.txt")
14
15 model=word2vec(train_file =
16   "review_merge_stem2.txt", output_file =
17   "vec_stem.bin", binary=1)
```

Figure 4: R code for Word2Vec

5. RESEARCH RESULT

5.1. Topic Modeling

As a result of conducting Topic Modeling, 10 topics were extracted and several top words of each topic are shown in Table 3.

Table 3
Result of Topic Modeling

<i>Variables</i>	<i>Meaning</i>	<i>Theme</i>	
Rockstar	Manufacturer of GTA	Topic 1: GTA	
Run	GTA4-can you run it version		
Ban	STEAM became an issue by preventing access to the game for a specific user		
Crash	Frequent crashes during download and connection		
Single	Single mod		
Consol	console version		
Hat	Compensation for quest performance	Topic 2: Dota2	
Russian	Ritmix Russian Dota 2 League		
Team	Updated by adding the functionality of “team fortress” to dota		
Fortress			
Dota	Game name		
Military	Military background	Topic 3: Arma3	
War	War		
Multi	Multi mod		
Realli	Gained a good reputation for realistic representation of graphics		
Graphic			
Shooter	Shooter, gun		
Strategy	Strategy/war game.	Topic 4: SMC	
War			
Turn	Turn Based Game		
Addict	‘Just one more turn’ Highly addictive		
Base	The first stage of the game		
Nuke	One of the civilizations in SMC		
DLC	Provide download content		
Gandhi	One of the civilizations in SMC		
Free	Free provided		Topic 5: War-Frame
Warframe	Game Name		
Grind	Raise the level by gaining weapons steadily		

Contd. table 3

<i>Variables</i>	<i>Meaning</i>	<i>Theme</i>
Weapon	Weapons acquisition is related to the game level rise	
Space	Game's concept	
Ninja		
Shooter	Gun	
Skyrim	One of the series	Topic 6: The Elder Scroll
Valv	Distributor: Valve corporation	
Bethesda	Manufacturer of The Elder Scroll	
RPG	The Elder Scroll's genre	
Dragon	Quest to get a dragon	
Quest		
Ding	KING SING RING DING Mini-games from Bethesda	
Garri	Garry's Mod	Topic 7: Garry's Mod
Sandbox	A system that allows you to do anything you want	
Addon	Add-on Features	
Anything	Users can create everything they want	
Manag	Football manager	Topic 8: Football Manager
Year	Age of character (player)	
Footbal	Football manager	
Match	Match making system	
Make		
Version	The game engine update is in progress	
Engin		
Counter	Counter Strike	Topic 9: Counter Strike
Strike		
Classic	Classic version of counter strike	
Year	Game of the year edition :half-life version	
Shooter	Shooter, gun	
Csgo	Abbreviation of Counter Strike Global Offensive	
School	HESL(High School e-Sports League)	
Back	Back to the basic : update version	
Rockstar onlin updat money price buy	Overall, words that represent online games	Topic 10: Etc.

The 10 topics summarized with the words that appeared in each topic. From topic 1 to topic 8, one game is allocated to one topic as shown in the table, but in the case of topic 9, counter strike and counter strike global offensive formed one topic together.

Since counter strike global offensive is the upgraded version of counter strike, this result is understandable.

Topic 10, however, is consisted of words that generally encompass the online games of STEAM, not a specific game, such as rock star, online, update, and buy.

Since Team Fortress could not form one topic, we can infer that Team Fortress has weaker identity than that of other's. It is assumed that since the genre of Team Fortress is FPS and 5 out of 11 games have FPS properties, such as Arma3, Counter Strike, Counter Strike Global Offensive, Dota2, the topics were probably included in other same genre games.

5.2. Word2Vec

The results of Word2Vec are shown in Table 4. We use k-means clustering algorithm to find 10 cluster of word vectors. 10 vectors and several words closest to the center of cluster are presented in Table 4.

The 10 clusters created by Word2Vec are described below.

Table 4
Result of Word2Vec

<i>Variables</i>	<i>Meaning</i>	<i>Theme</i>
Valve be the sda, gtaonlin, payment, websit, servic	Overall, words that represent online games	Cluster 1: Etc.
rpgs ofp Xbone Standalon	Genre of Garry's Mod Flight operation Console model Stand alone'	Cluster 2: Garry's Mod
Civilis Brazil Alexand England Portugues Aztec Napoleon Allianc Conquest	Civilization One of the civilizations in SMC One of the civilizations in SMC One of the civilizations in SMC civilizations in SMC One of the civilizations in SMC One of the civilizations in SMC Napoleon Alliance Conquest	Cluster 3: SMC Cluster3
Communist Spain Merciless Imperi Rogu Attila Orc Judaism	Communist 'Spain Dota 2' professional game team Merciless Empire, absolute power 'Rogue Omen': Game items King of the Xiongnu Orc Related to dota game items	Cluster 4: Dota2

Contd. table 4

<i>Variables</i>	<i>Meaning</i>	<i>Theme</i>
Grenad	Game items	Cluster 5: The Elder Scroll
Parachut	Parachut	
Allahu	Characters	
Minigun	Minigun	
Missil	Missile function	
Vampir	Apprring from series 2	
Rebel	Rebel, 'The Treason of the Rebel'	
Villag	Tribal village	
Aesthet	Aesthetic	Cluster 6: Arma3
Dialogu	Arma3's initial motif 'Weapons secret negotiation'	
Terrain	Various terrain appeared in Arma 3	
Multitud	The crowd is divided into several teams	
Orient	Orient countries	
Combo	Weapon Combo Function	
Goof	Go off	
Deathrun	Deat run Mod	
Jailbreak	Jailbreak Mod	
Heap	Computer data structure, Counter Strike Frequent errors occurred when running by version	
Glich	Glich	Cluster 8: GTA
Premis	There was a lot of controversy about the GTA's violent premise	
Spooki	Appears in Halloween Version	
Rec	GTA's REC Server	
Migrat	'Migration control' Mission	
Blob	de blob : similar game to GTA	
Framer	War frame	
Geforc	Promoting Warfare Game Money for GeForce buyers	
Leak	Coolant Leak	
Radeon	Promotion to present Radeon when purchasing a graphics card	
Spike	Missil	
Bout	Bout, Match	Cluster 10: Football Manager
Colleg	College	
Yrs	Age of character (player)	
Suspend	Suspended Game	
Thirteen	In the West, there is fear of 13. So the players do not have 13 as their number.	
Mark	man to man	
Troubleshoot	Trouble shooting	

The following Figure 5 represents the comparison of two analysis results.

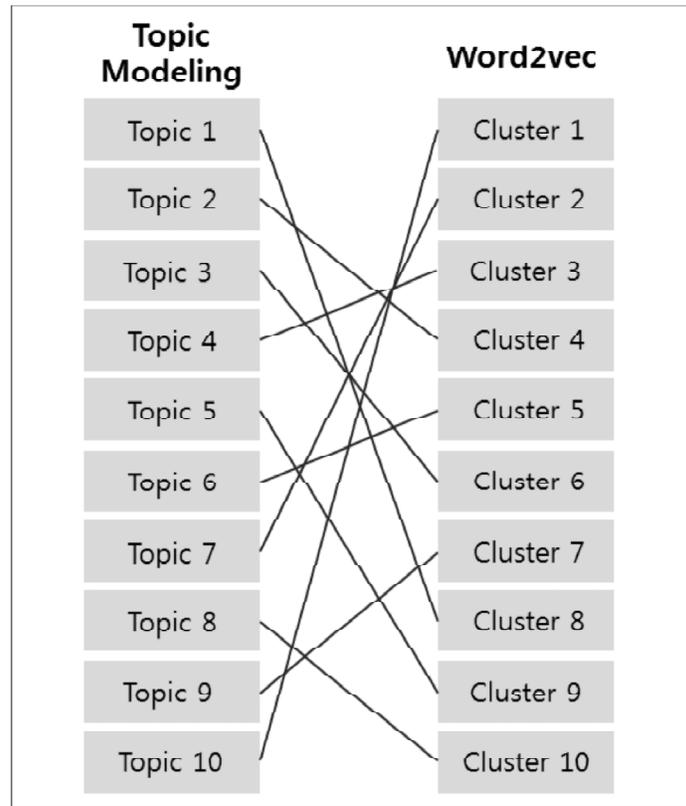


Figure 5: Comparison of Two Model

As shown in Figure 5, all topics founded in Topic Modeling are matched with the unique cluster founded in Word2Vec.

We can discover that the results of two techniques are similar and we can use basic characteristics of vector calculation for word analogue

Some analogue is shown in equation (2), (3) and (4)

$$\text{Counter Strike : Kill} = \text{GTA : } () \quad (2)$$

$$\text{Counter Strike : Target} = \text{GTA : } () \quad (3)$$

$$\text{Sport : Team} = \text{Counter Strike : } () \quad (4)$$

First, the relationship of Counter Strike and Kill was linked to ‘die, rob, chase’ in GTA.

The main attack of counter strike is to kill the opponent with a gun, and in GTA, the major attack is to usually chase and plunder the target.

In addition, the relationship of Counter Strike and Target was linked with ‘ride, trevor, station’ in GTA. It seems that the usual plundering method or the target was linked based on the attack nature of GTA.

Finally, the relationship of Sport and Team was linked to “Teamwork” in Counter Strike. This represents the importance of teamwork in FPS games just as much as in sports games.

6. CONCLUSION

In this research, we suggest a research framework for mining text document to understand users’ opinion using Topic Modeling Word2Vec. We also performed a case study of STEAM game review.

10 topics and 10 vector clusters are extracted by Topic Modeling and Word2Vec respectively. After comparing the clusters and topics of two models, 10 out of 11 games (Arma3, Dota2, Counter Strike, Counter Strike Global Offensive, The Elder Scroll, GTA, Warframe, SMC, Football Manager, Garry’s Mod) were extracted into 10 topics and clusters. Both models had Counter Strike and Counter Strike Global Offensive extracted into one topic and cluster. However, one cluster and topic were formed with words that generally encompass online games, and out of 11 games, only Team Fortress was unable to be extracted as one topic.

We also discover the similarity of mined results of two techniques and the possibility of complementary use of two techniques.

This research, however, has some limitations in examining the general relationship between Topic Modeling and Word2Vec mathematically and generalizing the results in other contexts.

REFERENCES

- Lee, J. H., and Lee, H. G. (2015), A Study on Customer Reviews about Domestic and Imported Clothes Products through Opinion Mining. *The Journal of Internet Electronic Commerce Research*, 15(3): 233-234.
- Hwang, H. S. (2016), Clustering Corporate Brands based on Opinion Mining: A Case Study of the Automobile Industry. *Journal of the Korea Academia-Industrial cooperation Society*, 17(11): 453-462.
- Lee, J. H., Lee, W. S., Park, J. W., and Choi, J. H. (2014), The Blog Polarity Classification Technique using Opinion Mining. *Journal of Digital Contents Society*, 15(4): 559-568.
- Kim, I. J., and Lim, J. Y. (2011), The way to measure trust ratio of text in Opinion Mining. In Proceedings of the Korean Information Science Society Conference. Korean Institute of Information Scientists and Engineers.
- Liu, B., and Zhang, L. (2012), A survey of opinion mining and sentiment analysis. Springer US, 415-463
- Chen, Y., and Xie, J. (2008), Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management science*, 54(3): 477-491.
- Kim, Y., and Jeong, S. R. (2013), Intelligent VOC Analyzing System Using Opinion Mining. *Journal of Intelligence and Information Systems*, 19(3): 113-125.
- Godes, D., and Mayzlin, D. (2004), Using online conversations to study word-of-mouth communication. *Marketing science*, 23(4): 545-560.
- Kim, Y. S., Kim, N. G., and Jeong, S. R. (2012), Stock-index invest model using news big data opinion mining. *Journal of Intelligence and Information Systems*, 18(2): 143-156.
- Kang, H. N., Yong, H. R., and Hwang, H. S. (2017), A Study of Factors Influencing Helpfulness of Game Reviews: Analyzing STEAM Game Review Data. *Journal of Korea Game Society*, 17(3): 33-44.
- Kwon, S. J. (2016), Sentiment Analysis of Movie Reviews using the Word2vec and RNN. Department of Computer Science and Engineering Graduate School of Dongguk University, 1-45.

- Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781.
- Seok, M., Song, H. J., Park, C. Y., Kim, J. D., and Kim, Y. S. (2015), Comparison of NER Performance Using Word Embedding. In The 4th International Conference on Artificial Intelligence and Application.
- O, D. S., Kang, S. W., and Seo, J. Y. (2016), An Iterative Approach to Graph-based Word Sense Disambiguation Using Word2Vec. *Journal of Cognitive Science*, 27(1): 43-60.
- Kim, J. Y., and Park, E. H. (2017), e-Learning Course Reviews Analysis based on Big Data Analytics. *Journal of the Korea Institute of Information and Communication Engineering*, 21(2): 423-428.
- Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003), Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan): 993-1022.
- Kim, D. Y. (2017), Design and Implementation of Patent Trend Analysis Tool based on LDA. Department of Computer Science and Engineering Graduate School of Soongsil University, 1-35.
- Blei, D. M. (2012), Probabilistic topic models. *Communications of the ACM*, 55(4): 77-84.
- Grimmer, J. (2010), An introduction to Bayesian inference via variational approximations. *Political Analysis*, 19(1): 32-47.
- Bae, J. H., Han, N. G., and Song, M. (2014), Twitter Issue Tracking System by Topic Modeling Techniques. *Journal of Intelligence and Information Systems*, 20(2): 109-122.
- Kang, B., Song, M., and Jho, W. (2013). A study on opinion mining of newspaper texts based on topic modeling. *Journal of the Korean Society for Library and Information Science*, 47(4): 315-334.
- Mimno, D., and McCallum, A. (2008), Topic models conditioned on arbitrary features with Dirichlet-multinomial regression, The 24th Conference on Uncertainty in Artificial Intelligence.
- Heo, C., and Ohn, S. Y. (2017), A Novel Method ofr Constructing Sentiment Dictionaries using Word2vec and Label Propagation. *Korean Institute of Next Generation Computing*, 13(2): 93-101.
- Lee, J. W., Yoo, K. H., and Mun, B. M., and Bae, S. J. (2017), Informal Quality Data Analysis via Sentimental analysis and Word2vec method. *J Korean Soc Qual Manag*, 45(1): 117-127.