International Journal of Database Theory and Application
Vol.9, No.9 (2016), pp.11-18
http://dx.doi.org/10.14257/ijdta.2016.9.9.02

Personal Preference Based Movie Recommendation System

Sang-Hyun You[1], Jeawon Park[2] and Jaehyun Choi[3*]

[1,2,3] *Graduate school of Software, Soongsil University, Seoul, Korea*
evillusi@gmail.com[1], jwpark@ssu.ac.kr[2], jaehyun@ssu.ac.kr[3]

Abstract

Recommendation systems sort out the information of user's concerns for supporting decision-making. Today, recommendation systems have a very close relationship with our modern society. However, despite the large amount of information available due to information technological advancement, finding information specific to the user's concern is getting more difficult. In order to handle such issues, the importance of the recommendation system has become apparent. Collaborative filtering is one of the referral systems, which automatically predicts the users' interest based on the information on preference collected from a considerable number of people. However, accuracy issues come to the fore as an insufficient amount of information collected. This paper derived a regression equation using collaborative filtering of user preference information and official movies information to solve the problems, thereby proposing a movie recommendation system. By adding user preference information to the regression equation using only objective movie information, accuracy has been increased by 20%, and the recall ratio by 9%. It has been shown that utilizing preference information increases accuracy for recommendation of movies.

Keywords*: recommendation system, collaborative filtering, regression analysis*

1. Introduction

Because of the introduction of the spread of the Internet and e-commerce, users have been able to obtain information about the item that they want faster. However, due to the rapid increase in the spread of various information devices and information, users were only making a lot more effort than before in finding the information they want in the overflowing information. Companies are using the recommendation system to address these problems and ensure a large number of users. The recommendation system helps users in giving them attention and helping them easily find a desired item. This system recommends online news, movies and a variety of web resource forms in using web sites, and has been used by many e-commerce sites such as Amazon, CDNow, DangDang, Singforyou [1]. Due to the recent development of IOT technology and popularization of smartphones, users can now access a variety of information. However, searching for the information according to the user's interest has become more difficult. In order to handle such issues, a recommendation system has emerged, and the most widely used method for the system is collaborative filtering.

Theater ticket sales last year, according to the 2014 Survey of Film Council, has increased by 7.3% as compared to the sales in 2013, 1.6 trillion won, which was the highest record so far. Moreover, the number of audience has been increased by 0.8% as compared to the year 2013, reaching 215 million. According to the record, each person goes for 4.19 movies every year, keeping the audience record above 200 million for the consecutive years [2]. It is obvious that watching movies for one's leisure has become a trend. However, despite the increasing trend, a movie theater's recommendation system or

* Corresponding author

ISSN: 2005-4270 IJDTA
Copyright ⓒ 2016 SERSC

promotions are kept in an old, one-way form. As a result, customers have to manually search for movie information of their interest, and the problem lies within the time and effort of customers still seeking for what they want in this environment wealth of information. Therefore, a movie recommendation system that can advise the film to the users according to their tendency and preference is found to be more necessary.

This paper is used in a weight by the user preference information through a method film to subject the proposed techniques using the genre information and the information of the user in order to increase the accuracy to reduce rare to cause a significant problem to the accuracy of the recommended [3]. Deriving a regression equation of the user's preferences and the objective film information through the collaborative filtering method, we propose a system like the user for the film.

2. Related Research

M2M technologies enable automated remote monitoring and controlling of smart objects or smart devices (also called smart things). In addition, these technologies provide connectivity between smart objects, facilitating communications between them and resulting in the formation and foundation of the Internet of Things (IOT) [4]. Due to the development of IOT technology and popularization of smartphones, users can now access variety of information. However, searching for information that is of interest to the user has become more difficult. In order to handle such issues, a recommendation system has emerged, and the most widely used method for the system is collaborative filtering.

2.1. Recommendation System

The recommendation system selectively screens for information based on the user's favorites or preference and provides information to the customer. Therefore, the recommendation system can be applied to companies for competitive advantage, such as customer care or sales improvement compared to other companies, and do not provide unnecessary information to improve customer satisfaction and minimize waste of resources [5].

2.2. Collaborative Filtering

Collaborative filtering is one of the main methods applied. For analysis, user evaluation on the item of recommendation is received, and the recommendation is done for users with similarities [6]. Collaborative filtering [7] is largely divided into two, the user information-based collaborative filtering technique gives collect evaluation or preference of the user and connects people with similar tendencies, and provides the user with this information. Item-based collaborative filtering technique [8] uses the information in the items evaluated by users and offers the information of the similar products that tend to favor the use place have to the products favored in the past.

2.3. Cold-Start

Cold-Start [9] is generated in an automated data model, typically caused by problems in the system and the like, of new user information generated when the recommendation information to the new user in the system because there is not enough recommendation. Collaborative filtering methods in screening information based on preference information and items that the user is identified, but this information is not enough if the Cold-Start occurs. If the content-based filtering techniques and the selection information from the user, or information based on the assessment of the product, but not enough, then the Cold-Start is generated [10].

 Copyright ⓒ 2016 SERSC

2.4. Multiple Regression Analysis

TCA-LI (linear interpolation algorithm) has good accuracy for single missing sensor data elements in TCA, but for continuous missing sensor data, TCA-LI cannot provide good estimation data. Therefore, in this section, we will introduce the multiple regression algorithm (TCA-MR) to estimate the continuous missing sensor data of the TCA model [11].

Regression analysis is a statistical process for estimating the relationships among variable factors. It includes many techniques for modeling and analyzing several variables, when the focus is on the relationship between a dependent variable and one or more independent variables (or predictors).

However, this study uses multiple regression analysis and objective information on the film is because two or more independent variables by user preference information. Multiple regression analysis can reduce the error value by introducing additional independent variables, incorrect setting of the model, as well as to eliminate the bias of the coefficient estimator. In general, the multiple regression equation of Y on X1, X2, …, Xk is given by: Figure 1

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \ldots\ldots\ldots\ldots\ldots\ldots + b_k X_k$$

Figure 1. Multiple Regression Model

3. Suggesting System

We use the objective information on the movie and personal tendencies to offer a personalized movie recommendation system. The procedure of the suggested system is as shown in Figure 2.

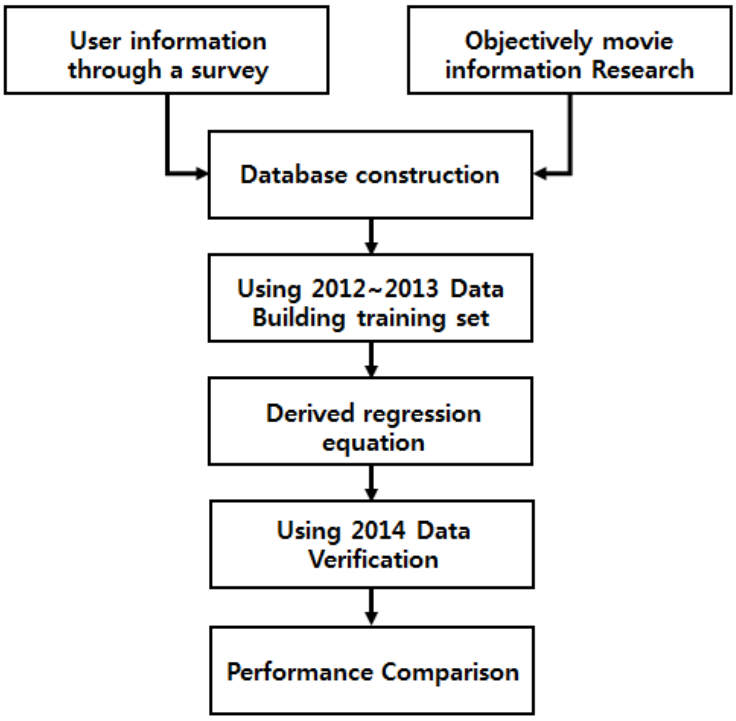

Figure 2. Procedure of the Suggested Movie Recommendation System

Copyright ⓒ 2016 SERSC

The first is to bring objective information about the movie from the Movie Database page on the site Film Council to build movie information database. Objective movie information database included genre, movie release season, country made from, and online reviews score of specialists and netizens. The procedure is shown in Figure 3. The genre information is used to make a single merged genre code.

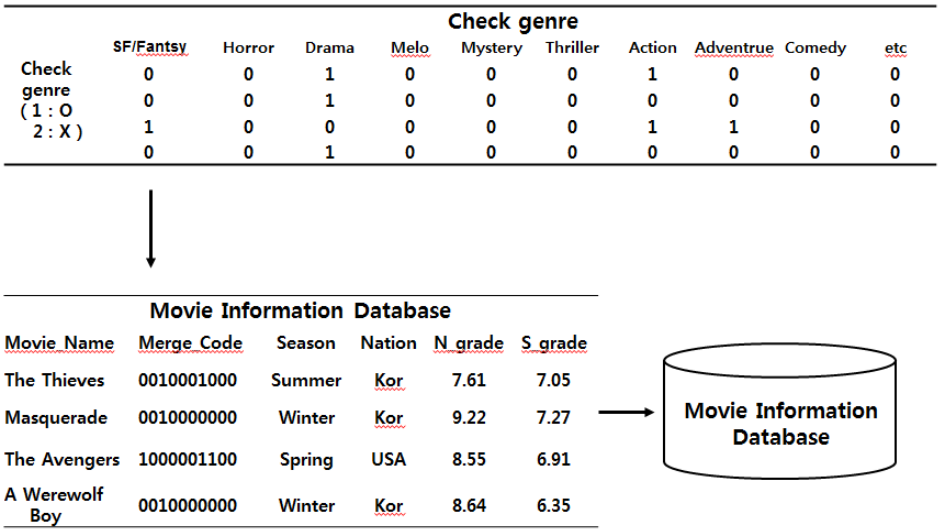

Figure 3. The Structure of Movie Information Database

Also, user data is generated according to the user's survey information and consists of 12 pieces, such as name, gender, blood type, living area, age, favorite nation of making film, released season of film, favorite genre, watching movie with who, prefer movie score, influence the degree by actor, and user`s type. In order to apply regression analysis, if the user saw the movie expressed as 1, another expressed as 0. The procedure is shown in Figure 4.

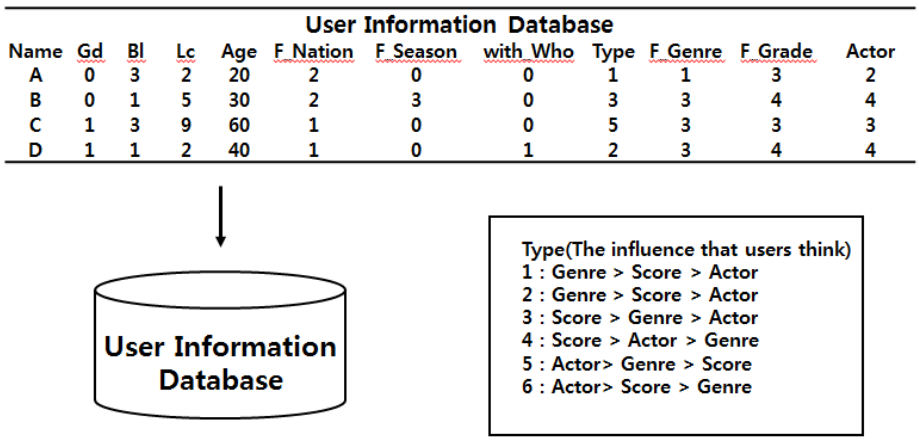

Figure 4. The Structure of User Information Database

Finally, total movie information is created by combining user information and movie information for regression analysis. The procedure is shown in Figure 5. Total movie information is divided into training set and validation set. The regression equations are derived from the training set and are applied to a validation set to predict the probability that the customer will see the movie.

 Copyright ⓒ 2016 SERSC

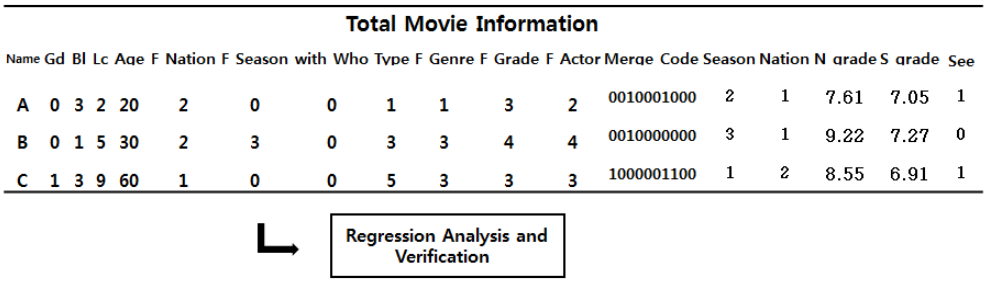

Figure 5. Total Data Combined Movie Information Database and User Information Database

3.1. Data

Objective movie data and user`s preference information have been used for prediction of the probability that the user will watch the movie. Users' preferences information is collected from the survey of 110 people's preferences, and objective movie information has been collected from the Korea box office top 80 films during 2012-2014.

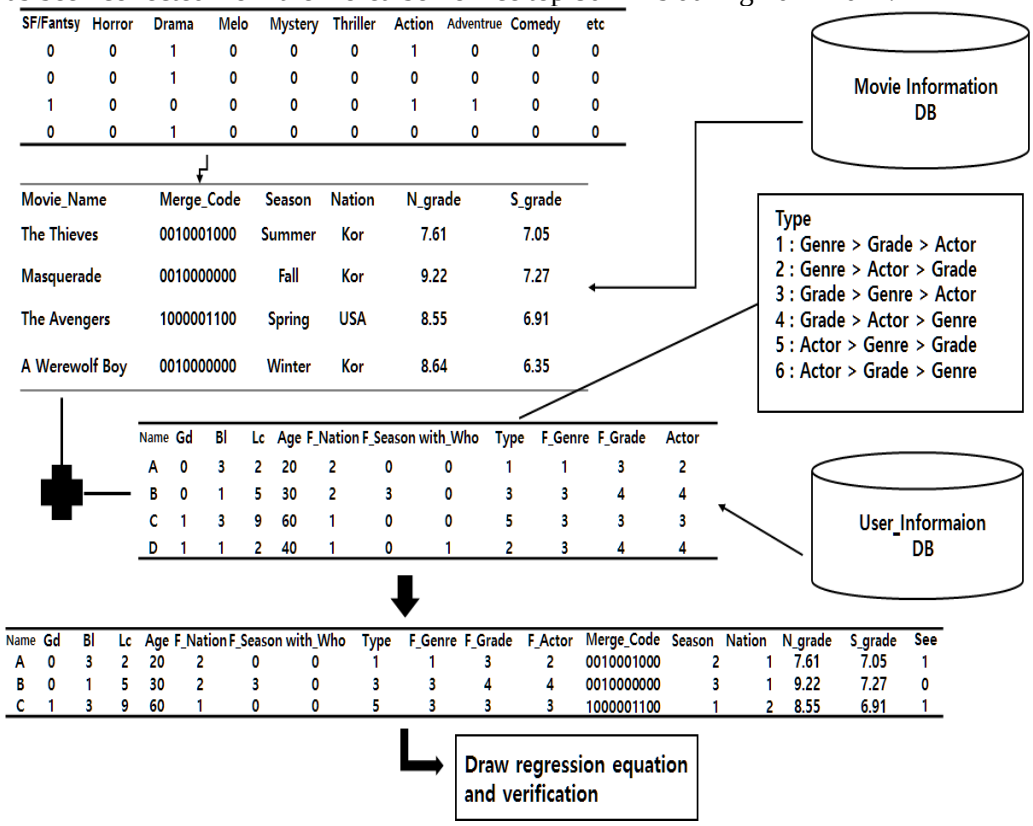

Figure 6. Movie Recommendation System

3.2. Probability Prediction

As shown in Figure 2, the movie information database has been created using objective movie information data, and user information database has been created using data collected from the survey. 26400 data records have been generated using preference information from 110 people and objective movie information on 240 movies from the Korea box office. The regression equation has been derived using 17600 records from

movies watched during year 2012 and 2013, and 8800 records on movies watched during year 2014 are used for verification.

4. Result

For the experiment, we objectively used movie data from the 2012-2014 Korea box office Top 80 movie data from the Film Council web site. Also, from the 110 people receiving preference information related to the film, it was used as the user data. 110 people's preferences, information on whether he/she watched the movies during year 2012-13, and objective movie information have been used as training set. Same people's preferences and information on whether he/she watched the movies during 2014 have been used as validation set. Table 1 shows the contents of the training set and the validation set used. The regression equation was derived using 17600 records, and 8800 records are used for verification. As a result, regression equation derived using both objective movie information and user's preferences has shown higher accuracy and recall ratio, as shown in Figure 3.

Table 1. The Amount of Data

	Objectively Movie Information	Personal tendencies Information	Merge Data
Training Set	**800**	**1320**	**17600**
Verification Set	**400**	**1320**	**8800**
Total	**1200**	**2640**	**26400**

5. Conclusion

Regression equation has been derived using objective movie information and preference information, and has been applied to the proposed system in this paper. By adding user preference information to the regression equation using only objective movie information, accuracy has been increased by 20% and the recall ratio by 9%. It has been shown that utilizing preference information increases accuracy for the recommendation of movies. However, from 2012 to 2014 movie box office top 80, despite a total of 240 to derive a regression equation, with the values of precision and recall data rather than saw the movie expecting a little less grungy it seems less out.

 Copyright ⓒ 2016 SERSC

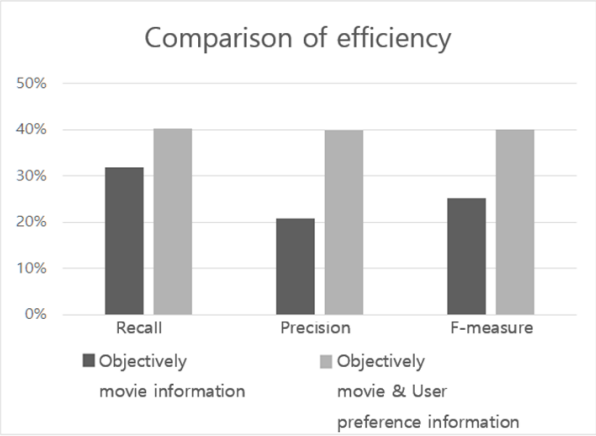

Figure 7. Comparison of Efficiency

References

[1] T. H. Song and Y. H. Wu, "A Collaborative Filtering Recommendation Algorithm Based On Item Classification", Pacific-Asia Conference on Circuits, Communications and System, **(2009)**, pp. 694–697.
[2] Film Commission, "Korea film industry settlement of accounts", **(2014)**.
[3] S. Y. Yun and S. D. Youn, "A Recommendation Technique using Weight of User Information", Journal of Information and Communication Convergence Engineering, **(2011)**, pp. 877-885.
[4] S. Husain, A. Prasad, A. Kunz, A. Papageorgiou and J. S. Song, "Recent Trends in Standards Related to the Internet of Things and Machine-to-Machine Communications", Journal of information and communication convergence engineering, vol. 12, no. 4, **(2014)**, pp. 12, 228-236.
[5] Z. Zui, "A hybrid fuzzy-based personalized recommender system for telecom products/services", Information Sciences, vol. 235, **(2013)**, pp. 117-129.
[6] S. Lee, H. Lee and H. Kim, "Dynamic Recommender on user taste tendency model: Focusing on Movie Recommender Systems", K.I.S.S. Journal, vol. 31, no. 2, **(2004)**, pp. 153-163.
[7] S. Badrul, "Item-based collaborative filtering recommendation algorithms", Proceedings of the 10th international conference on World Wide Web, **(2001)**.
[8] W. Suyun, "Item-based collaborative filtering recommendation algorithm combining item category with interestingness measure", Computer Science & Service System (CSSS), **(2012)**.
[9] B. Justin and T. Hofmann, "Unifying collaborative and content-based filtering", Proceedings of the twenty-first international conference on Machine learning, **(2004)**.
[10] I. Y. Jeong, X. Yang and H. K. Jun, "A Study on Movies Recommendation System of Hybrid Filtering-Based", (J. Korea Inst. Inf. Commun. Eng.), vol. 19, no. 1, **(2015)**, pp. 113-118.
[11] X. Ren, H. T. Sug and H. J. Lee, "A New Estimation Model for Wireless Sensor Networks Based on the Spatial-Temporal Correlation Analysis", Journal of information and communication convergence engineering, vol. 13, no. 2, **(2015)**, pp. 105-112.

Authors

Sang-Hyun You, received his BCS degree in year 2013 in Computer Science from Myongji University, Korea. Currently, he is a graduate student with the Graduate School of Software, Soongsil University, Korea. His research area includes Knowledge Discovery in Databases, Software Engineering and data mining.

Copyright ⓒ 2016 SERSC

Jeawon Park, received his Ph.D. degree in Computer Science from Soongsil University in Korea, 2011. He is a professor at the Graduate School of Software, Soongsil University. His research interests are in areas of Software Testing, Software Process, Web Services, and Project Management.

Jaehyun Choi, received his Ph.D. degree in Computer Science from Soongsil University in Korea, 2011. He is a professor at the Graduate School of Software, Soongsil University. His research interests are in areas of Data Processing, Service Engineering, Software Engineering, and Text Mining.

Copyright ⓒ 2016 SERSC