Mining User Requirements from Online Reviews

分析線上評論以挖掘顧客需求

Ming Min Hsu
Information Management
National Taiwan University
r09725012@ntu.edu.tw

Yen Hsin Chen
Information Management
National Taiwan University
r09725001@ntu.edu.tw

Yu Ting Lin
Information Management
National Taiwan University
r09725004@ntu.edu.tw

ABSTRACT

Our study focused on analyzing customer satisfaction based on users reviews obtained from the Steam platform. We selected the popular game group from Top Sellers' Board and filtered the rest of games as the unpopular game group. After getting the two groups of data, we conducted two kinds of analysis. First of all, we tried to extract the patterns of two groups and compared them by using the tf-idf method. Second, we mined the user requirements by extracting the topics from users reviews of the two groups, and interpreted the result based on the concept of the Kano model. From the Kano model, we expected to analyze what factors are important to game developers in the development process of a new game, and provide companies some suggestions or directions to improve the sales of current games.

KEYWORDS

Information Retrieval, Topic Extraction, Reviews Analysis, Sentiment Analysis, Kano model

1 Introdiction

Based on the latest, August 2020, Global Video Gamer Segmentation report by games industry market research firm DFC Intelligence, the global gaming population has now surpassed three billion players. It indicates that about 40% of the global population are used to playing video games in a certain way or making related purchases. About 48% of the gamer population are PC gamers, and 8% are console consumers. However, this consumer group has the highest per user spending.

In the vast sea of video game titles, there's a need to distinguish which games are interesting. Just like any product review, it is clear that the main purpose of game reviews is to give valuable advice on whether a game is worth the player's time and money. This purpose may be a good thing for gamers, but for companies, the importance of game reviews can be questionable, as it is uncontrollable whether certain reviews give justice to the games they developed and affect their income in any way.

In the context of an online community, reviewer disclosure of identity-descriptive information is used by consumers to supplement or replace product information when making purchase decisions and evaluating the helpfulness of online reviews. For the game industry, there are a lot of aspects that may have contributed to the game's success, and reviews might be a contributing factor. Game reviews are usually composed of every concept of a game.

Reviewers cover the graphics, music, cinematics, story, content, controls, and gameplay along with their own opinions for each of those elements. Readers would rather see critical or controversial perspectives instead of reviews that are generic, and do not include any individual opinion. Such reviews may give the impression to players that the totality of the game was not well-analyzed. The quality of a review thus may lead to readers disregard the video game itself.

Reviews are made not to dictate customers what to play, but to help them see what they could find interesting and invest in. Game reviews are basically just opinions, but what influence do these opinions hold, especially for game companies and potential buyers? Therefore, we want to find out how game reviews actually affect the degree of customer satisfaction about video games.

2 Related Works

Ahn et al. [2] proposed Word2vec, PCA, K-means clustering, word cluster creation, and characterization, based on the Kano model for text mining to analyze user reviews on Steam. They firstly classify the games distributed that have led to the growth of the gaming industry in two groups – popular games and unpopular games – using the Bass diffusion model. Then they propose to use Word2vec to analyze online game reviews from Steam.

Eberhard et al. [4] apply a statistical hypothesis test and conduct a prediction experiment. They find that there are significant differences between helpful and unhelpful reviews. e.g., review length and time spent playing a game strongly. These reviews tend to be more critical towards the product and go into greater detail about the individual aspects. Despite that, they also find a number of reviews that differ from these patterns. First, they discover reviews with a large number of helpfulness votes but with a short or meaningless text, where the number of votes is derived from humor, an outside source (e.g., a video on an external site), or the author being a popular personality. Second, they find reviews of very similar style to the most helpful ones but without any votes, potentially because of bad timing.

Lin et al. [6] performed an empirical study on the reviews of 6,224 games on the Steam platform. They studied the number and the complexity of reviews, the type of information that is provided in the reviews, and the number of playing hours before posting a review. In addition, they study the relation between several game-specific characteristics and the fluctuations of the number of reviews that are received on a daily basis. They then focus on the useful information that can be acquired from reviews by studying the major concerns that users express in their reviews, and the amount of play time before players post a review. They find that the number of playing hours before posting a review is a unique and helpful attribute for developers that is not found in mobile app reviews.

Song et al. [9] combine the data mining technology with the Kano model. They analyze the sentiment of the comment through machine learning to acquire the parameters of the Kano model such as the initial importance, then obtain the real demand of multiple customers as well as the weight of demand. They focus on the discovery of different types of users' demand for products, and using the transformation function analysis to obtain the weight of various customer needs. At the same time in mining the needs of customers, they also explore the differences between the needs.

Zhang et al. [11] proposed an aspect sentiment collaborative filtering algorithm (ASCF), which combines sentiment analysis with a fuzzy Kano model. It obtains the users' different attitudes toward aspects of the product by fine-grained sentiment analysis from the user's purchase records, and then analyzes the user's degree of desire and importance for each feature based on the fuzzy Kano model, proposing a novel similarity measure method with user preferences for a collaborative filtering algorithm.

Min et al. [7] categorizes customer reviews into two types — positive reviews and supplementation-required reviews — and suggests a five-section framework according to the frequency of each review type. They define characteristics of each section from the perspective of the Kano model. Based on this framework, they analyze the dynamics of customer requirements in the online businesses, for which customer reviews are the main indicator of service quality.

3 Proposed Method

This study consists of two analysis parts. We selected the popular game group from Top Sellers' Board and filtered the rest of games as the unpopular game group. First of all, we tried to extract the patterns of two groups and compared them. Second, we mined the user requirements by extracting the topics from users reviews of the two groups, and interpreted the result based on the concept of the Kano model.

3.1 **Patterns Extraction**

3.1.1 TF-IDF

In this part, we used the tf-idf method trying to extract some obvious patterns that are different between two groups that are popular games on top sellers board and unpopular games among tens of thousands of PC games on Steam. To utilize this method, we have to preprocess the data so that we can consume the analysis.

A. Tokenization, Stemming and Lemmatization
First of all, we pick all the reviews of popular and unpopular games that are selected equally on scale. Secondly, we use the casual_tokenize function provided by nltk to split review content into separate words. Then, we sequentially conduct Port stemming method and lemmatization on the tokens of reviews.

B. Stopwords Removing and Repeating Characters Removing
This step is really important since many players can post so many common vocabularies or meaningless words with repeating characters such as "Ohhhhh" or "Rrrrrrrr". As a consequence, apart from removing stopwords, we also handle the situation mentioned above by removing repeating characters more than 3 times within a word.

C. TF-IDF Vectorizer and Ranking
As for the last step, we use the tf-idf vectorizer originated from scikit-learn that helps us convert text into vectors. Then, we summarize the importance of all the terms within reviews of popular and unpopular games. We select some representative terms that seem to be patterns of its group. The result will be shown in the next part describing our overall results.

3.2 **User Requirements Mining**

3.2.1 Topic Extraction

LDA model is a method of modeling documents, topics and words, which transforms the traditional word vector expression into the topic vector expression . The benefit of doing this is obvious that the expression based on topics reduces the dimension of feature space. The LDA model has a clear logical structure, containing the document layer,the topic layer and the word layer. Each layer is adjusted by variables and parameters.

We used the semi-supervised learning method based on the LDA theme model to extract the features. A document can contain multiple topics, each of which is generated by one of the topics in the document. The LDA can give the subject of each document in the document set as a probability distribution.

3.2.2 Kano Model Interpretation

We used the Kano model to get further understanding of the characteristics of the popular and unpopular gaming products in the topics we extracted. According to the study by a research group led by Kent Thorén at the Royal Institute of Technology in Sweden, the Kano model defines the key success factor by associating the three main types of factors, M, O, and A, with the gaming industry. [2] The mandatory requirement, M, represents the Basic (Must-be) Attributes of the gaming product, and, if associated with the gaming product, can be categorized as technical functionality, motivator, and challenge. O, which is a proportional requirement, represents the Performance Attributes, an indicator of the product's performance level. Examples of performance attributes are price, audiovisual aesthetics, story and/or narrative, and gameplay. The attractive requirements, A, represent the Excitement Attributes and are properties. The words which describe three main types of factors, M, O, and A, with the gaming industry is shown in Table 1.

Table 1. Words Describing Three Main Types of Factors

M	**Technical functionality**	bug, bugs, problem, problems, needs, want, wants, conflict, conflicts
	Motivator	motivator, achieve, compete
	Challenge	challenge, challenges, difficult, difficulty, easy
O	**Price**	price
	Audiovisual aesthetics	audio, music, sound, graphic, visual
	Story and/or narrative	story, storyline, campaign
	Gameplay	gameplay, gameplays
A	**Emotional connection**	positive, negative, emotional, connection
	Uniqueness	unique, uniqueness, original

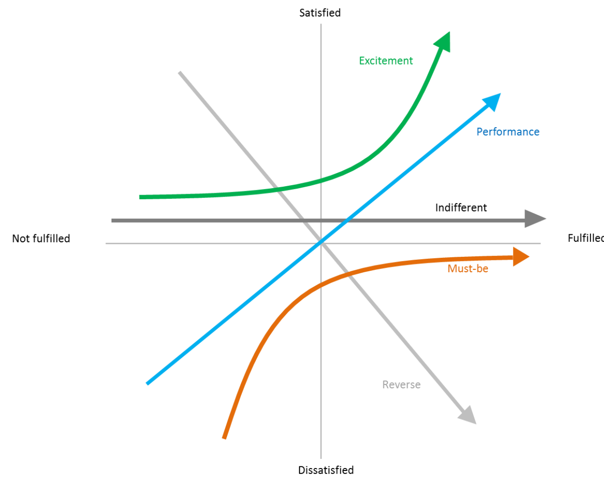

Figure 1. Kano model

3.2.2.1 Similarity Comparison

Similarity is determined by comparing word vectors or "word embeddings", multi-dimensional meaning representations of a word. Word vectors can be generated using an algorithm like Word2Vec which is a more recent model that embeds words in a lower-dimensional vector space using a shallow neural network. The result is a set of word-vectors where vectors close together in vector space have similar meanings based on context, and word-vectors distant to each other have differing meanings. We use Word2vec, the closer the distance between the vectors is, the higher the correlation between them is. To find the differences between the two groups, we calculate the correlation between two chosen words of each factor of the Kano model for each group.

3.2.2.2 Sentiment Analysis

Sentiment analysis refers to the extraction of opinions, topic analysis, and sentiment mining on text information such as news, and products reviews. The related methods of textual emotion analysis are mainly divided into two categories, one is the sentiment classification method based on the sentiment dictionary, the other is the sentiment classification method based on machine learning. We used the sentiment dictionary launched by HowNet [12], which includes sentiment texts of evaluation, sentiment, proposition, and degree, to match the emotion words of the text, aggregate the emotion words and score, and finally get the emotional tendency of the text. Both the evaluation and sentiment vocabulary have positive and negative sentiment dictionaries, we merged the evaluation vocabulary and the sentiment vocabulary into emotional words. And, the degree words contained in the degree vocabulary are classified according to grades:most, very, more, slightly, insufficiently, over.

A. Corpus Construction

Get the text corpus after the sentence segmentation, word segmentation, part-of-speech tagging identification, and stopwords removal. Then, extract the positive emotional words, negative emotional words, and the degree words to further sentiment score calculation.

B. Sentiment Score Calculation

According to the positive and negative emotional words list, the emotional score of each word in a review is given: 1 for positive emotion, 0 for neutral emotion and -1 for negative emotion. And it is also necessary to determine whether there is a degree word, or called an adverb, in front of the emotion word. If it exists, it is necessary to give different weights, which is shown in Table 2, according to the type of degree adverb and multiply it by the number of emotion words. After all, statistically calculate the sentiment score of the whole review to get the sentiment tendency of the review.

Table 2. The Weight of 6 Kinds of Degree Words

degree words	most	very	more	slightly	insufficiently	over
weight	2	1.75	1.5	1.25	0.5	-1

C. Sentiment Extraction of Three Main Types of Factors

To find the sentiment difference between the two groups, we calculated the average sentiment score of the reviews containing two chosen words in each factor of the Kano model for each group.

4 Result and Evaluation

4.1 Data

Since our inspection is centered on PC games, we choose Steam platform, which is the largest digital distribution platform for PC gaming, as our data source. Steam is a digital game distribution platform, developed by Valve Corporation. Steam is considered to be one of the largest digital distribution platforms for PC gaming, with over 8,000 games available and over 184 million active users. Steam offers digital rights management (DRM), multiplayer gaming, and social networking services, through two major components of the Steam platform: the Steam Store, and the Steam Community.

4.1.1 Web Crawling

Steam provides several APIs for the public, including the whole games IDs on it and the reviews of each game. However, the API provided by Steam has some limits. For example, Steam limits the API by providing reviews of each game no more than 100. This could potentially be a problem since the data quantity seems far less than the quantity we need. Hence, we develop our own dataset by crawling the Steam website to download games' information and reviews. Table 3 and Table 4 shows the features of games' information and reviews.

Table 3. Data Features of Games' Information

Games' Information	
Feature	**Description**
game_id	the ID of the game
game_name	the name of the game
release_date	the release date of the game
price	free (True) or not (False)
price_initial	the pricing of product
discount_perc	the discount percentage
label	the tags of the games
description	the description of the game

Table 4. Data Features of Games' Review

Reviews	
Feature	**Description**
game_id	the ID of the game
useful	number of people that found the review to be useful
funny	number of people that found the review to be funny
username	username of the reviewer
num_owned_games	number of games owned by the reviewer
num_written_reviews	number of reviews written by the reviewer
recommended	1=recommended, -1=not recommended
hours_played	hours played by the reviewer on the game
post_date	date of creation of the review
content	text of the review

4.1.2 Exploratory Data Analysis

There are 61,294 games with complete information, and 56.3% (34,503) of them are needed to be paid. Among the paid games, there are 8% (2,821) of games on the Top Sellers board, we dropped out those games without reviews and categorized these games as the popular game group. The rest of games are sorted by the number of reviews in descending order, and the top of 2,639 games are seen as the unpopular game group.

We crawled over 10 millions reviews from the Steam website, the popular game group contains 4,702,951 reviews while the unpopular game group contains 3,833,522 reviews. For further model construction, we set rules

- number of people that found the review to be useful ≧ 10
- hours played by the reviewer on the game ≧ 5

to filter the reviews data. After filtering by the rules, the number of the reviews in both groups are approximately equal to 106,500.

4.2 **Patterns Extraction**

4.2.1 Similar Terms Removing

Due to the incomplete property of the stop-words file, the result of important terms of popular group and unpopular group are somehow very similar. Take their top-10 results for example.
Both the top-1 terms of two groups are "game", and the top-2 terms of that are "play". This indicates that they are difficult to distinguish to some extent. Therefore, we decide to pick their representative terms by comparing the terms at the same rank. If both of terms at same rank are equal, then they won't be selected to the waiting list. For the final step, we manually pick seven terms of popular group and unpopular group respectively to represent the patterns of their belonging group.

4.2.2 Representative Terms

According to Table 5, we could find some clues regarding the patterns of these two groups. For the popular group, "friend", "multiplay" and "online" perhaps mean that players can make some friends via the game and play it together. "DLC" and "update" may refer to the meaning that the game developer releases an updated version to improve the playing experience or DLC (no matter if it needs additional pay or not) to let players experience more contents of the game.

As for the unpopular group, "bored", "simulate" and "short" can somehow explain that the game content is few and boring. Furthermore, maybe it simulates another game's content -- which is the last thing players would love to see. From the terms "bug" and "crash", we can guess that the game is full of bugs and it is easy to crash during playing. That is, the playing experience of the game is obviously bad and players don't like this phenomenon without doubt.

Table 5. Patterns of the Two Groups

Popular games	Unpopular games
friend	bored
DLC	bug
update	puzzle
survive	old
multiplay	crash
beautiful	simulate
online	short

4.3 **User Requirements Mining**

4.3.1 Topic Extraction

The area of the circle represents the importance of each topic over the entire corpus, the distance between the center of circles indicates the similarity between topics, shown in Figure 2 and Figure 3. There is some overlapping between topics, but generally, the LDA topic model can help us grasp the key topics of the group.

In LDA models, each review is composed of multiple topics. However, typically only one of the topics is dominant. We extract the dominant topic for each review and calculate the weight of the topic and the keywords. This way, we can know which review belongs predominantly to which topic. The LDA model is built with 8 different topics where each topic is a combination of keywords and each keyword contributes a certain weightage to the topic. We extract 20 keywords of each topic which shows in Figure 4 and Figure 5.

Figure 2. Topic Clustering Of Popular Game Group

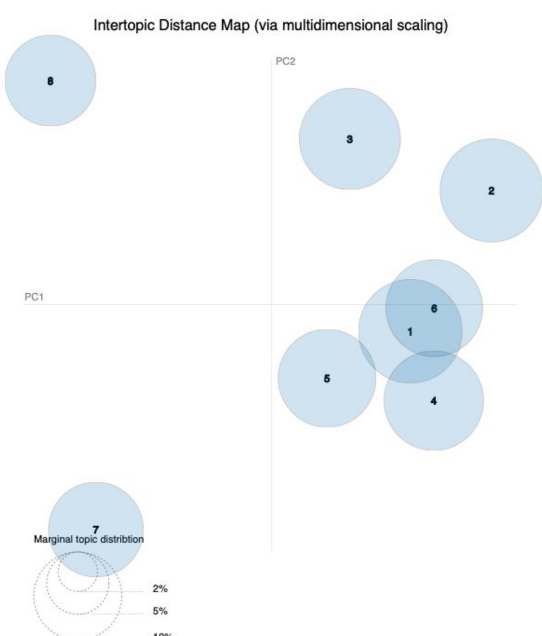

Figure 3. Topic Clustering Of Unpopular Game Group

Figure 4. Popular Game Group Topic Keywords

Figure 5. Unpopular Game Group Topic Keywords

After topic clustering, the results of popular and unpopular game groups both show that some of the clusters contain the words describing three main types of factors of the Kano model. This section is similar in the case of both the groups and can be explained as a result of the unpopular game group. It can be seen that three major types of factors are distributed in topics 2, 3, 4, 5, 6, and 8. The attributes of types M, O, and A are not grouped together in one topic, but are slightly mixed (M: topic 2, 3; O: topic 2, 4, 6, 8; A: topic 5, 8). However, three major types of factors are not distributed in topics 1 and 7, which could be interpreted as indifferent (I) or reverse (R) of the Kano model.

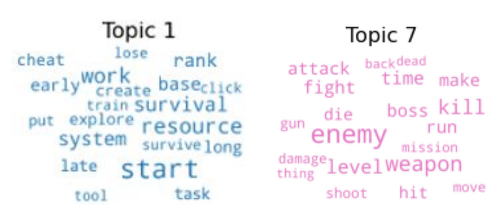

Figure 6. Topics without Three Major Types of Factors in Unpopular Game Group

Figure 6 shows the topics without three major types of factors in the unpopular game group. In topic 1, we could see that the words such as “task”, “rank”, “lose”, and “cheat”, which could be interpreted as the factors influencing game results and result in dissatisfaction with the game among the users. Therefore, we viewed this cluster as having the elements of type R (reverse) satisfaction. In topic 7, the characters such as “enemy” and “boss”, and the game arms items and actions such as “weapon”, “gun”, “attack”, “fight”, “shoot”, and “kill” were dominant. These factors were found in other games with the same labels, so it could be seen as type I (indifferent) satisfaction, which is not related to the satisfaction of the game.

The topic extraction result shows that the reviews on Steam platform contains all five types of product development and customer satisfaction discussed in the Kano model. Therefore, we interpreted the topics obtained from Steam user reviews and the difference between the popular and the unpopular game groups based on the concept of the Kano model.

4.3.2 Kano Model Interpretation

To find the differences between the popular and the unpopular game group, we calculate the correlation and average sentiment score between two chosen words of each factor of the Kano model for each group. The similarity difference and the average sentiment score is shown in Table 6.

Table 6. Similarity Difference and Sentiment Scores between Two Group

Kano Model Factor		Similarity Difference		Sentiment Scores	
		Game Group		Game Group	
		Popular	Unpopular	Popular	Unpopular
M	Technical functionality	0.5382	0.4765	75.9180	63.4963
O	Price	0.3176	0.2498	101.0785	63.7210
	Audiovisual aesthetics	0.4991	0.4864	58.2281	17.3032
A	Uniqueness	0.6840	0.3856	78.8603	71.2361

Firstly, the similarity between “bug” and “fixed” associated with the “Technical functionality” factor, which are 0.5382 and 0.4765 for the popular and unpopular game groups, respectively, shows that more technical issues are being fixed in the popular group. And the average sentiment score, which are 75.9180 and 63.4962 for the popular and unpopular game groups, respectively, also supports that the sentiment of the “Technical functionality” factor is more positive in the popular group.

Secondly, the association between “price” and “reasonable” of the “Price” factor, which are 0.3176 and 0.2498 for the popular and unpopular game group, respectively, implies that the games of the popular group meet the consumer expectations in terms of pricing better than the games of the unpopular group. And the average sentiment score, which are 101.0785 and 63.7210 for the popular and unpopular game groups, respectively, also supports that the sentiment of the “Price” factor is more positive in the popular group.

Thirdly, the distances between “graphic” and “smooth” of the “Audiovisual aesthetics” factor, which are 0.4991 and 0.4864, for the popular and unpopular game group, respectively, implies that the graphic is smoother in the case of popular games. And the average sentiment score, which are 58.2281 and 17.3032 for the popular and unpopular game groups, respectively, also supports that the sentiment of the “Audiovisual aesthetics” factor is much more positive in the popular group.

Finally, the distances between “uniqueness” and “gamestyle” associated with the “Unique” factor, which are 0.6840 and 0.3856 for the popular and unpopular game group, respectively, shows that the unpopular games are required to have more original gamestyle elements. And the average sentiment score, which are 78.8603 and 71.2361 for the popular and unpopular game groups, respectively, also supports that the sentiment of the “Unique” factor is a little more positive in the popular group.

5 Conclusion

Our study focused on analyzing customer satisfaction based on users reviews obtained from the Steam platform. We selected the popular game group from Top Sellers’ Board and filtered the rest of games as the unpopular game group. After getting the two groups of data, we conducted two kinds of analysis. First of all, we tried to extract the patterns of two groups and compared them by using the tf-idf method. Second, we mined the user requirements by extracting the topics from users reviews of the two groups, and interpreted the result based on the concept of the Kano model. From the Kano model, we analyzed what the game developers or companies cared about and they could do some improvement from their games, thus gaining more customer satisfaction.

In our observation, reviews are important and informative to players, however, the “real” useful and helpful ones are considerably little. Players usually choose to read a long review that detailedly describe the game and show the pros and cons of the game objectively. This kind of review shows comprehensive patterns of the game, but the amount of this kind of review is too small. Other reviews usually contain little contents such as “good game” and “nice to play”, they did not show the details or patterns of the game. As a consequence, when we account for all the reviews to analyze our issue, we can hardly know what are the true differences between popular games and unpopular games. That is, the reviews with short content might be a kind of noise to our models.

Our study is expected to have managerial implications for the gaming industry like understanding the factors that are important in the development of a new game or providing the suggestions or directions to improve the sales of current games. However, there are also some shortcomings of our research such as not all reviews are informative, the label of different games should be considered in the part of user requirements mining and adjust the constraints to filter the reviews. There is still much room for improving the models and the filtering rules, and we could design an adjusted model to improve these shortcomings in our future work.

6 Member's Workload

Member	Workload
Ming Min Hsu	Data Crawling, User Requirement Mining, Presentation
Yen Hsin Chen	Patterns Extraction, Report Summarization
Yu Ting Lin	User Requirement Mining, Report Summarization

7 References

[1] A. (n.d.). aesuli/steam-crawler. GitHub. Retrieved December 25, 2020, from https://github.com/aesuli/steam-crawler

[2] Ahn, Sangho, Juyoung Kang, and Sangun Park. "What makes the difference between popular games and unpopular games? analysis of online game reviews from steam platform using word2vec and bass model." ICIC Express Letters 11.12 (2017): 1729-1737.

[3] Che, Wanxiang, et al. "N-LTP: A Open-source Neural Chinese Language Technology Platform with Pretrained Models." arXiv preprint arXiv:2009.11616 (2020).

[4] Eberhard, Lukas, et al. "Investigating helpfulness of video game reviews on the steam platform." 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS). IEEE, 2018.

[5] Ganesan, K. (2020, July 31). Gensim Word2Vec Tutorial – Full Working Example. Kavita Ganesan, Ph.D. Retrieved January 10, 2021, from https://kavita-ganesan.com/gensim-word2vec-tutorial-starter-code/#.X_pmN2QzaWA

[6] Lin, Dayi, et al. "An empirical study of game reviews on the Steam platform." Empirical Software Engineering 24.1 (2019): 170-207.

[7] Min, H.; Yun, J.; Geum, Y. "Analyzing Dynamic Change in Customer Requirements: An Approach Using Review-Based Kano Analysis." Sustainability 2018, 10, 746.

[8] Prabhakaran, S. (2020, September 17). Topic modeling visualization - How to present results of LDA model? ML+. Retrieved January 10, 2021, from https://www.machinelearningplus.com/nlp/topic-modeling-visualization-how-to-present-results-lda-models/

[9] Song, Huaming, Chao Chen, and Qin Yu. "Research on Kano models based on online comment data mining." 2018 IEEE 3rd International Conference on Big Data Analysis (ICBDA). IEEE, 2018.

[10] Summarization of 2020 DFC Intelligence's report https://www.podcastgameconsultant.com/reports-research/dfc-intelligence-more-than-40-of-world-population-now-play-video-games/

[11] Zhang, Jinming, Dexin Chen, and Min Lu. "Combining sentiment analysis with a Fuzzy Kano model for product aspect preference recommendation." IEEE Access 6 (2018): 59163-59172.

[12] Zhendong Dong and Qiang Dong, "HowNet - a hybrid language and knowledge resource," International Conference on Natural Language Processing and Knowledge Engineering, 2003. Proceedings. 2003, Beijing, China, 2003, pp. 820-824, doi: 10.1109/NLPKE.2003.1276017.