ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2019 December: 25(4): 89~103 http://dx.doi.org/10.13088/jiis.2019.25.4.089

# Evaluating the Quality of Recommendation System by Using Serendipity Measure

Tserendulam Dorjmaa
Wonju University Innovation Support Project Team
Yonsei University MIRAE Campus
(*martius1025@gmail.com*)

Taeksoo Shin
Division of Business Administration
College of Government and Business
Yonsei University
(*tsshin@yonsei.ac.kr*)

Recently, various approaches to recommendation systems have been studied in terms of the quality of recommendation system. A recommender system basically aims to provide personalized recommendations to users for specific items. Most of these systems always recommend the most relevant items of users or items. Traditionally, the evaluation of recommender system quality has focused on the various predictive accuracy metrics of these. However, recommender system must be not only accurate but also useful to users. User satisfaction with recommender systems as an evaluation criterion of recommender system is related not only to how accurately the system recommends but also to how much it supports the user's decision making. In particular, highly serendipitous recommendation would help a user to find a surprising and interesting item. Serendipity in this study is defined as a measure of the extent to which the recommended items are both attractive and surprising to the users. Therefore, this paper proposes an application of serendipity measure to recommender systems to evaluate the performance of recommender systems in terms of recommendation system quality. In this study we define relevant or attractive unexpectedness as serendipity measure for assessing recommendation systems. That is, serendipity measure is evaluated as the measure indicating how the recommender system can find unexpected and useful items for users. Our experimental results show that highly serendipitous recommendation such as item-based collaborative filtering method has better performance than the other recommendations, i.e. user-based collaborative filtering method in terms of recommendation system quality.

**Key Words** : Recommendation system, Serendipity measure, Unexpectedness, Relevance, Quality

Received : June 11, 2019 Revised : November 21, 2019 Accepted : December 6, 2019
Publication Type : Regular Paper Corresponding Author : Taeksoo Shin

## 1. Introduction

In the age of big data or information overload, the volume of available product and user's digital information has exponentially increased. This kind of information in terms of recommendation systems are useful in providing the most relevant items with users. Therefore a lot of attention about personalized recommendation system has been increasing for selecting the suitable items for their personal taste and needs. A recommender system aims to provide personalized recommendations to users for specific items. Several techniques for developing the recommendation systems have been studied (Aggarwal et al., 2002; Smith and Linden, 2017; Su and Khoshgoftaar, 2009). They include collaborative filtering, content based filtering, hybrid filtering, and association rule or sequential pattern analysis and so on. Among them, the collaborative filtering technique has been widely used in high recommendation accuracy in Amazon.com, Netflix.com, MovieLens.com, and so on.

Most of these systems always provide the most relevant items of users or items. However, recent research trends also consider other factors besides accuracy that contributes towards the quality of recommendation for users' satisfaction (Zhang et al., 2011). According to the previous researches, the evaluation of recommender system quality has focused on the various predictive accuracy metrics of these. But the accuracy is not the only metric to evaluate recommender systems. There are another important aspects that we need to focus on in future evaluations of recommender systems. Accurate recommendations may not provide the most useful items to the users (Ge et al., 2010; Zhang et al., 2011). User satisfaction with recommender systems as an evaluation criterion of recommender system is related not only to how accurately the system recommends but also to how much it supports the user's decision making. Therefore, recommender system must be not only accurate but also useful to users because there are still some unsearched but unexpected items useful for users. The evaluation metrics such as serendipity measure for measuring user satisfaction with recommender systems move beyond accuracy metrics to evaluate recommender systems. Serendipity measure in our study is defined as a measure of the extent to which he recommended items are both attractive and surprising to the users (Gemmis et al., 2015; Herlocker et al., 2004; McNee et al., 2006). Highly serendipitous recommendation would help a user to find a surprising and interesting item (Murakami et al., 2007; Herlocker et al., 2004). Therefore, the purpose of this paper is to suggest a serendipity measure as a new method to evaluate the performance of two competitive collaborative filtering methods in terms of the quality of recommendation systems.

## 2. Several issues in Recommender System

Until now, recommender systems are a popularly accepted technology used by market leaders in several industries (e.g., by Amazon, Netflix and Pandora, etc.). Recommender systems apply statistical and knowledge discovery techniques to the problem of making product recommendations based on previously recorded data (Sarwar et al., 2000; Hahsler, 2014). Such recommendations can help the customer to buy products they want to buy faster by promoting cross-selling by suggesting additional products, and then can improve customer loyalty through creating the value-added relationship (Schafer et al., 2001).

Recommender systems can be mainly categorized into content-based approach and collaborative filtering methods. Content-based approaches are based on the idea that if we can elicit the preference structure of a customer (user) concerning product (item) attributes then we can recommend items which rank high for the user's most desirable attributes (Hahsler, 2014). Collaborative filtering uses given rating data by many users for many items as the basis for predicting missing ratings and/or for creating a top-N recommendation list for a given user, called the active user. Collaborative filtering algorithm suggests new items or to predict the utility of a certain item for a particular user based on the user's previous likings and the opinions of other like-minded users (Sarwar et al., 2001). Collaborative filtering algorithm has are two main kinds of collaborative filtering recommender systems such as item-based and the user-based collaborative filtering algorithms (Surti, 2011).

However, recommender systems have several issues such as cold start, sparsity, the gray sheep users who has low correlation with almost users, and serendipity of recommendation (Huang et al., 2004; Adomavicius & Tuzhilin 2005). In order to solve these problems, several studies proposed several methods as follows.

First of all, many researchers have attempted to alleviate the sparsity problem. Sarwar et al. (2001) proposed an item-based approach to addressing both the scalability and sparsity problems. Furthermore, for the cold-start problem considered as a special instance of the sparsity problems, many methods have been proposed to deal with the cold-start problem of new or rare users or items (Adomavicius and Tuzhilin, 2005; Bernardi et al., 2015; Chen et al., 2011).

Kim and Im (2014) proposed a new algorithm that utilizes social network analysis techniques to resolve the gray sheep problem. They utilized degree centrality in SNA to identify gray sheep users with unique preference. Degree centrality in social network analysis refers to the number of direct links to and from a node. In a network of users connected through common preferences or tastes, those with unique tastes have fewer links to other users and they are isolated from other users. Therefore gray sheep can be identified by calculating degree centrality of each node. Ghazanfar and Prugel-Bennet (2014) also proposed a new method to solve gray-sheep user problem in recommendation system. Gray sheep users problem is responsible for the increased error rate in collaborative filtering based recommender systems. They identified the gray-sheep users by using the clustering algorithms and offered a hybrid recommendation algorithm to make reliable recommendations for gray-sheep users. They claimed that the presence of a large number of gray-sheep users might significantly affect the recommendations quality.

In addition to the above issues, a few recent studies argued that accuracy is not the only metric for evaluating recommender systems and that there are other important aspects we need to focus on in future evaluations (Herlocker et al., 2004; McNee et al., 2006; Ge et al., 2010). Ultimately, recommender system should provide personalized recommendations so as to improve users' satisfaction (Ge et al., 2010). Furthermore, several researchers began to consider serendipity problem in the context of

〈Table 1〉 Previous Research on Serendipity Measures for Recommendation Systems

| Author | Content | Serendipity measure |
|---|---|---|
| Murakami et al. (2007) | They proposed metrics, i.e. unexpectedness and unexpectedness r for measuring the serendipity of recommendation lists produced by recommender systems. Basic idea of proposed metrics is that unexpectedness is considered to be the deviation from the result obtained from primitive prediction method (PPM). | $\text{unexpectedness} = \frac{1}{N}\sum_{i=1}^{N}\max(\Pr(s_i) - Prim(s_i), 0) * isrel(s_i)$ <br> $\text{unexpectedness r} = \frac{1}{N}\sum_{i=1}^{N}\max(\Pr(s_i) - Prim(s_i), 0) * isrel(s_i) * count(i)$ |
| Gemmis et al. (2015) | Serendipity is defined as relevant and unexpectedness items of the recommended item list. Unexpectedness is evaluated by popularity and average rating. Popularity is number of users who rated on item. Unexpected items are defined as those lower than average rating and popular items. Relevant items of user are those greater than the average value of all ratings provided by the user. | $\text{Reference@N} = \frac{\sum_{i \in L} R(i)}{N}$ <br> where $R(i) = 1$ $if\ is\ relevant; 0\ is\ otherwise$ <br> $\text{unexpectedness@N} = \frac{\sum_{i \in L} U(i)}{N}$ <br> where $U(i) = 1$ $if\ is\ unexpected; 0\ is\ otherwise$ <br> $\text{Serendipity@N} = \frac{\sum_{i \in L} s(i)}{N}$ <br> where $S(i) = 1$ $if\ is\ serendipitous,\ 0\ is\ otherwise$ |
| Chiu et al. (2011) | Serendipitous items are defined as "items which are not yet accessed or rated, but the target user's friends have accessed before". They propose a social network-based serendipity (SNS) recommender system that uses interactive information from the social network to find out which items are interesting for users but hard to discover by themselves. | $Sv = \left(\frac{\sum_{i=1}^{k} Fp'_i}{k}\right) * \alpha + \left(\frac{\sum_{j=1}^{m} Ry'_j}{k}\right) * \beta + \left(\frac{\sum_{i=1}^{k} Fs'_i}{k}\right) * \gamma 1 + (Ff') * \gamma 2 + (Fd') * \gamma 3$ <br> $\text{SRDP(Serendipity)} = \frac{\text{(Useful Items of Unexpected Recommendations)}}{\text{Total Unexpected Recommendations}}$ |
| Ge et al. (2010) | Serendipity item should be not yet discovered and not be expected by the user and the item should also be interesting, relevant and useful to the user. The serendipity metric catches the essential aspects of serendipity, unexpectedness and usefulness. | $SRDP(Serendipity) = \frac{\sum_{i=1}^{N} u(RS_i)}{N}$ <br> $UNEXP = RS \backslash PM$ <br> *RS is the recommendations generated by a recommender system.* <br> *PM is a set of recommendation generated by a primitive prediction model(PPM).* <br> $RS_i$ *is an element in UNEXP.* <br> $u(RS_i) = 1$ *means that* $RS_i$ *is useful to the user.* <br> $u(RS_i) = 0$ *means that* $RS_i$ *is useless to the user.* |
| Zhang et al. (2011) | Serendipity represents the unusualness or surprise of recommendations. Unlike novelty, serendipity encompasses the semantic content of items, and canbe imagined as the distance between recommended items and their expected contents. | $\overline{Unserendipity} = \sum_{u \in S} \frac{1}{\|S\|\|H_u\|} \sum_{h \in H_u} \sum_{i \in R_{u},20} \frac{CosSim(i,h)}{20},$ <br> *S is set of all users,* <br> $R_{u,n}$ *is the top n recommended items for user u.* <br> $H_u$ *is non-withheld item history of user u.* |
| Adamopoulos and Tuzhilin, (2014) | They proposed a method to improve user satisfaction by generating unexpected recommendations based on the utility theory of economics. Unexpectedness is defined as deviation from the expectations of users. | $UNEXPECTED = \sum_{u} \frac{\|RS_u \setminus E_u\|}{\|N\|}$ <br> $RS_u$ *is the generated recommendation list* <br> $E_u$ *is the set of expected movies for a user.* |

recommender systems (Murakimi et al., 2007; Chiu et al., 2011; Sridharan, 2014; Gemmis et al., 2015). Therefore, this study focuses on serendipity of recommendation considered as a critical quality issue in recommendation systems. Serendipity is concerned with the novelty of recommendations and in how far recommendations may positively surprise users (Ge et al., 2010). In recommender systems, it is defined as a measure that indicates how the recommender system can find unexpected and useful items for users (Sridharan, 2014). Table 1 summarizes prior studies about several serendipity approaches applied to recommendation systems.

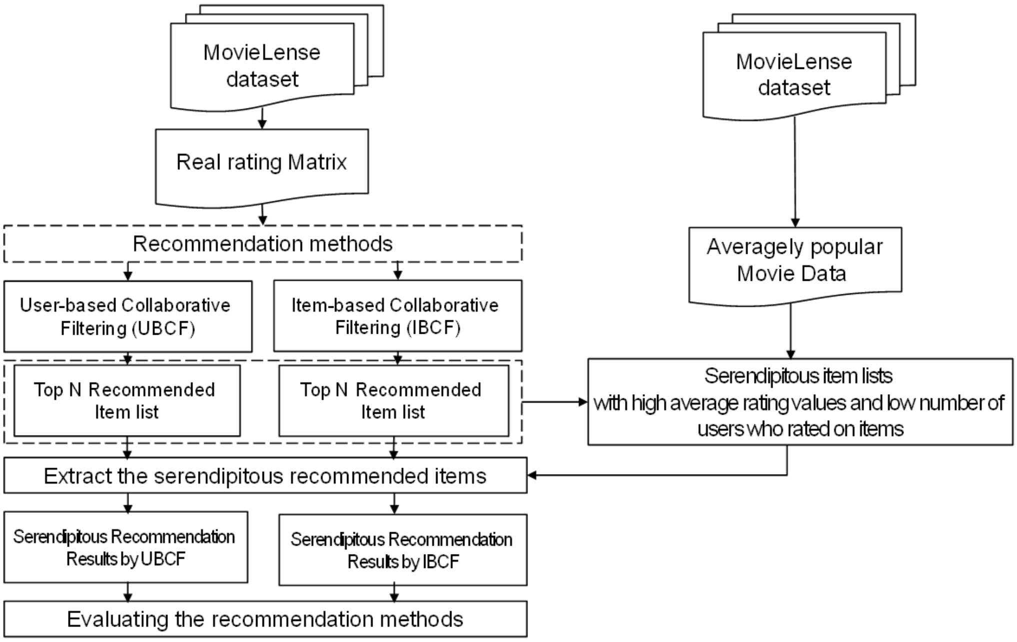
# 3. Research methodology

## 3.1 Evaluating the Quality of Recommendation System by Using Serendipity Measure

Serendipity is the most closely related concept to unexpectedness and both of them involve a positive emotional response of the user about a recommended item (Adamopoulos and Tuzhilin, 2014). Serendipity is concerned with the novelty of recommendations and it characterized by interestingness of items and the surprise for users who get unexpected suggestions (Sridharan, 2014; Gemmis et al., 2015). Murakami et al. (2007) proposed a metric for evaluating unexpectedness in a recommendation item list selected from overall items in database. According to their research, unexpectedness is defined as the deviation from obtained result of the other benchmarking model, i.e. primitive prediction model (PPM) or primitive recommendation method based on average ratings and number of users for each item (Murakami et al., 2007; Sridharan, 2014; Gemmis et al., 2015). In other words, the recommendation list consists of items with a high degree of expectedness. Adamopoulos and Tuzhilin (2014) proposed evaluation method of recommendation system based on unexpectedness and utility measure. They defined unexpectedness as departing from the users' expectations. Gemmis et al. (2015) proposed also unexpectedness measure for the evaluation of recommendation system. They adopted two criteria, i.e. a popularity and an average rating of a recommended item for defining unexpectedness. Popularity of the item is defined as the numbers of users who rated. They defined unexpected items as items with lower rating than average rating and lower popularity. However, this study defines unexpected items as items with only lower popularity. Previous researches related with serendipity measures are showed in Table 1.

This paper proposes a serendipity measure to recommender systems to evaluate the performance of recommender systems in terms of the quality of recommendation systems as shown in Figure 1.

In this study, serendipity of recommendation is evaluated by the relevance and unexpectedness measures similar to those proposed by the Gemmis et al. (2015). We also use user-based collaborative filtering and item-based collaborative filtering methods as recommendation models for




〈Figure 1〉 Research model

comparative experimental analysis. Relevance in the recommendation can be modeled as a binary concept: either an item is liked by a user or not (Gemmis et al., 2015). In this study, an item i is defined as relevant to a user if the average rating value on i is greater than the average rating value of total items provided by all the users. A relevant item in recommendation list is calculated by Equation (1) defined such that R(i) is 1 if i is relevant and 0 otherwise. L is recommendation list and N is total number of recommended items.

$$Relevance = P(relevant\ \ or\ \ attractive\ \ items|L) = \frac{\sum_{i \in L} R(i)}{N} \quad (1)$$

An unexpected item in recommendation list is calculated by Equation (2) defined such that U(i) is 1 if i is unexpected and 0 otherwise. Unexpected item is assumed to have lower popularity in this study.

$$Unexpectedness = P(unexpected\ \ items|L) = \frac{\sum_{i \in L} U(i)}{N} \quad (2)$$

Serendipity means how many relevant and unexpected items exist in overall recommendation items from traditional recommendation methods such as user-based collaborative filtering method and item-based collaborative filtering. Serendipity measure is calculated by the Equation (3). S(i) is a serendipity value of a recommended item. S(i) is defined as R(i)*U(i). N is total number of recommended items (Gemmis et al., 2015).

$$Serendipity = P(Relevance, Unexpected\ items|L) = \frac{\sum_{i \in L} S(i)}{N} = \frac{\sum_{i \in L} R(i) * U(i)}{N} \quad (3)$$

Serendipitous items are those relevant and unexpected at the same time. Relevant items are defined as those whose ratings are greater than the average rating value of all the users. Serendipity is defined as Equation (3).

### 3.2 Collaborative filtering algorithms

The goal of a collaborative filtering algorithm is to suggest new items or to predict the utility of a certain item for a particular user-based on the user's previous likings and the opinions of other like-minded users (Sarwar et al., 2001). There are two kinds of collaborative filtering recommender systems such as item-based and user-based collaborative filtering algorithms (Surti, 2011).

User-based collaborative filtering recommends to a user the items that are the most preferred by similar users. User-based collaborative filtering method is based on rating data from each users. The assumption is that users with similar preferences will rate items similarly. User-based collaborative filtering method finds another users whose past rating behavior is similar to that of current user and use their ratings on other items to predict what the current user will like (Hahsler, 2014; Sridharan, 2014). Fundamental ingredients of the user-based collaborative filtering method are (1) rating matrix that specifies the item, user, and rating, (2) a similarity function between users, and (3) a recommendation method for user-based similarities and rating information (Sridharan, 2014).

This user-based collaborative filtering is popular, but it suffers from scalability issues related with the frequent computation of similarity between users. That is, when a user changes the rating of times frequently, the rating vector of such a user changes which modifies the similarity with others. So the user neighborhood N for a given user cannot be precomputed but has to be evaluated whenever recommendations are needed. It can be a big computational bottleneck for large datasets (Sridharan, 2014).

Item-based collaborative filtering recommends to a user the items that are most similar to the user's purchases. Item-based collaborative filtering algorithm assumes similarities between each item that was rated from the same users in data sets (Linden at al., 2003; Surti, 2011). In other words if two items tend to have the preferences from the same users, then they are similar items (Sridharan, 2014). Item-based collaborative filtering is more efficient than user-based collaborative filtering since the model (reduced similarity matrix) is relatively small and can be fully precomputed (Hahsler, 2014). So item-based collaborative filtering is successfully applied in large scale recommender systems (e.g., by Amazon.com, Neflix.com, etc.). However, Item-based collaborative filtering is known to be of lower-quality than user-based collaborative filtering (Deshpande & Karypis, 2004; Hahsler, 2014).

## 4. Experimental Results

We used MovieLens dataset of University of Minnesota for our experimental analysis. These data sets contain 943 users, 1,682 movies, and 100,000 ratings on an integer scale 1 to 5. In our data sets, average rating is 3.25 and averaged rating count value of each movie is 31. In this paper, we used the serendipity measure similar to it proposed by the Gemmis et al. (2015) for the evaluation of serendipitous recommendation and applied this measure to user-based collaborative filtering and item-based collaborative filtering as recommendation models for comparative experimental analysis. Table 2 and Table 3 show the number of the serendipitous items and serendipity values in each recommendation systems. The result shows that item-based collaborative filtering method recommends more serendipitous items than user-based method. In F-measure perspective, user-based collaborative filtering methods more accurately recommended items than item-based collaborative filtering methods for the users without considering the serendipity value. This study uses F-measure as recommendation accuracy measure. F-measure is used to measure the precision of the whole recommendation and the recall of the whole recommendations. F-measure is evaluated by combining the values of precision and recall. The larger F-measure means the more useful items that the recommendation system can recommend (Chiu et al., 2011). That is, precision is defined as (highly-rated items recommended)/(total highly-rated items), recall as (highly-rated items recommended)/(total items recommended), and F-measure as 2*(precision*recall)/(precision+recall). Figure 2 shows the relationship between serendipity value and F-measure as recommendation accuracy. Figure 3 and 4 show the relationship between the number of Top N recommendation items and serendipity, and the relationship between the number of Top N recommendation items and F-Measure. First of all, Figure 2 shows that the relationship between serendipity value and F-measure is positive only in item-based collaborative filtering method. As shown in Figure 3, the relationship between the number of Top N recommendation items and serendipity exists slightly only in item-based collaborative filtering method. Item-based collaborative filtering also has higher serendipity value than user-based collaborative filtering in all the number of Top N recommendation items. Finally Figure 4 shows that the relationship between the number of Top N recommendation items and F-measure are positive and user-based collaborative filtering has higher performance than item-based collaborative filtering in terms of recommendation accuracy (i.e. F-measure).

In summary, these results have the following implications. First, the serendipity value has a positive influence on recommendation accuracy such as F-measure particularly in item-based collaborative filtering method, but user-based collaborative filtering method doesn't have any relationship between the serendipity value and the F-measure. Second, the number of Top N recommendation items has a positive influence on

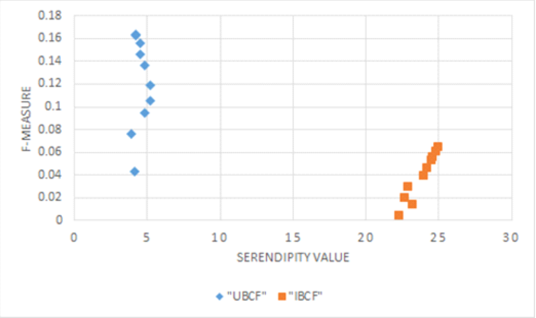
〈Table 2〉 Evaluation Result of User-based Collaborative Filtering Method

| # of Top recommended items for each user | # of serendipitous items | Serendipity value (%) | Precision | Recall | Accuracy | F |
|---|---|---|---|---|---|---|
| 1 | 39 | 4.1 | 0.349 | 0.023 | 0.986 | 0.043 |
| 2 | 73 | 3.9 | 0.310 | 0.043 | 0.986 | 0.076 |
| 3 | 135 | 4.8 | 0.284 | 0.057 | 0.985 | 0.095 |
| 4 | 197 | 5.2 | 0.258 | 0.066 | 0.985 | 0.105 |
| 5 | 244 | 5.2 | 0.251 | 0.078 | 0.985 | 0.119 |
| 6 | 272 | 4.8 | 0.245 | 0.094 | 0.984 | 0.136 |
| 7 | 294 | 4.5 | 0.238 | 0.106 | 0.984 | 0.146 |
| 8 | 342 | 4.5 | 0.227 | 0.119 | 0.984 | 0.156 |
| 9 | 353 | 4.2 | 0.220 | 0.129 | 0.983 | 0.163 |
| 10 | 392 | 4.2 | 0.213 | 0.134 | 0.983 | 0.164 |

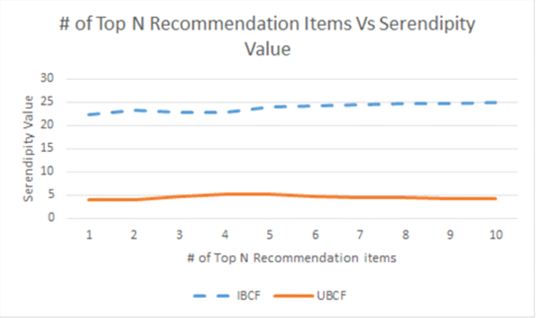
〈Table 3〉 Evaluation result of Item-based Collaborative Filtering Method

| # of Top recommended items for each user | # of serendipitous items | Serendipity value (%) | Precision | Recall | Accuracy | F |
|---|---|---|---|---|---|---|
| 1 | 210 | 22.3 | 0.084 | 0.004 | 0.986 | 0.004 |
| 2 | 437 | 23.2 | 0.123 | 0.014 | 0.986 | 0.014 |
| 3 | 641 | 22.7 | 0.116 | 0.020 | 0.985 | 0.020 |
| 4 | 879 | 22.9 | 0.118 | 0.030 | 0.985 | 0.030 |
| 5 | 1130 | 24.0 | 0.117 | 0.039 | 0.985 | 0.039 |
| 6 | 1372 | 24.2 | 0.117 | 0.046 | 0.984 | 0.046 |
| 7 | 1614 | 24.5 | 0.112 | 0.053 | 0.984 | 0.053 |
| 8 | 1856 | 24.6 | 0.107 | 0.056 | 0.983 | 0.056 |
| 9 | 2104 | 24.8 | 0.108 | 0.061 | 0.983 | 0.061 |
| 10 | 2359 | 25 | 0.106 | 0.065 | 0.982 | 0.065 |

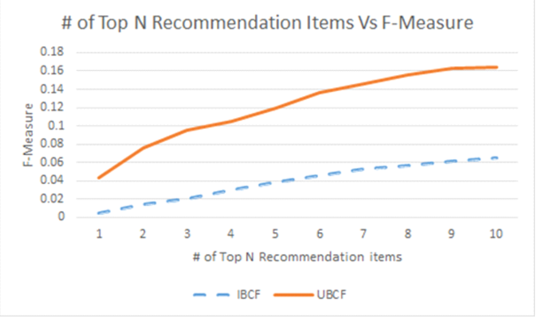
serendipity value and F-measure only in item-based collaborative filtering method. That is, the more the number of Top N recommendation items is, the higher serendipity value is in item-based collaborative filtering method. The higher serendipity value is, the higher F-measure is. Third, user-based collaborative filtering has better performance than item-based collaborative filtering in terms of recommendation accuracy (i.e. F-measure), but on the other hand, item-based collaborative filtering has better performance than user-based collaborative filtering in terms of serendipity value.




〈Figure 2〉 Serendipity Value vs. F-Measure




〈Figure 3〉 # of Top N Recommendation items vs. Serendipity Value




〈Figure 4〉 # of Top N Recommendation Items vs. F-Measure

## 5. Conclusion

This study defined serendipity measure for evaluating recommendation systems in serendipity perspective and reviewed related studies. Serendipity is defined by the measure indicating how the recommender system can find unexpected and relevant (or useful) items for users. We calculated unexpectedness and relevance (or usefulness) on recommendation list similar to the way presented by the Murakami et al. (2007) and Gemmis et al. (2015). That is, relevant items are defined as items with higher rating values and unexpected items are defined as items with low popularity in whole recommended items. If there are more relevant and unexpected items in recommendation list, the recommendation system will recommend more serendipitous items. According to our experimental results, the item-based collaborative filtering method recommended more serendipitous items than user-based collaborative filtering method for the users. It has an implication that the item-based recommendation method can be more effective than user-based method in recommending useful and novelty items to users in serendipity perspective. In addition, we found that the number of Top N recommendation items is positively related to serendipity value, and serendipity value is also partially related to F-measure only in item-based collaborative filtering method.

One of our interesting results from our experimental analysis is that item-based collaborative filtering method is better than user-based collaborative filtering method in terms of recommendation quality because item-based collaborative filtering method has higher serendipity value than user-based collaborative filtering method. That is, item-based collaborative filtering method is more likely to recommend relevant and unexpected, i.e. serendipitous items in comparison with used-based collaborative filtering method.

Serendipity value in this study is a measure of the extent to which recommended items are both attractive and surprising to users. It also provides a higher level of recommendation satisfaction. Therefore, it has implications that item-based collaborative filtering method with higher serendipity value than user-based collaborative filtering method can be more attractive to contents service providers or commerce companies such as Netflix, YouTube, Amazon.com, etc. Furthermore, user satisfaction on recommendation systems recommending to users for only serendipitous items among top N recommendation lists of each collaborative filtering method will be higher than that recommending for both serendipitous and non-serendipitous items. This recommendation is a kind of conditional or selective recommendation approach. Therefore, in viewpoint of customers, they are more likely to respond a promoted serendipitous items recommended by contents service providers or commerce companies through the selective recommendation approach.

Finally, the limitations and further research issues of this study are as follows. First, this study focused on only two popular collaborative filtering methods, but we need to consider alternative competitive recommendation methods including model-based collaborative filtering methods. Second, this study used serendipity value to analyze only posterior quality of recommendation systems of these two collaborative filtering methods, but we need to analyze fundamentally how to recommend only more attractive and unexpected or serendipitous items using these two collaborative filtering methods in order to enhance quality of recommendation systems. Third, we need to increase # of Top N recommendation lists and use statistical significance tests for getting significantly meaningful results of the relationship between serendipity measures and recommendation accuracy measures. Finally, the future research needs to develop more sophisticated serendipity measures for refining original serendipity measure into different serendipitous levels according to users' different preferences or their own tastes.

## References

Adamopoulos, P. and Tuzhilin, A., "On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected," *ACM Transactions on Intelligent Systems and Technology*, Vol.1, No.1(2014), 1~51.

Adomavicius, G. and Tuzhilin, A., "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No.6(2005), 734~749.

Aggarwal, C. C., Procopiuc, C., and Yu, P. S., "Finding Localized Associations in Market Basket Data," *IEEE Transactions on Knowledge* *and Data Engineering*, Vol.14, No.1(2002), 51~62.

Bernardi, L., Kamps, J., Kiseleva, J., and Müller, M., "The Continuous Cold Start Problem in e-Commerce Recommender Systems," *2nd Workshop on New Trends on Content-Based Recommender Systems*, (2015), 30~33.

Chen, Y., Wu, C., Xie, M. and Guo, X., "Solving the Sparsity Problem in Recommender Systems Using Association Retrieval," *Journal of Computers*, Vol.6, No.9(2011), 1896~1902.

Chiu, Y. S., Lin, K. H., and Chen, J. S., "A Social Network-based Serendipity Recommender System," *International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, (2011).

Deshpande, M. and Karypis, G., "Item-based top-N Recommendation Algorithms," *ACM Transactions on Information Systems*, Vol.22, No.1(2004), 143~177.

Ge, M., Delgado-Battenfeld, C., and Hannach, D., "Beyond Accuracy: Evaluating Recommender Systems by Coverage and Serendipity," *Proceeding of the fourth ACM conference on Recommender systems*, (2010), 257~260.

Gemmis, M. D., Lops, P., Semeraro, G., and Musto, C., "An Investigation on the Serendipity Problem in Recommendation System," *Information Processing and Management*, Vol.51(2015), 695~717.

Ghazanfar. M. A. and Prugel-Bennet, A., "Leveraging Clustering Approaches to Solve the Gray-Sheep Users Problem in Recommender Systems," *Expert Systems with Applications,* Vol.41(2014), 3261~3275.

Hahsler, M., "recommenderlab: A Framework for Developing and Testing Recommendation Algorithms," *Comprehensive R Archive Network*, 2014. Available at http://cran.r-project.org/web/packages/recommenderlab/vignettes/recommenderlab.pdf.

Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T, "Evaluating Collaborative Filtering Recommender Systems," *ACM Transactions on Information Systems*, Vol.22, No.1(2004), 5~53.

Huang, Z., Zeng, D., and Chen, H., "A Link Analysis Approach to Recommendation under Sparse Data," *Proceedings of Americas Conference on Information Systems*, 2004.

Kim, M. S. and Im, I., "Resolving the Gray Sheep Problem Using Social Network Analysis (SNA) in Collaborative Filtering (CF) Recommender Systems," *Journal of Intelligent Information System*, Vol.20(2014), 137~148.

Linden, G., Smith, B., and York, J., "Amazon.com Recommendation," *IEEE Internet Computing*, Vol.7, No.1(2003), 76~80.

McNee, S. N., Riedl. J., and Konstan, J. A., "Being Accurate is Not Enough: How Accuracy Metrics have hurt Recommender Systems," *Extended Abstracts on Human factors in Computing Systems, CHI(06)*, (2006), 1097~1101.

Murakami, T., Mori, K., and Orihara, R., "Metrics for Evaluating the Serendipity of Recommendation Lists," *Proceedings of the Conference on New Frontiers in Artificial Intelligence,* (2007), 40~46.

Sarwar, B., Karyps, G., Konstan, J., and Reidl, J, "Analysis of Recommendation Algorithms for e-Commerce," *Proceedings of the 2nd ACM conference on Electronic Commerce,* (2000), 158~167.

Sarwar, B., Karyps, G., Konstan, J., and Reidl, J, "Item-based Collaborative Filtering Recommendation Algorithms," *Proceedings of the 10th international conference on World Wide Web. ACM, New York, NY, USA*, (2001), 285~295.

Schafer, J.B., Konstan, J.A., and Riedl, J., "E-Commerce Recommendation Applications," *Data Mining and Knowledge Discovery*, Vol.5(1/2)(2001), 115~153.

Smith, B. and Linden, G., "Two Decades of Recommender Systems at Amazon.com," *IEEE Internet Computing*, Vol.21, No.3(2017), 12~18.

Sridharan, S., "Introducing Serendipity in Recommender Systems through Collaborative Methods," *Master of Science Thesis, University of Rhode Island,* 2014.

Su, X. and Khoshgoftaar, T. M., "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence,* Vol.2009(2009), 1~19.

Surti, T, "Social Recommender Systems: Improving Recommendations through Personalization," *Computer Science Department, Haverford College*, 2011.

Zhang, Y. C., Seaghdha, D. O., Quercia, D., and Jambor, T., "Auralist: Introducing Serendipity into Music Recommendation," *Research Note*, 2011.

국문요약

# 세렌디피티 지표를 이용한 추천시스템의 품질 평가

체렌돌람* · 신택수**

최근 추천시스템의 품질평가 관점에서 이에 대한 다양한 연구들이 진행되고 있다. 추천시스템은 기본적으로 사용자들에게 특정 아이템에 대한 개인화된 추천을 제공하는데 목적이 있으며, 대부분의 추천시스템은 항상 사용자 또는 아이템과 가장 관련 있는 아이템을 추천한다. 그리고 이러한 추천시스템의 성과는 전통적으로 다양한 예측정확도 등에 초점을 두어 왔다. 그러나, 추천시스템은 예측가능성 차원에서 정확해야 할 뿐만 아니라 사용자들에게 유용해야 한다. 특히 최근의 추천시스템에 대한 연구로서, 추천시스템의 평가기준에 속하는, 추천시스템에 대한 사용자 만족도(품질)는 추천시스템이 얼마나 정확하게 추천하느냐 뿐만 아니라 사용자의 의사결정에 얼마나 충분히 도움이 되는지와 관계가 깊다. 예를 들어, 특히 높은 수준의 세렌디티피한 추천은 사용자들이 뜻밖의 아이템이면서 흥미로운 아이템을 찾는데 도움이 된다. 여기서, 세렌디피티란 추천 아이템이 사용자에게 매력적인 동시에 뜻밖의(비기대성의) 아이템인 정도를 의미한다.

본 연구는 추천시스템의 성과를 나타내는 세렌디피티 지표를 추천시스템에 적용하여 추천시스템의 품질을 평가하는 것을 목표로 한다. 본 연구에서는 세렌디피티 지표는 관련성(매력)이 있는 동시에 뜻밖인(비기대성의) 아이템을 추천하는 정도로 정의하고, 이 세렌디피티 지표를 측정하기 위해, 추천시스템이 사용자들에게 예상치 못한 유용한 아이템을 찾을 수(또는 추천할 수) 있는 정도를 평가하였다. 본 연구의 주요 실증분석결과로는, 아이템기반 협력 필터링 기법이 사용자기반 협력 필터링 기법보다 더 높은 세렌디피티값을 가지며, 따라서, 추천시스템의 품질평가 차원에서 아이템기반 협력 필터링 기법은 사용자기반 협력 필터링 기법보다는 더 좋은 추천 품질을 갖고 있음을 보여 주었다.

**주제어** : 추천시스템, 세렌디피티 지표, 비기대성, 관련성, 품질

논문접수일 : 2019년 6월 11일　　논문수정일 : 2019년 11월 21일　　게재확정일 : 2019년 12월 6일
원고유형 : 일반논문　　　교신저자 : 신택수

* 연세대학교 미래캠퍼스 원주대학혁신지원사업단 교육전문연구원
** 교신저자 : 신택수
연세대학교 정경대학 경영학부 교수
26493 강원도 원주시 연세대길 1
Tel: +82-33-760-2335, Fax: +82-33-763-4324, E-mail: tsshin@yonsei.ac.kr

# 저 자 소 개

**체렌돌람**

현재 연세대학교 미래캠퍼스 원주대학혁신지원사업단의 교육전문연구원으로 재직하고 있다. 몽골 후레정보통신대학교에서 경영학 학사 학위를 받고, 2018년 연세대학교에서 경영학 박사학위를 받았다. 주요 관심분야는 빅데이터 분석, 소셜네트워크 분석, 추천시스템, 인공지능을 이용한 빅데이터 분석 등이다.

**신택수**

현재 연세대학교 정경대학 경영학부 교수로 재직하고 있다. 연세대학교 경영학과에서 학사 및 석사학위를 받고, KAIST에서 경영정보시스템으로 경영공학 박사학위를 받았다. 주요 관심분야는 빅데이터 분석, 소셜미디어 분석, 인공지능을 이용한 고객추천시스템 등이다.