Asia Pacific Journal of Information Systems
Vol. 29 No. 4 (December 2019), 856-874

ISSN 2288-5404 (Print) / ISSN 2288-6818 (Online)
https://doi.org/10.14329/apjis.2019.29.4.856

Special Issue: Data Analytics in Artificial Intelligence Era

The Effects of Content and Distribution of Recommended Items on User Satisfaction: Focus on YouTube

Janghun Jeong[a], Kwonsang Sohn[b], Ohbyung Kwon[c,*]

[a] *M.S. Student, School of Management, Kyung Hee University, Korea*
[b] *Ph.D. Candidate, School of Management, Kyung Hee University, Korea*
[c] *Professor, School of Management, Kyung Hee University, Korea*

A B S T R A C T

The performance of recommender systems (RS) has been measured mainly in terms of accuracy. However, there are other aspects of performance that are difficult to understand in terms of accuracy, such as coverage, serendipity, and satisfaction with recommended results. Moreover, particularly with RSs that suggest multiple items at a time, such as YouTube, user satisfaction with recommended results may vary not only depending on their accuracy, but also on their configuration, content, and design displayed to the user. This is true when classifying an RS as a single RS with one recommended result and as a multiple RS with diverse results. No empirical analysis has been conducted on the influence of the content and distribution of recommendation items on user satisfaction. In this study, we propose a research model representing the content and distribution of recommended items and how they affect user satisfaction with the RS. We focus on RSs that recommend multiple items. We performed an empirical analysis involving 149 YouTube users. The results suggest that user satisfaction with recommended results is significantly affected according to the HHI (Herfindahl-Hirschman Index). In addition, satisfaction significantly increased when the recommended item on the top of the list was the same category in terms of content that users were currently watching. Particularly when the purpose of using RS is hedonic, not utilitarian, the results showed greater satisfaction when the number of views of the recommended items was evenly distributed. However, other characteristics of selected content, such as view count and playback time, had relatively less impact on satisfaction with recommended items. To the best of our knowledge, this study is the first to show that the category concentration of items impacts user satisfaction on websites recommending diverse items in different categories using a content-based filtering system, such as YouTube. In addition, our use of the HHI index, which has been extensively used in economics research, to show the distributional characteristics of recommended items, is also unique. The HHI for categories of recommended items was useful in explaining user satisfaction.

Keywords: Recommender Systems, User Satisfaction, Herfindahl-Hirschman Index, Social Media

*Corresponding Author. E-mail: obkwon@khu.ac.kr Tel: 8229612148

Ⅰ. Introduction

Recommender systems (RS) can be defined as programs that attempt to recommend the most suitable items (products or services) to particular users (individuals or businesses) by predicting their interests based on related information about previously viewed items, their users, and the interactions between items and users (Lu et al., 2015). The aims of developing RSs are to reduce information overload and to facilitate retrieval of the most relevant information and services from a huge amount of data, thereby providing personalized services (Lu et al., 2015). RSs can provide content-based recommendations and collaborative filtering. More precise, personalized recommendations are possible when users give information, creating a personal profile or evaluating recommendation results (Moghavvemi et al., 2017). YouTube is a typical example of a platform on which RSs are utilized.

The performance of RSs has been measured in terms of recommendation accuracy using the RMSE (root mean squared error) or MAE (mean absolute error). Correspondingly, recommendation algorithms have been optimized for accuracy (Kaminskas and Bridge, 2014), which indicates their predictive power. However, recently the focus of the RS community has begun to shift towards factors other than accuracy (Kaminskas and Bridge, 2014), as accuracy alone does not always result in user satisfaction (Kotkov et al., 2018). In this regard, user satisfaction on the recommendation results is getting interested.

However, user satisfaction with RSs is not uniform; there are many dimensions such as information content, customized service, user interface, and system value in terms of utility or efficiency (Liang et al., 2006). Indirect surveys have been used to measure the concept of satisfaction (Ogara et al., 2014). However, controversy exists over the appropriate method by which to evaluate satisfaction (Gatian, 1994).

Several related studies have been conducted to identify factors affecting user satisfaction with recommendations, the results of which have important implications for improvement of the performance and design of RS. For example, personal characteristics such as social presence have been examined, and subjective assessments have been provided of parameters like trust (Choi et al., 2011), diversity, and serendipity of recommendations (Nguyen et al., 2108). However, recommendation accuracy has remained the main focus of study (Liang et al., 2006). Even though a wider range of factors should be considered such as content and distribution of recommended items, and system architecture, not just accuracy, very few empirical studies take these characteristics into account extensively: Most studies involve simulations (Shani and Gunawardana, 2011), which cannot demonstrate how the content or distribution of recommended items affect user satisfaction (Kotkov et al., 2018). Finally, user satisfaction with recommended outcomes does not necessarily coincide with algorithmic accuracy.

Hence, the purpose of this study is to empirically examine the effects of the content and distribution of recommendation items in a multi-RS setting. The main contribution of this study is that this study successfully presents a sophisticated model to understand how user satisfaction is formed for multi-RS. The results suggest that both the characteristics of the content and distribution of recommendation items influence user satisfaction with multi-RS. Second, to the best of our knowledge, this study is the first to investigate the effect of content characteristics of recommendation items of multi-RS on satisfaction. Third, we firstly invited Herfindahl-

Hirschman Index (HHI) as a compound measure of consistency and serendipity. The results show that HHI is a statistically significant indicator of satisfaction with multiRSs. Finally, this study firstly showed that content and distribution characteristics affecting satisfaction are different depending on the purpose of using multi-RS, say pleasure purpose and utility purpose.

Ⅱ. Theoretical Foundations

2.1. Definitions

RSs analyze data about a user's propensity and preferences, identifying desirable content. RSs are typically applied in a variety of areas with large services such as YouTube, Netflix, and Amazon, and have been studied since the 1990s (Hong and Kim, 2016; Zeng et al., 2015; Zhao et al., 2016). RSs utilize either content-based filtering, collaborative filtering, or a combination of the two. Content-based filtering is a technique in which information about the age, sex, occupation, and preferences of users is analyzed and content is recommended (Yoo, 2016). Collaborative filtering is a method by which users' historical characteristics are analyzed and grouped, after which content is recommended that is enjoyed by groups with similar preferences to target users (Kumar et al., 2015). Hybrid RS is a combination method involving more than one filtering method, thereby overcoming the limitations of the other two methods, such as cold start, inaccuracy, and so on (Thorat et al., 2015). For example, the YouTube RS, which combines content-based filtering and collaborative filtering, helps users find specific content using deep neural networks; this is a successful case of hybrid filtering (Abbas et al., 2017).

2.2. Performance Measures for RSs

RS performance has traditionally been evaluated in terms of accuracy. To determine overall accuracy, MAE (Anand and Bharadwaj, 2011), RMSE (Anand and Bharadwaj, 2011), NRDM (normalized distance-based performance measure; Yao, 1995), and ROC (receiver operating characteristic) (Herlocker et al., 2004) are suitable metrics. RS evaluation metrics can be used to evaluate the accuracy of prediction of continuous data or the classification of categorical data according to the type of data, or to evaluate diversity, recommendation accuracy, etc. according to the purpose of evaluation. However, predictive accuracy metrics treat all measurements the same, so it is difficult to know the accuracy (Lu et al., 2012) and classification accuracy metrics have the disadvantage that the system may not recognize the correct item (Herlocker et al., 2004). In other words, there is no single measurement item that can measure the performance of the evaluation metrics efficiently (Jalili et al., 2018), so there is a need to consider a variety of factors. In particular, because Youtube, the subject of this study, contains various factors that affect the individual characteristics of the user or the user (Qin et al., 2010), We used diversity and serendipity to see if Youtube's various factors and user characteristics influence satisfaction (Kotkov et al., 2016; Kunaver et al., 2017). In addition, scholars have begun to show interest in how diverse and appropriate recommendations are made based on other parameters. Examples of this include coverage, diversity, and serendipity. Coverage and serendipity are crucial evaluation metrics for evaluation of RS performance. While coverage concerns the degree to which recommendations cover the set of available items and the extent to which recommendations can reach all potential users, diversity focuses on how different the recommdned items are with respect to each other (Vargas and Castells, 2011). Finally, serendipity is concerned with the novelty of recommendations and the extent to which recommendations may positively surprise users (Ge et al., 2010).

First, coverage reflects the degree to which the generated recommendations cover the catalog of available items (Adomavicius and Kwon, 2011; Ge et al., 2010; Herlocker et al., 2004; Kaminskas and Bridge, 2017). Higher coverage may benefit both system users and business owners, and exposing the users to a wider range of recommended items may increase both their satisfaction with the system (Adomavicius and Kwon, 2011) and overall product sales (Anderson, 2006). RSs with high coverage provide a more detailed and careful investigation of the product space; therefore, high coverage is an indicator of quality.

Second, diversity refers to the system's ability to recommend different items to different users, or to the relevant portion of the item catalog recommended across all users (Kaminskas and Bridge, 2017). Ziegler et al. (2005) observed that a more diverse recommendation list can lead to higher user satisfaction, despite its lower accuracy. Diversity has frequently been measured as the average or aggregate dissimilarity of items in the recommendation list. For example, Smyth and McClave (2001) suggested measuring the diversity of a recommendation list R (|R| > 1) as the average pairwise distance between items in the list:

$$\mathrm{Diversity(R)} = \frac{\sum_{i\in R}\sum_{j\in R\setminus\{i\}} dist(i,j)}{|R|(|R|-1)}$$

Due to the nature of these metrics, diversity often has a trade-off relationship with accuracy.

Third, serendipity is the deviation from the "natural" prediction (Murakami et al., 2007). Hence, serendipity is frequently associated with novelty or unexpectedness (Kotkov et al., 2016). Novelty is indicated by lack of ratings by target users regardless of familiarity (Shani and Gunawardana, 2011). Serendipity is defined as follows (Sridharan, 2014, p. 2):

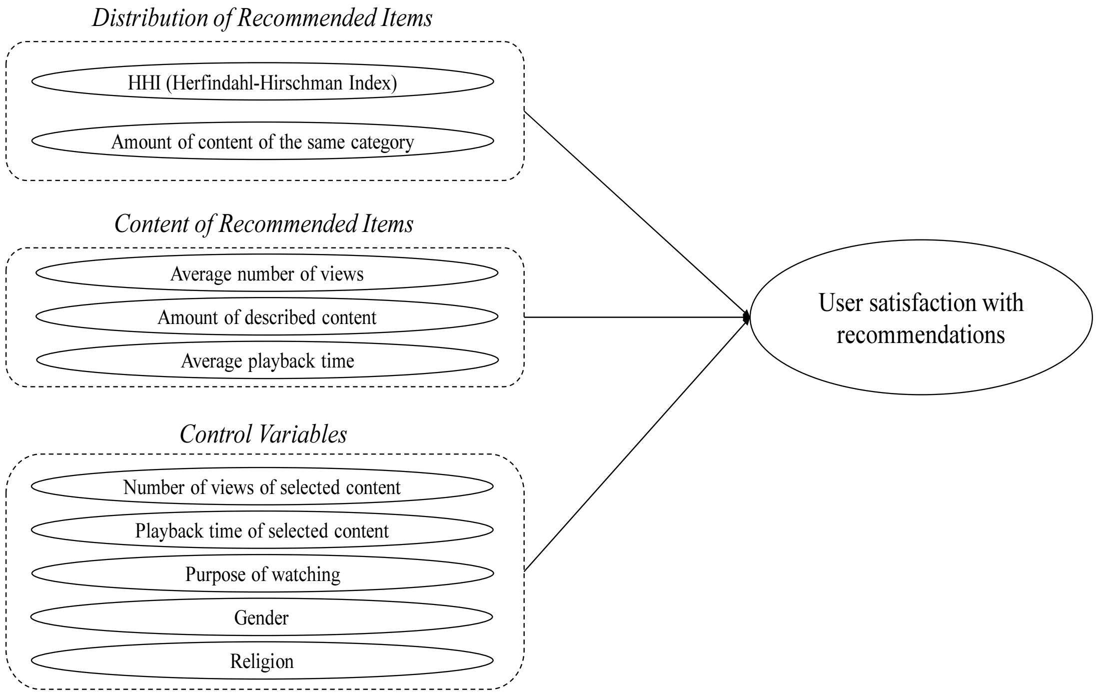
> … the accident of finding something good or useful while not specifically searching for it. Serendipity is thus closely related to unexpectedness and involves a positive emotional response of the user about a previously unknown item. It measures how surprising the unexpected recommendations are (Shani and Gunawardana, 2011) serendipity is concerned with the novelty of recommendations and in how far recommendations may positively surprise users.

A serendipitous recommendation helps the user to find a surprisingly interesting item that she might not have otherwise discovered (Iaquinta et al., 2010). However, serendipity is challenging to investigate and hard to measure in a simulation because it includes an emotional dimension (Kotkov et al., 2016). Though several studies have suggested metrics of serendipity (Adamopoulos and Tuzhilin, 2015; Ge et al., 2010; Kaminskas and Bridge, 2014; Murakami et al., 2007; Vargas and Castells, 2011), they measure unexpectedness separately from relevance, which might result in error.

An item can be considered serendipitous if a classifier is uncertain about its relevance for the user (Iaquinta et al., 2008), if the item is different from those found in the user's profile (Adamopoulos and Tuzhilin, 2015), or if the item is associated with a distinct area in a user-item graph (Nakatsuji et al., 2010; Onuma et al., 2009; Zhang et al., 2012). In this study, recommended items not belonging to the same category of content initially selected by the user were considered serendipitous because YouTube utilizes content-based filtering, not collaborative filtering. In this case, we examine how many items of varying content were recommended, assuming that user satisfaction with recommendations may differ if multiple items are recommended in one category or few items are recommended in each of the different categories. Thus, we focus on aspects of serendipity that have not been considered in other studies.

2.3. User Satisfaction

In addition to studies on the performance of RSs, research has also been conducted on factors affecting user satisfaction with them. First, personal characteristics such as personality, taste, preference, and interests can affect satisfaction (Ferwerda et al., 2106). The characteristics of recommended items may also affect satisfaction. For example, the higher the number of views of content and the shorter the playback time, the more positive the user's experience (Meseguer-Martinez et al., 2017), the more time spent interacting with the content, and the greater the satisfaction (Mehrotra et al., 2018). The more diversified the content provided, the greater the user's satisfaction with specific content (Ferwerda et al., 2106). However, there have been few empirical studies on how the distribution of recommended items. In this study, we examine this phenomenon, including the relevant factors of coverage, diversity, and serendipity in our model, and investigating how they affect user satisfaction with recommendations.



<Figure 1> Research Model

Ⅲ. Research Model

For RSs that recommend several items, we propose the research model as shown in <Figure 1>, which illustrates factors affecting user satisfaction with recommendations. Recommendation characteristics are divided into three groups: content characteristics, distribution characteristics, and characteristics of first selected content.

3.1. User Satisfaction with Recommendations

A RS may be considered successful when it delivers accurate recommendations, or user satisfaction may be considered as the main indicator of success. Most studies consider accuracy as a key performance factor indicating the success of RSs. However, the ultimate goal of a RS is to satisfy its users and alleviate the information overload problem, because user satisfaction is the predecessor of other goals such as increased sales. Although accurate recommendations are important, other aspects also determine user satisfaction. For example, an RS providing recommendations of TV programs would be very accurate in recommending 10 episodes of *Friends* when a user is known to like this television series. It is unlikely, however, that the user will be pleased with this list, however accurate it may be. In addition to accuracy, satisfaction with the recommendation may be affected by topic diversification, coverage, or serendipity (Ziegler et al., 2005). Since the RS considered in this study provides multiple recommendations for platforms like YouTube, user satisfaction with the recommendations is measured by examining the top *n* recommended items. We expect that the recommended content, especially that listed first, will have a greater impact on the level of user satisfaction (Bobadilla et al., 2013). In fact, accuracy is merely a property of a single recommendation (McNee et al., 2006), while diversity is a property of recommendation lists (Ziegler et al., 2005). Thus, we do not include accuracy as a dependent variable in our analysis, but rather directly measure user satisfaction based on the recommended items.

3.2. Distribution of Recommended Items

Coverage, serendipity, and diversity are the factors related to the distribution of recommended items. However, measurement of these factors cannot be grasped from the viewpoint of the user, and serendipity and diversity do not lend themselves to a distributional configuration of recommended items belonging to two or more categories. Therefore, on a platform such as YouTube, there are limits to our ability to evaluate an RS that can recommend multiple items in two or more categories. In this study, accordingly, the HHI was selected as the measure with which to elucidate the distributional characteristics of items belonging to three or more categories.

Originally, the HHI was determined as the sum of the squares of the percentage of sales of competitors within an industry (Lu et al., 2017). The smaller the sum of the squares, the fiercer the market competition. Therefore, the higher the HHI, the more items there are in the same category, and because items of the same category are mainly recommended along with general content searched by users, a high HHI means that a recommendation consists of items in the same categories as the searched content. Conversely, a low HHI means that the recommendation provides a similar number of items across multiple categories. Therefore, we can posit that the value of HHI will be low when serendipity or diversity is high.

In this study, the HHI was examined in three types of distribution: the distribution of items by category (HHI_C), the distribution of items by number of views (HHI_V), and the distribution of items by playback time (HHI_T). In addition, the HHI considers customized content among the five recommended items listed at the top of the RS screen, which can be interpreted as an auxiliary distribution of recommended items. Accordingly, we examine whether the content listed at the top of screen belongs to the same category has the selected content (1: same, 2: different).

3.3. Content of Recommended Items

We included the average number of views and the average playback time of all the content recommended as two characteristics of recommendations. First, views are defined as the number of click-throughs by other users of a recommended item; we use this as a proxy for peer evaluation. Scholars have traditionally regarded user ratings as an important measure for evaluating recommendations (Salter, 2006). Generally, users of an RS are influenced by others' choices. However, in the case of a social media platform such as YouTube, views can be considered as a useful proxy for peer reviews if there are no user ratings of specific content.

Many views also represent network externality. Network externality is defined as follows: "the utility that a user derives from consumption of the good increases with the number of other agents consuming the good" (Katz and Shapiro, 1985, p. 424). Just as an increased number of users provides added benefits to existing users, a high number of views provides benefits to users of an RS. That is, the network externality that appears as a high number of views is attractive for hedonic purposes, as it enhances social ties (Wang and Chen, 2012) and boosts the perceived value of the content (Yoo et al., 2010). However, evaluation of RS performance in terms of the number of views of certain recommended items is limited because with platforms such as YouTube, in which this study is interested, the RS must recommend multiple items. Therefore, in this study, we measure the content of multiple recommended items as the average number of views within the whole list of recommended results.

Another valuable way to measure RS performance is to examine playback time, which can affect user satisfaction with recommendations (Zhang et al., 2018). On the other hand, playback time may vary depending on the context of viewing, such as user preference (Wallner et al., 2019) or the purpose of viewing (Li and Zhou, 2018). In this study, the mean value of playback time was considered because of our focus on multiple recommendations.

3.4. Control Variables

As a control variable, we considered the characteristics of previously selected content. In a content-based filtering system such as that used with YouTube, recommendations to users are highly related to previously searched and selected content. Therefore, it should be controlled because of its potential effect on user satisfaction. In this study, we considered the number of views and the playback time of the selected content as characteristics of the selected content.

The purpose of watching a given video may be utility-oriented or hedonic. Users may seek recommendations in order to solve a problem or acquire knowledge. Hedonic purposes include joy or emotional reaction. Some studies have shown an association between hedonic purposes and higher levels of satisfaction with recommended outcomes (Lee and

Hosanagar, 2016). Therefore, this factor must also be controlled.

Finally, gender and religion are controlled. Several studies have demonstrated a gender difference regarding the negative impact of various media. For example, Cooper and Mackie (1986) argued that girls are more affected than boys after playing violent games. One recent study, however, obtained mixed results (Tang and Wanoto, 2016). We postulate that there is a gender difference in terms of recommendations, in particular with regard to ethical issues. This may also occur with religion, as recommendations for hedonic purposes may raise ethical concerns among members of some highly religious groups. Therefore, we control this factor because religiosity can affect satisfaction with recommendations.

<Table 1> summarizes the variables that are expected to affect user satisfaction with recommendations.

Ⅳ. Research Methodology

4.1. Experimental Design

The RS of the YouTube platform helps users find customized content by continuously increasing content data (Covington et al., 2016) and providing information of various types (Zhou et al., 2016), such as number of views, playback time, comments, and number of subscribers. In this study, we designate the characteristics of distribution of suggested content, the characteristics of recommended content, and selected content, gender, and religion as control variables influencing user satisfaction.

We conducted a survey over a period of 4 months, from December 2018 to March 2019, to measure user satisfaction with recommended results on YouTube. In a sample of undergraduate students,

<Table 1> Independent Variables

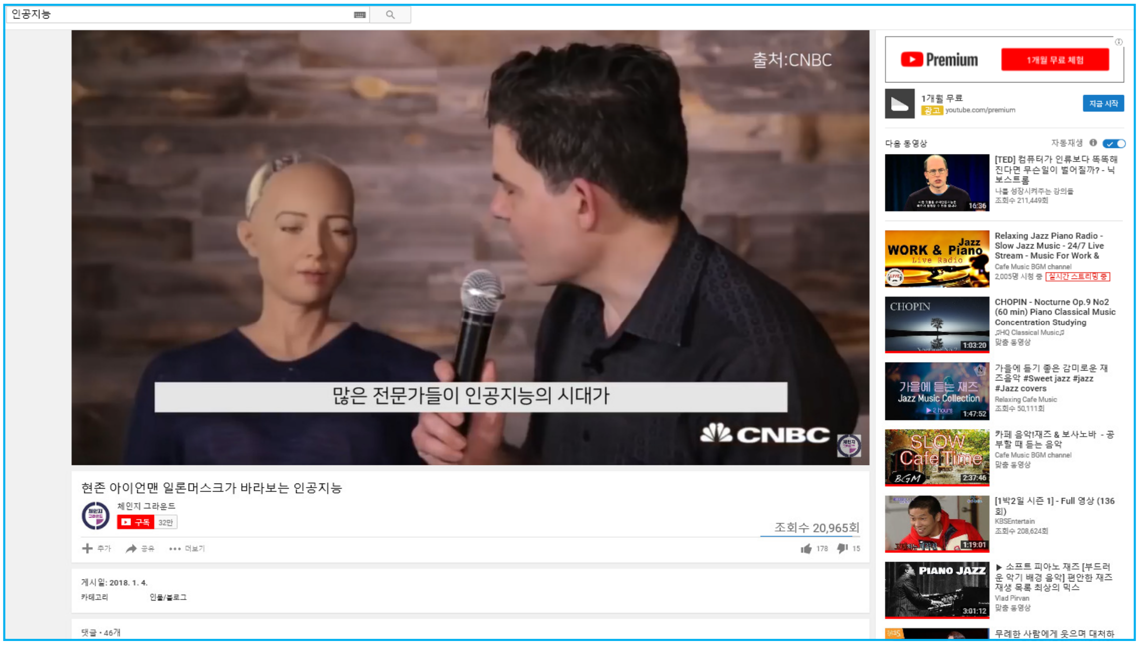
Category	Variable	Definition
Distribution of recommended items	HHI_C	Concentration of content categories (HHI) $\sum_{i=1}^{m} S_i^2$ m<n, where i is the category and Si is the number of recommended items belonging to the i-th category
	HHI_V	Concentration of each recommended item (number of views)
	HHI_T	Concentration of each recommended item (playback time)
	N_CUST	Amount of customized content among the top 5 recommended items
	N_BUND	Amount of bundled content among the top 5 recommended items
	CAT_F	Top recommended item belongs to the same category as the selected content (1: same, 2: different)
Content of recommended items	VIEW_AVG	Average number of views of all recommended items
	TIME_AVG	Average playback time of all recommended items
Control variables	VIEW	Views of selected content
	TIME	Playback time of selected content
	PURPOSE	1: Utilitarian Purpose, 2: Hedonic Purpose
	GENDER	1: Male, 2: Female
	RELIGION	1: Yes, 2: No

those in their 20s and 30s had higher usage rates when compared to other age groups (Moghavvemi et al., 2017). Therefore, we focused on user satisfaction with recommendations of YouTube videos provided by the RS to undergraduates in their 20s. Data for 149 respondents were used in the analysis, and items with low response rates were excluded.

The experiment was conducted in a computer lab, and one PC was assigned per participant. Participants voluntarily partook in the experiment, receiving a participation fee of 5,000 won at the end of the experiment. The questionnaire started with an explanation of user satisfaction with the suggested content provided by the RS on YouTube and a description of the study in order to facilitate understanding of the purpose of the survey. Participants ran YouTube on a web browser using the PC offered according to the guidance of the experimenter. First, the participants determined whether they were going to search for content for the purpose of completing tasks (work, information, project execution, etc.) or pleasure (entertainment, diversion, etc.), selecting relevant keywords. Then, the users decided whether to scroll down to see more content and to which

<Table 2> YouTube Categories

1 Entertainment	7 Comedy	13 Travel & Events
2 Music	8 How-to & Style	14 Pets & Animals
3 People & Blogs	9 News & Politics	15 Nonprofits & Activism
4 Film & Animation	10 Sports	16 Shows
5 Gaming	11 Science & Technology	17 Movies
6 Education	12 Autos & Vehicles	18 Trailers



<Figure 2> Example of Multiple RS

category the selected content belonged. For reference, YouTube classifies content into 18 categories (<Table 2>) (https://www.blogkens.com/blog/youtube-video-tatistics-infographic/). The classification was suggested by Vargas et al. (2014) and has been used for defining the diversity of an item list on the premise that genre diversity corresponds to users' perceptions of diverse recommendations. Dissimilarities between items on the list were considered to represent category heterogeneity.

Next, the participants described why they chose selected content in response to questions about multiple attributes including images, image titles, text titles, views, and playback time. They also responded to questions about satisfaction with the selected content based on ratings, video technologies, expert evaluations, and comments. Responses were scored on a 7-point Likert scale.

<Table 3> Results of Supplementary Regression Analysis

Purpose	Total		Utilitarian Purpose		Hedonic Purpose	
Model	(1)	(2)	(3)	(4)	(5)	(6)
HHI_C	**0.433***** *(6.620)*	**0.428***** *(6.858)*	**0.426***** *(4.314)*	**0.410***** *(4.380)*	**0.491***** *(4.954)*	**0.474***** *(5.322)*
HHI_V	-0.081 (-1.406)	-0.084 (-1.500)	-0.007 (-0.065)	-0.033 (-0.364)	**−0.160*** *(-1.948)*	**−0.148*** *(-1.929)*
HHI_T	0.050 (0.820)	0.055 (1.006)	0.138 (1.179)	0.097 (1.176)	-0.039 (-0.439)	-0.012 (-0.156)
N_CUST	-0.002 (-0.025)	-0.002 (-0.040)	-0.102 (-0.979)	-0.093 (-0.921)	0.117 (1.388)	0.097 (1.207)
N_BUND	0.040 (0.662)	0.036 (0.650)	0.004 (0.040)	-0.020 (-0.229)	0.136 (1.488)	0.127 (1.628)
CAT_F	**−0.432***** *(-7.084)*	**−0.435***** *(-7.344)*	**−0.451***** *(-4.896)*	**−0.446***** *(-5.014)*	**−0.405***** *(-4.616)*	**−0.424***** *(-5.016)*
VIEW_AVG	-0.011 (-0.082)		0.056 (0.492)		0.037 (0.178)	
TIME_AVG	-0.005 (-0.053)		-0.073 (-0.623)		-0.171 (-0.967)	
VIEW	-0.005 (-0.036)		-0.141 (-1.196)		-0.052 (-0.250)	
TIME	-0.012 (-0.133)		-0.114 (-1.169)		0.977 (0.332)	
SAT	0.024 (0.422)		0.068 (0.709)		1.060 (0.293)	
PURPOSE	**0.134**** *(2.381)*	**0.129**** *(2.410)*				
GENDER	-0.073 (-1.284)	-0.074 (-1.328)	**−0.169*** *(-1.845)*	-0.446 (-5.014)	-0.023 (-0.282)	-0.023 (-0.296)
RELIGION	**−0.110*** *(-1.979)*	**−0.112**** *(-2.111)*	-0.037 (-0.418)	-0.134 (-1.526)	-0.102 (-1.282)	-0.104 (-1.393)
R^2	0.583	0.597	0.542	0.553	0.586	0.600
F-value	15.797	25.389	7.276	11.680	9.486	15.653

Because YouTube changes the number of views of a given video in real time (e.g., 20 items appear in personal computer while 16 items are displayed in smartphones), participants submitted information regarding the number of views and playback time of their suggested content using the screen capture function. <Figure 2> is a typical example of a screen capture image of YouTube's recommended results.

After classifying 20 recommended items into two groups (8 items at the top) and scrolling down to view all recommended content, participants selected an item according to their preference of the categories offered by YouTube, recording the number of items in the same category with the selected items among those recommended. Finally, they rated their satisfaction with all recommended content. User satisfaction was measured both with and without scrolling.

4.2. Results

In this study, Models 1 and 2 identified the significant factors influencing user satisfaction with YouTube's RS by distinguishing the distribution of recommended items (HH_C, HHI_V, HHI_T, N_CUST, N_BUND, CAT_F), the content of recommended items (VIEW_AVG, TIME_AVG), and control variables (VIEW, TIME, PURPOSE, GENDER, RELIGION). Since the factor representing the purpose of viewing garnered the most significant results, we conducted a supplementary analysis on male and female groups separately, the results of which are presented in Models 3 – 6. The models included the variables related to distributional characteristics and control variables. Content characteristics were excluded in order to isolate factors relevant to the RS for each model. The results of the multiple regression analysis are listed in <Table 3>.

Ⅴ. Discussion and Results

5.1. Main Findings

In the first model proposed in this study, HHI_C (0.433) and PURPOSE (0.134) had a significant positive impact at the 1% and 5% levels, respectively, and CAT_F (-0.432) and RELIGION (-0.110) had a significant negative impact at the 1% and 10% levels, respectively. These results showed high satisfaction with viewing for hedonic purposes for religious users. High satisfaction was also evident when there were only a few categories provided by the RS and when the content at the top was in the same category as the selected content. In the second model, although the significant variables were the same as in the first model, religion declined in importance to under 5%. Users with religious tendencies seemed to feel more satisfied when content characteristics were excluded. Moreover, R^2 values for these two models were 0.583 and 0.597, respectively, explaining 58.3% and 59.7% of the variance, respectively. In the third model focusing on utilitarian purposes for viewing, HH_C (0.426) had a significant positive impact at the 1% level, and CAT_F (-0.451) and GENDER (-0.169) had a significant negative impact at the 1% and 10% levels, respectively. Thus, respondents expressed high satisfaction when the purpose of viewing was utilitarian, when there were only a few categories provided by the RS, and when the top content was in the same category as the selected content. This was especially true for male participants. In the fourth model, HHI_C (0.410) and CAT_F (-0.446) had a significant positive impact at the 1% level and a significant negative impact at the 1% level, respectively, as in the third model. However, when selected characteristic variables were excluded, gender had no significant impact. It appears that gender does not affect satisfaction if content characteristics are excluded when the purpose of viewing is utilitarian. R^2 values were 0.542 and 0.553, respectively, explaining 54.2% and 55.3% of the variance, respectively. In the fifth model, where viewers searched for hedonic purposes, HHI_C (0.491) had a significant positive impact at the 1% level, and HHI_V (-0.160) and CAT_F (-0.405) had a significant negative impact at the 10% and 1% levels, respectively. These results showed high satisfaction when the purpose of viewing was hedonic, there were fewer categories provided by the RS, and when the items at the top were in the same category as the selected content. These results indicated high satisfaction when the categories of recommended content were similar. In the sixth model, as in the fifth model, HHI_C (0.474) had a significant positive impact at the 1% level, and HHI_V (-0.148) and CAT_F (-0.424) had a significant negative impact at the 10% and 1% levels, respectively. Thus, the characteristics of recommended content had no impact when the purpose of viewing was hedonic. R^2 values were 0.586 and 0.600, respectively, explaining 58.6% and 60.0% of the variance, respectively.

5.2. Contributions

This study has several theoretical and practical contributions, as follows. First, this study utilizes the HHI to measure the characteristics of recommended items; as a result, we discovered that the category of recommended items was a particularly significant factor explaining satisfaction with recommendations. Coverage measures the number of items recommended by the RS, and serendipity represents the situation when users are exposed to different items from selected categories unexpectedly. By themselves, these variables do not provide enough information on the value of multiple recommended items. On the other hand, information provided by the HHI reveals the best content ratio in each category. Although the HHI was originally applied in other fields such as economics to measure parameters such as industry concentration (Qiao and Li, 2018), market share, or sales volume, and it is a frequently utilized index in research on market structure (Bremus and Buch, 2015), social network structure (Li et al., 2019), and portfolio configuration (Zhang et al., 2018), our study is the first use the HHI to explain the distribution of items recommended by an RS.

For websites like YouTube that recommend multiple items in diverse categories using content-based filtering systems, we discovered that the category concentration of recommended items actually affects user satisfaction with the recommendations. In addition, the results of our analysis confirmed differences in the influence of various factors according to the purpose of viewing. Utilitarian and hedonic purposes of viewing have different functions; therefore, the RS must evaluate the purpose of the recommended content, changing recommendations accordingly and evaluating content in terms of the potential pleasure it may bring the user (Huang, 2016). The results for users viewing for hedonic purposes are consistent with those of previous research stating that consumers are more satisfied with recommendations of hedonic products than with utility-oriented products (Lee and Hosanagar, 2016). Research has shown that utilitarian goods are those for which consumption is cognitively driven, instrumental, and goal-oriented, intended to accomplish a functional or practical task (De et al., 2010; Hirschman and Holbrook, 1982; Meseguer-Martinez et al., 2017; Strahilevitz and Myers, 1998). On the other hand, hedonic goods are those whose consumption is primarily characterized by an affective and sensory experience of aesthetic or sensual pleasure, fantasy, and fun. The object of this study is to evaluate the ability of a RS to recommend items, including distinguishing between hedonic and utilitarian products.

This study also has practical implications. First, the strategy for revealing of recommended items should differ according to the purpose of viewing. Although satisfaction increased for users viewing for both utilitarian and hedonic purposes, the increase was greater for those viewing for hedonic purposes. Thus, items in similar categories should be recommended to those that users are already consuming. Surprisingly, multiple recommendations had little influence on satisfaction for participants viewing for utilitarian purposes. This result seems to reflect a tendency away from diversity in problem solving.

Second, We can establish a strategy that the first content at the top of the recommended content list should be the same category as the content selected by the user. Content belonging to the same category is highly interrelated, and when it is exposed to users, users may have a positive perception of the content (Bayer and Stubber, 2010). This can be used not only in the YouTube subject to this study, but also in various areas where recommended systems can be applied, such as fashion, advertising, and social media (Lee et al., 2012).

Third, the results suggest that recommending items based on the content viewed by many other users increases network externality for users viewing for hedonic purposes. This result is consistent with those in previous literature, which indicated that network externalities can enhance perceived usefulness and enjoyment for SNS users (Lin and Bhattacherjee, 2008; Lin and Lu, 2011; Zhou and Lu, 2011). Therefore, RSs can improve user satisfaction and enhance network externality by recommending items with a large number of views and suggesting diverse content for users viewing for hedonic purposes.

5.3. Limitations

This study was conducted using a sample of adult undergraduates in a computer lab. Despite efforts to ensure the reliability of the results by randomly dividing the sample according to viewing purpose and precisely measuring the HHI via screenshots without recording the recommended items, we did not consider participants of various ages. In addition, this experiment was conducted in a specific place, not an actual use environment. For example, in some working environments, the number of views might be higher for utilitarian purposes; in this study, students in the utilitarian purpose group were given assigned tasks which may or may not accurately simulate a real use environment. Future researchers should try to reproduce the results of this study using data from other groups beside students. For all these reasons, generalizations of the results of this research should be made with caution. In this study, we conducted an empirical analysis of an RS recommending videos on YouTube. Another interface other than YouTube might alter the attitudes or emotional reactions of users, which would also affect their evaluation of the RS. In addition, the configuration of the recommended items may affect satisfaction with recommendations in another interface. Future studies should consider not only recommended items, but also users' emotions in relation to the configuration of recommended items.

5.4. Conclusion

RSs that recommend multiple items in diverse categories have been increasing. The HHI showed that this factor had a significant impact on user satisfaction with recommended items. In addition, differences among factors in terms of degree of influence on user satisfaction were found according to the purpose of viewing. In the case of viewing for hedonic purposes, the number of views was not concentrated, and satisfaction was higher when there was a variety of configurations.

Although satisfaction increased in both the utilitarian and hedonic groups, the increase was larger for those viewing for hedonic purposes. This study is the first to discover that the category concentration of items actually impacts user satisfaction in the case of websites operating RSs which suggest various items in different categories using content-based filtering systems, such as YouTube. Moreover, this study utilized the HHI to measure the distributional characteristics of recommended items, revealing the category HHI of recommended items as a crucial variable explaining user satisfaction with recommendations.

Acknowledgements

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2017S1A3A2066740).

<References>

[1] Abbas, M., Riaz, M. U., Rauf, A., Khan, M. T., and Khalid, S. (2017, December). Context-aware Youtube recommender system. In *2017 International Conference on Information and Communication Technologies (ICICT)* (pp. 161-164). IEEE.

[2] Adamopoulos, P., and Tuzhilin, A. (2015). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST), 5*(4), 54.

[3] Adomavicius, G., and Kwon, Y. (2011, October). Maximizing aggregate recommendation diversity: A graph-theoretic approach. In *Proc. of the 1st International Workshop on Novelty and Diversity in Recommender Systems (DiveRS 2011)* (pp. 3-10).

[4] Anand, D., and Bharadwaj, K. K. (2011). Utilizing various sparsity measures for enhancing accuracy of collaborative recommender systems based on local and global similarities. *Expert systems with applications, 38*(5), 5101-5109.

[5] Anderson, C. (2006). *The long tail: Why the future of business is selling less of more*. Hachette Books.

[6] Baier, D., and Stüber, E. (2010). Acceptance of recommendations to buy in online retailing. *Journal of Retailing and Consumer Services, 17*(3), 173-180.

[7] Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-based systems, 46*, 109-132.

[8] Bremus, F., and Buch, C. M. (2015). Banking market structure and macroeconomic stability: Are low income countries special? *Pacific Economic Review, 20*(1), 73-100.

[9] Choi, J., Lee, H. J., and Kim, Y. C. (2011). The influence of social presence on customer intention to reuse online recommender systems: The roles of personalization and product type. *International Journal of Electronic Commerce, 16*(1), 129-154.

[10] Cooper, J., and Mackie, D. (1986). Video games and aggression in children 1. *Journal of Applied Social Psychology, 16*(8), 726-744.

[11] Covington, P., Adams, J., and Sargin, E. (2016, September). Deep neural networks for Youtube recommendations. In *Proceedings of the 10th ACM conference on recommender systems* (pp. 191-198). ACM.

[12] De, P., Hu, Y., and Rahman, M. S. (2010). Technology usage and online sales: An empirical study. *Management Science, 56*(11), 1930-1945.

[13] Ferwerda, B., Graus, M. P., Vall, A., Tkalcic, M.,

and Schedl, M. (2016, August). The influence of users' personality traits on satisfaction and attractiveness of diversified recommendation lists. In *EMPIRE@ RecSys* (pp. 43-47).

[14] Gatian, A. W. (1994). Is user satisfaction a valid measure of system effectiveness? *Information & Management, 26*(3), 119-131.

[15] Ge, M., Delgado-Battenfeld, C., and Jannach, D. (2010, September). Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of the fourth ACM conference on Recommender systems* (pp. 257-260). ACM.

[16] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS), 22*(1), 5-53.

[17] Hirschman, E. C., and Holbrook, M. B. (1982). Hedonic consumption: emerging concepts, methods and propositions. *Journal of marketing, 46*(3), 92-101.

[18] Hong, S. E., and Kim, H. J. (2016, July). A comparative study of video recommender systems in big data era. In *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)* (pp. 125-127). IEEE.

[19] Huang, L. T. (2016). Exploring utilitarian and hedonic antecedents for adopting information from a recommendation agent and unplanned purchase behaviour. *New Review of Hypermedia and Multimedia, 22*(1-2), 139-165.

[20] Iaquinta, L., De Gemmis, M., Lops, P., Semeraro, G., Filannino, M., and Molino, P. (2008, September). Introducing serendipity in a content-based recommender system. In *2008 Eighth International Conference on Hybrid Intelligent Systems* (pp. 168-173). IEEE.

[21] Iaquinta, L., de Gemmis, M., Lops, P., Semeraro, G., and Molino, P. (2010). Can a recommender system induce serendipitous encounters?. In *E-commerce*. IntechOpen.

[22] Jalili, M., Ahmadian, S., Izadi, M., Moradi, P., and Salehi, M. (2018). Evaluating collaborative filtering recommender algorithms: a survey. *IEEE Access, 6*, 74003-74024.

[23] Kaminskas, M., and Bridge, D. (2014, October). Measuring surprise in recommender systems. In *Proceedings of the workshop on recommender systems evaluation: dimensions and design* (Workshop programme of the 8th ACM conference on recommender systems).

[24] Kaminskas, M., and Bridge, D. (2017). Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems (TiiS), 7*(1), 2.

[25] Katz, M. L., and Shapiro, C. (1985). Network externalities, competition, and compatibility. *American Economic review, 75*(3), 424-440.

[26] Kotkov, D., Konstan, J. A., Zhao, Q., and Veijalainen, J. (2018, April). Investigating serendipity in recommender systems based on real user feedback. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing* (pp. 1341-1350). ACM.

[27] Kotkov, D., Wang, S., and Veijalainen, J. (2016). A survey of serendipity in recommender systems. *Knowledge-Based Systems, 111*, 180-192.

[28] Kumar, A., Gupta, S., Singh, S. K., and Shukla, K. K. (2015, August). Comparison of various metrics used in collaborative filtering for recommendation system. In *2015 Eighth International Conference on Contemporary Computing (IC3)* (pp. 150-154). IEEE.

[29] Kunaver, M., and Požrl, T. (2017). Diversity in recommender systems:A survey. *Knowledge-Based Systems, 123*, 154-162.

[30] Lee, D., and Hosanagar, K. (2016, April). When do recommender systems work the best?: The moderating effects of product attributes and consumer reviews on recommender performance. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 85-97). International World Wide Web Conferences Steering Committee.

[31] Lee, J. K., Wang, J., and Lee, W. (2012). How is strategic brand alliance communicated to consumers? A content assessment of print ads. *The Journal of Advertising and Promotion Research, 1*(2), 5-40.

[32] Li, C., and Zhou, H. (2018). Enhancing the efficiency of massive online learning by integrating intelligent analysis into MOOCs with an application to education of sustainability. *Sustainability, 10*(2), 468.

[33] Li, W., Xu, G., Xing, Q., and Lyu, M. (2019). Effect of directors' social network centrality on corporate charitable donation. *Social Behavior and Personality: An international journal, 47*(3), 1-9.

[34] Liang, T. P., Lai, H. J., and Ku, Y. C. (2006). Personalized content recommendation and user satisfaction: Theoretical synthesis and empirical findings. *Journal of Management Information Systems, 23*(3), 45-70.

[35] Lin, C. P., and Bhattacherjee, A. (2008). Elucidating individual intention to use interactive information technologies: The role of network externalities. *International Journal of Electronic Commerce, 13*(1), 85-108.

[36] Lin, K. Y., and Lu, H. P. (2011). Intention to continue using Facebook fan pages from the perspective of social capital theory. *Cyberpsychology, Behavior, and Social Networking, 14*(10), 565-570.

[37] Lu, J., Wu, D., Mao, M., Wang, W., and Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems, 74*, 12-32.

[38] Lü, L., Medo, M., Yeung, C. H., Zhang, Y. C., Zhang, Z. K., and Zhou, T. (2012). Recommender systems. *Physics reports, 519*(1), 1-49.

[39] Lu, C., Qiao, J., and Chang, J. (2017). Herfindahl–Hirschman Index based performance analysis on the convergence development. *Cluster computing, 20*(1), 121-129.

[40] McNee, S. M., Riedl, J., and Konstan, J. A. (2006, April). Being accurate is not enough: how accuracy metrics have hurt recommender systems. In *CHI'06 extended abstracts on Human factors in computing systems* (pp. 1097-1101). ACM.

[41] Mehrotra, R., McInerney, J., Bouchard, H., Lalmas, M., and Diaz, F. (2018, October). Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* (pp. 2243-2251). ACM.

[42] Meseguer-Martinez, A., Ros-Galvez, A., and Rosa-Garcia, A. (2017). Satisfaction with online teaching videos: A quantitative approach. *Innovations in Education and Teaching International 54*(1), 62-67.

[43] Moghavvemi, S., Sulaiman, A. B., Jaafar, N. I. B., and Kasem, N. (2017, July). Facebook and YouTube addiction: The usage pattern of Malaysian students. In *2017 international conference on research and innovation in information systems (ICRIIS)* (pp. 1-6). IEEE.

[44] Murakami, T., Mori, K., and Orihara, R. (2007, June). Metrics for evaluating the serendipity of recommendation lists. In *Annual conference of the Japanese society for artificial intelligence* (pp. 40-46). Springer, Berlin, Heidelberg.

[45] Nakatsuji, M., Fujiwara, Y., Tanaka, A., Uchiyama, T., Fujimura, K., and Ishida, T. (2010, October). Classical music for rock fans?: novel recommendations for expanding user interests. In *Proceedings of the 19th ACM international conference on Information and knowledge management* (pp. 949-958). ACM.

[46] Nguyen, T. T., Harper, F. M., Terveen, L., and Konstan, J. A. (2018). User personality and user satisfaction with recommender systems. *Information Systems Frontiers, 20*(6), 1173-1189.

[47] Ogara, S. O., Koh, C. E., and Prybutok, V. R. (2014). Investigating factors affecting social presence and user satisfaction with mobile instant messaging. *Computers in Human Behavior, 36*, 453-459.

[48] Onuma, K., Tong, H., and Faloutsos, C. (2009, June). TANGENT: A novel, 'Surprise me', recommendation algorithm. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 657-666). ACM.

[49] Qiao, Z., and Li, Z. (2018). Do foreign institutional investors enhance firm innovation in China? *Applied Economics Letters*, 1-4.

[50] Qin, S., Menezes, R., and Silaghi, M. (2010, August). A recommender system for youtube based on

its network of reviewers. In *2010 IEEE Second International Conference on Social Computing* (pp. 323-328). IEEE.

[51] Salter, J., and Antonopoulos, N. (2006). CinemaScreen recommender agent: Combining collaborative and content-based filtering. *IEEE Intelligent Systems, 21*(1), 35-41.

[52] Shani, G., and Gunawardana, A. (2011). Evaluating recommendation systems. In *Recommender systems handbook*(pp. 257-297). Springer, Boston, MA.

[53] Sridharan, S. (2014). *Introducing serendipity in recommender systems through collaborative methods.* Master's Theses. University of Rhode Island.

[54] Strahilevitz, M., and Myers, J. G. (1998). Donations to charity as purchase incentives: How well they work may depend on what you are trying to sell. *Journal of consumer research, 24*(4), 434-446.

[55] Tang, T. Y., and Winoto, P. (2016). I should not recommend it to you even if you will like it: The ethics of recommender systems. *New Review of Hypermedia and Multimedia, 22*(1-2), 111-138.

[56] Thorat, P. B., Goudar, R. M., and Barve, S. (2015). Survey on collaborative filtering, content-based filtering and hybrid recommendation system. *International Journal of Computer Applications, 110*(4), 31-36.

[57] Vargas, S., and Castells, P. (2011, October). Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 109-116). ACM.

[58] Vargas, S., Baltrunas, L., Karatzoglou, A., and Castells, P. (2014, October). Coverage, redundancy and size-awareness in genre diversity for recommender systems. In *Proceedings of the 8th ACM Conference on Recommender systems* (pp. 209-216). ACM.

[59] Wallner, G., Schinnerl, C., Schiller, M. H., Calvo, A. M., Pirker, J., Sifa, R., and Drachen, A. (2019). Beyond the individual: Understanding social structures of an online player matchmaking website. *Entertainment Computing, 30*, 100284.

[60] Wang, E. S. T., and Chen, L. S. L. (2012). Forming relationship commitments to online communities: The role of social motivations. *Computers in Human Behavior, 28*(2), 570-575.

[61] Yao, Y. Y. (1995). Measuring retrieval effectiveness based on user preference of documents. *Journal of the American Society for Information Science, 46*(2), 133-145.

[62] Yoo, W. S., Lee, Y., and Park, J. (2010). The role of interactivity in e-tailing: Creating value and increasing satisfaction. *Journal of retailing and consumer services, 17*(2), 89-96.

[63] Yoo, S. J. (2016). Reliability-based similarity evaluation model for collaborative filtering. *Journal of Advanced Information Technology and Convergence, 14*(12), 141-146.

[64] Zeng, A., Yeung, C. H., Medo, M., and Zhang, Y. C. (2015). Modeling mutual feedback between users and recommender systems. *Journal of Statistical Mechanics: Theory and Experiment, 7*, 07020.

[65] Zhang, S., Zhao, T., and Xie, B. C. (2018). What is the optimal power generation mix of China? An empirical analysis using portfolio theory. *Applied energy, 229*, 522-536.

[66] Zhang, W., Gu, T., Sun, W., Phatpicha, Y., Chang, L., and Bin, C. (2018, September). Travel attractions recommendation with travel spatial-temporal knowledge graphs. In *International Conference of Pioneering Computer Scientists, Engineers and Educators* (pp. 213-226). Springer, Singapore.

[67] Zhang, Y. C., Séaghdha, D. Ó., Quercia, D., and Jambor, T. (2012, February). Auralist: Introducing serendipity into music recommendation. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 13-22). ACM.

[68] Zhou, R., Khemmarat, S., Gao, L., Wan, J., and Zhang, J. (2016). How YouTube videos are discovered and its impact on video views. *Multimedia Tools and Applications, 75*(10), 6035-6058.

[69] Zhou, T., and Lu, Y. (2011). Examining mobile instant messaging user loyalty from the perspectives of network externalities and flow experience.

Computers in Human Behavior, 27(2), 883-889.

[70] Ziegler, C. N., McNee, S. M., Konstan, J. A., and Lausen, G. (2005, May). Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web* (pp. 22-32). ACM.

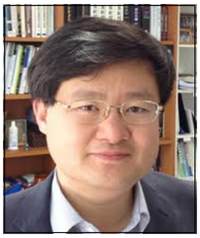
◆ About the Authors ◆

Janghun Jeong

Janghun Jeong is received a Bachelor's degrees in School of Management from Mokwon University, Korea. He is currently taking a master's course major in Big Data Management at Kyung Hee University, Korea. And also working at the Research Center for Advanced Information Technology (CAITech). His research interests include big data analytics and data mining.

Kwonsang Sohn

Kwonsang Sohn is a Ph.D student in the School of Management at Kyung Hee University, Korea. He also working at the Research Center for Advanced Information Technology (CAITech) as a business analytics researcher. He received a Master's degree in Entrepreneurship from Inha University, Korea. His research interests include business applications of artificial intelligence, intelligent systems, and data-driven decision-making. His research has been published in *information systems review.*



Ohbyung Kwon

Ohbyung Kwon is presently a full professor at School of Management, Kyung Hee University, Korea, where he initially joined in 2004. In 2002, he worked Institute of Software Research International (ISRI) at Carnegie Mellon University to perform a project on context-aware computing, web service and semantic web. He received BA at Seoul National University and Ph.D. degree at KAIST in 1988 and 1995, respectively. He was also an adjunct professor at San Diego State University (SDSU). His current research interests include data science, human robot interaction, context-aware services, e-commerce and DSS. His research has been published in *Journal of Management Information Systems, Decision support systems, Expert Systems with Applications, Telematics & Informatics, Electronic Commerce Research and Applications, Computers in Human Behaviour, Computers in Industry, International Journal of Information Management, Journal of Computer Information Systems,* and many other journals.

Submitted: May 11, 2019; 1st Revision: September 3, 2019; Accepted: September 25, 2019