IIAI Open Conference Publication Series
IIAI Letters on Informatics and Interdisciplinary Research
Vol.006, LIIR329
DOI: https://doi.org/10.52731/liir.v006.329

Optimization of Recommendation System by Improving Serendipity and Grouping Users Based on Their Number of Data Points

Haruto Domoto *, Tetsuya Nishibe *,
Takahiro Uchiya *, Ichi Takumi *

Abstract

Information recommendation systems aim to deliver optimal content to users, but conventional methods often only suggest similar items, leading to user boredom and reduced recommendation effectiveness. This study addresses this limitation by focusing on "serendipity", enhancing the unexpectedness of recommendations. We compared conventional methods, existing techniques, and six newly proposed approaches. Users were grouped into three categories based on the number of data points they evaluated to analyze the impact on recommendation performance. For users with fewer data points, the best approach was to recommend items significantly different from the average user preferences. For users with more data points, recommending items that other users disliked but held high value for the target user was most effective. These strategies improved diversity and unexpectedness without sacrificing usefulness, thereby successfully enhancing serendipity. This method shows promise in increasing user satisfaction by providing a more engaging recommendation experience.

Keywords: Movie Recommendations, Recommendation Systems, Serendipity, User-based Collaborative Filtering

1 Introduction

In recent years, with the rapid spread of the Internet, the number of digital contents available online has increased rapidly, exacerbating the problem of information overload. As a result, useful information for users tends to be buried in the network, making it difficult for them to find it easily [1][2]. Against this background, information recommendation systems [3][4] such as collaborative filtering [5] and content-based recommendation [6] are used to provide appropriate information to each user.

However, conventional systems have a tendency to recommend only items that are similar to items that users already know, and this problem causes many challenges. As a result, the discovery of novel and diverse content is limited. This phenomenon, known as the filter bubble effect [7], has a significant negative impact on the user experience because it reduces the diversity of recommendations and can lead to user disengagement [8].

To address this problem, researchers have begun to introduce the concept of serendipity [9] in recommendation systems. Serendipity is not just about the accuracy of the recommendation, but also provides an element of pleasant surprise or unexpectedness for the user. This perspective is

* Nagoya Institute of Technology, Nagoya, Japan

very important because it not only improves user satisfaction, but also triggers users to explore new content that would have been difficult to find using traditional methods [10]. The challenge here is to balance the trade-off between recommendation accuracy and serendipity. By providing users with unexpected yet relevant recommendations, our method is expected to open up new possibilities in the field of recommendation systems [11].

In this study, we focus on serendipity and introduce six new methods designed to improve serendipity without degrading the accuracy of recommendations. The effectiveness of the proposed method is then confirmed by evaluating the performance of each method from the perspective of the three evaluation metrics of serendipity: unexpectedness, diversity, and usefulness, while taking into account the impact of differences in the number of data available for each user.

This study made academic contributions in the following three ways: (1) it proposed a new method to improve serendipity, (2) the number of available user data has a significant impact on both unexpectedness and usefulness, and (3) the optimal method to improve serendipity for each number of available user data was identified.

In this paper, we first present related research in Section 2. Next, the proposed method is presented in Section 3. Furthermore, experimental results and analysis are presented in Section 4. Finally, conclusions are presented in Section 5.

2 Related Work

This section presents two papers related to improving serendipity in recommendation systems.

2.1 Domoto et al.'s Approach [12]

Through the development of a serendipity-enhancing recommendation system, Domoto et al. sought to eliminate user boredom caused by the continuous recommendation of similar items. They employed a method that utilizes the popularity ranking of items to improve serendipity. Specifically, they tabulated the number of times each item was rated by all users and created a popularity ranking based on that total. Items with a low number of ratings, i.e., items that are not popular, are unlikely to be selected by the conventional recommendation algorithm. These items are thought to contain many items that are highly valuable as chance discoveries, i.e., items with high serendipity. Therefore, they attempted to suppress excessive recommendation of popular items (popularity bias) and improve serendipity by weighting items with low evaluation frequency and recommending them preferentially.

The flow of their experiment is shown in Figure. 1. First, using the user's evaluation data, the predictive evaluation values were calculated using the conventional recommendation method of user-based collaborative filtering. We assigned weights to these predictive evaluation values, and then recommended items in the order of highest weighted predictive evaluation value. In doing so, we set multiple maximum values for the weights and made comparisons. The results showed two important results: (1) serendipity increased when the impact of the weights was equal to or greater than that of the conventional recommendation method, and (2) recommendation accuracy decreased when the maximum value of the weights was increased.

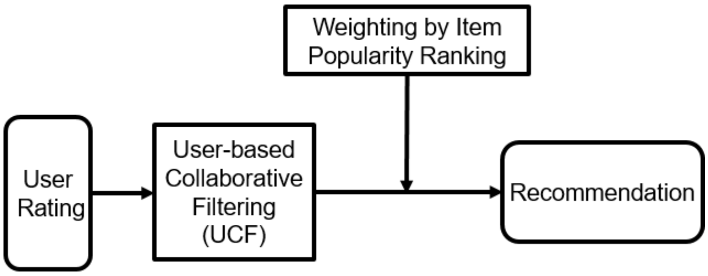

Figure 1: The Flow of Domoto's System

Their results indicate that serendipity can be improved by weighting items with fewer evaluations. However, in their experiments, the comparison is limited to conventional user-based collaborative filtering only, which is not a comprehensive comparison. In addition, only diversity is used as a measure of serendipity, and it is questionable whether it accurately evaluates serendipity. Furthermore, although the genres of recommended items are checked in order to evaluate diversity, quantitative evaluation is not conducted by simply visually checking the graphs.

2.2 Ren et al.'s Approach [13]

Ren et al. focus on the problem of popularity bias in recommendation systems, noting that this bias often limits the availability of items to users and may result in reduced serendipity. They proposed a method to mitigate popularity bias by focusing on unbalanced interactions and utilizing a gradient approach. Specifically, their method increases the chances that less popular items will be recommended by reweighting the loss function during the training process of the recommendation model. The effectiveness of their method was demonstrated by experiments that showed improved recommendation performance and reduced popularity bias. The experimental flow of Ren et al.'s method is as follows:

1. **Popularity Bias Identification**: First, the interaction data are analyzed to identify the popularity bias of the items. This involves evaluating the distribution of interactions per item and identifying which items are unequally supported.

2. **Re-weighting Loss Functions**: Re-weight the loss functions used to train the recommendation model according to the popularity of the item. In particular, for less popular items, a higher weight is set to increase the probability of being recommended.

3. **Training of recommendation model**: train the recommendation model using an adjusted loss function that incorporates the reweighted item interactions.

As a result, their proposed method effectively mitigated popularity bias and improved the balance of item exposure, thereby achieving an improvement in overall recommendation performance. However, while their method focuses on reducing popularity bias and improving recommendation performance, it does not actively enhance serendipity through measures such as diversity and novelty.

In this study, we propose six new methods for improving serendipity, and compare and evaluate seven methods, including a method for mitigating popularity bias, with the conventional method of user-based collaborative filtering. In related research, the number of times each item was rated is taken into account, but the number of times a user rated an item is not taken into account. Therefore, we group users according to the number of times they rated items, and check the differences in the results for each group.

3 Methodology

In this section, we propose a new method that aims to improve serendipity while maintaining recommendation accuracy. The flow of the proposed method is shown in Figure. 2. First, using the conventional method User-based Collaborative Filtering, the predicted evaluation values of items that have not yet been evaluated by the target user are calculated from the actual values of the target user. Then, at the "Change Point" in Figure. 2, in order to improve serendipity, the recommendation content is changed using a total of seven methods, including one existing method and six methods proposed in this study, and the top 100 items are recommended to the user.

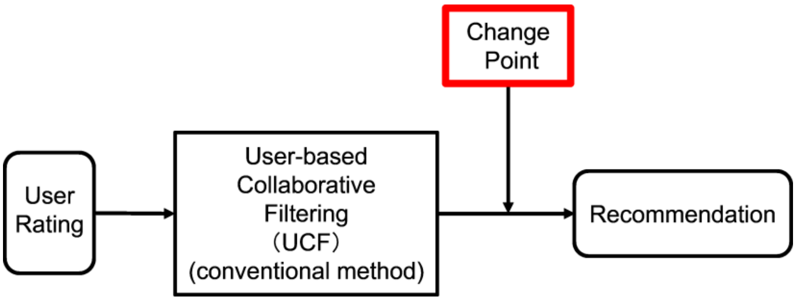

Figure 2: System Flow

3.1 Dataset

The dataset used is Movielens [14], which contains data on user ratings for movies in 18 different genres (e.g., action, comedy, etc.) It contains 610 users and about 9,700 movies, with each user rating a minimum of 20 and a maximum of 2,698 movies. Rating scores range from 0.5 to 5, with users rating in 0.5 increments within this range.

There is a problem called the cold start problem [15][16] in recommendation systems, where it is not possible to accurately predict user preferences from existing data. This problem refers to the inability of the system to make efficient recommendations to cold users (or new users) who have not rated any or only a very few items. This usually occurs when new users are added to the system or when new items (products) are added to the database. To check the impact of this problem on the proposed method, we conduct an experiment by dividing users into three groups based on the number of times they rated a movie, as shown in Table. 1.

Table 1: User Grouping

Group Name	Minimum Data Volume	Maximum Data Volume	Number of Users
Group A	20	913	592
Group B	914	1806	13
Group C	1807	2698	5

3.2 User-based Collaborative Filtering (UCF) [5]

User-based Collaborative Filtering (UCF) is a conventional method of finding other users with similar preferences to the target user and recommending items preferred by those users. Users with similar preferences are identified by calculating similarity using the Pearson correlation coefficient based on the users' item ratings. The system then recommends items that have not yet been evaluated by the target user based on the evaluation data of similar users. Since this method selects recommended items based on the user's evaluation data, it is possible to make recommendations that match the user's preferences. However, since it relies on data from users with similar preferences, it tends to result in the recommendation of only similar items. Therefore, this study proposes a new method to reduce recommendation bias.

3.3 Proposed Methods

The six methods are shown using the data in Table. 2. Measured values are the values that users rated for the movies they watched, as recorded in the data set, and predicted evaluation values are the values that UCF calculated for movies that users did not rate.

3.3.1 Existing method: Popularity Ranking [12]

We control for popularity bias by utilizing the popularity ranking used in previous studies [12]. This ranking is based on the number of times each item is rated. Since items with a low number of evaluations are unlikely to be recommended by conventional methods, there is a high possibility that items with low name recognition but high value exist. Therefore, by recommending such items, the popularity bias can be reduced and the problems of the existence of items with high value but low recognition and low market diversity can be solved. Therefore, the popularity ranking method weights these low valued items and recommends them preferentially. The weighting's maximum value is the one that increased serendipity the most in previous studies.

3.3.2 Proposal 1 : $max(pre\_target - \overline{pre\_all})$

In this method, items with a large difference between the target user's predicted evaluation value and the average of all users' predicted evaluation values are recommended. Items with high predictive evaluation values for the target user and low predictive evaluation values for other users are less likely to be recommended than items with both high predictive evaluation values for the target user and for other users. Therefore, this system aims to improve serendipity by preferentially recommending items with high predictive evaluation values for the target user and low predictive evaluation values for all users.

3.3.3 Proposal 2 : $sim(max(actual\_target - \overline{actual\_all}))$

In this method, items that are close to the item with a large difference between the target user's actual measured value and the average of all users' actual measured values are recommended.

Items with high actual measured values for the target user and low actual measured values for other users are less likely to be recommended than items with both high actual measured values

Table 2: Data Expression

DATA	Description
actual_all	Actual measured values for all users
actual_target	Measured values of target users
pre_all	Predicted evaluation values for all users
pre_target	Predicted evaluation values of the target user
count	Number of times each movie was rated by all users
Like	Movies for which the measured value is greater than 3.5

for the target user and high actual measured values for other users. Therefore, serendipity can be improved by preferentially recommending items with high actual measured values for the target user and low actual measured values for all users. However, the measured value is assigned to items that have already been viewed by the target user. Since we want to recommend items that the target user does not yet know, we recommend items that are similar to items with high actual value for the target user and low actual value for all users. The similarity of items is calculated using the Euclidean distance [17], which is the square root of the sum of the squares of the differences between the ratings of two users.

3.3.4 Proposal 3 : $count = 1 \land max(pre\_target)$

This method recommends items that have been evaluated few times by all users but have high predictive evaluation values for the target user. Normally, items with high predictive evaluation values but few evaluations by all users are rarely recommended. Therefore, by preferentially recommending such items, we aim to improve serendipity.

3.3.5 Proposal 4 : $sim(count = 1 \land max(actual\_target))$

This method recommends items that are similar to items that are rated few times by all users but have high actual values for the target user. Normally, items that have been evaluated few times by all users and have high actual measured values are rarely recommended. Therefore, by preferentially recommending such items, we aim to improve serendipity. Next, we want to recommend items that have not yet been evaluated, so we recommend items similar to these items. The method for calculating similarity is the same as in Proposal 2.

3.3.6 Proposal 5 : $Like < 50\% \land max(pre\_target)$

This method recommends items with a low Like rate but a high predictive evaluation values; items with a low Like rate, i.e., items that are not liked by many users, are less likely to be recommended, even if their predictive evaluation values is high. Therefore, we aim to improve serendipity by preferentially recommending these items.

3.3.7 Proposal 6 : $sim(Like < 50\% \land max(actual\_target))$

This is a method of recommending items similar to items with low Like ratios but high actual values; items with low Like ratios, i.e., items not liked by many users, are less likely to be recommended even if they have high actual values. Therefore, we aim to improve serendipity by preferentially recommending such items. And since we want to recommend items that have not yet been evaluated, we recommend items similar to these items. The method for calculating similarity is the same as in Proposal 2.

3.4 Evaluation Index

Serendipity can be classified into pseudo-serendipity, which is the accidental discovery of what the user is looking for, and true serendipity, which is the accidental discovery of what the user is not looking for [18]. In this study, we aim to solve the problem of recommending items that are too similar to what the user already knows by focusing on and evaluating "true serendipity", which is the accidental discovery of something that the user is not looking for.

Serendipity is commonly evaluated as a construct consisting of multiple elements [19]. In this experiment, we evaluate three elements that can be evaluated in this experiment: unexpectedness, diversity, and usefulness.

Unexpectedness is evaluated using the Equation. (1) based on the idea that the more items that are not easily recommended by conventional methods are recommended, the higher the unexpectedness [20]. S_i is the ith item from the top of the recommendation list. Unexpectedness exists only when the predictive evaluation value of the proposed method ($Pr(S_i)$) is higher than that of the conventional method ($Prim(S_i)$). Also, $count(i)$ is the number of items in the 1 to ith that are suitable for the user's preferences. Ensure that items at the top of the recommendation list have a greater impact.

$$Unexpectedness = \frac{1}{N}\sum_{i=1}^{N} max(\, Pr(S_i) - Prim(S_i), 0\,) \cdot \frac{count(i)}{i} \tag{1}$$

Diversity is evaluated by entropy, which is the variation in the genre of the recommended items [21]. Entropy is calculated using the Equation. (2). $P(x_i)$ is the number of items recommended from genre x_i divided by the total number of items recommended. Since there are 18 genres, the maximum value of entropy is 4.1670.

$$Diversity = -\sum_{i=1}^{n} P(x_i) log_2\, P(x_i) \;\; [bit] \tag{2}$$

Usefulness is defined as the percentage of recommended items with a UCF predictive rating greater than 3.5 (Like) (Equation. (3)) [22]. Higher usefulness indicates higher recommendation accuracy and performance comparable to UCF.

$$Usefulness = \frac{Number\ of\ Like\ items}{Total\ number\ of\ recommended\ items} \times 100[\%] \tag{3}$$

4 Experiment and Analysis

The results of the experiment are shown in Table. 3.

4.1 Unexpectedness

All groups confirmed that proposals 1, 2, 4, 5, and 6 are more surprising than the conventional method. Among them, proposal 6 has the highest value among all groups. In other words, these proposed methods recommended many items that were not recommended by the conventional methods. On the other hand, the existing method and proposal 3 had a value of almost zero, indicating that they recommended only items similar to those recommended by the conventional method. This may be attributed to the fact that the method gave priority to recommendations with high predictive evaluation values.

4.2 Diversity

All methods in all groups showed higher diversity than conventional methods. With the exception of Proposal 2 in Group A, the diversity of the results improved by approximately 5% to 14% compared to the conventional method. These results can be attributed to the fact that the conventional method mostly recommended items from popular movie genres, but each of the proposed methods recommended items that were not recommended in the conventional method, which led to the recommendation of items from multiple genres, thereby improving the diversity of genres. Among the proposed methods, proposal 5 had the highest value among all groups.

4.3 Usefulness

Except for Proposal 1 and Group A's Proposals 4 and 6, the usefulness values exceeded 90%. In particular, Proposal 3 and 5 in Group B, Proposal 2, 3, 4, and 5 in Group C, and the existing methods had usefulness values of 100%. This means that these proposed methods can make recommendations with the same performance as UCF, even though the items to be recommended were changed from UCF. This result may be attributed to the fact that these proposed methods preferentially recommend items with high predictive evaluation values. On the other hand, Proposal 1 has low values for all groups. The reason for the lower recommendation ac-

Table 3: Experimental Results

	Unexpectedness			Diversity [bit]			Usefulness [%]		
Group	A	B	C	A	B	C	A	B	C
Conventional method	-	-	-	3.194	3.167	3.152	-	-	-
Existing method	0.0	0.0	0.0	3.492	3.560	**3.581**	**100**	**100**	**100**
Proposal 1	16.6	20.6	19.3	3.528	3.538	3.452	82.2	68.6	69.8
Proposal 2	20.6	19.2	17.3	3.267	3.473	3.461	94.6	96.4	**100**
Proposal 3	1.8	0.0	0.0	3.382	3.350	3.330	99.8	**100**	**100**
Proposal 4	23.9	18.4	12.7	3.446	3.438	3.386	72.9	98.0	**100**
Proposal 5	9.9	14.2	14.2	**3.547**	**3.575**	3.554	99.8	**100**	**100**
Proposal 6	**24.5**	**23.4**	**21.9**	3.471	3.540	3.377	78.6	91.6	97.5

curacy can be attributed to the fact that items with a large difference between the target user's and other users' predictive evaluation values were recommended, and thus items with low predictive evaluation values were also recommended. In addition, for almost all of the proposed methods, the order of usefulness was Group C, Group B, and Group A, in that order. In other words, it was shown that the greater the number of data points on which the user's items were evaluated that can be used to calculate the predictive evaluation value, the more accurately the user's preferences can be analyzed and the higher the recommendation accuracy.

4.4 Trade-off between Unexpectedness and Usefulness

Proposals 1, 2, 4, and 6, which have relatively high unexpectedness, have lower usefulness, while the existing methods and Proposals 3 and 5, which have relatively low unexpectedness, have higher usefulness. From this result, it can be said that there is a trade-off between usefulness and unexpectedness. In other words, a highly unexpected recommendation is not necessarily useful to the user, and likewise, a highly useful recommendation is not necessarily unexpected to the user. The key to optimizing the user experience is to find the right balance between unexpectedness and usefulness, depending on the purpose of the recommendation system.

4.5 Overall Analysis

Proposal 6 and Proposal 4 are the most unexpected in Group A, in that order, but their usefulness is low. Proposal 2, which is the next most unexpected, has a high usefulness of about 95%, and its diversity is also higher than that of the conventional method. Therefore, considering the balance, Proposal 2, which recommends items similar to those with a large difference between the target user's actual measured value and the average of all users' actual measured values, is considered the optimal choice for Group A. Groups B and C also have a higher surprising value of more than 90% for Proposal 6, which is the most surprising, and a higher diversity than the conventional method. Therefore, for Groups B and C, Proposal 6, which recommends items similar to those with a low percentage of Likes but high measured values, is considered the optimal choice.

5 Conclusion

In this study, we proposed a new method focusing on serendipity in order to eliminate user boredom with recommendations and improve the effectiveness of recommendation systems. Experiments were conducted using unexpectedness, diversity, and usefulness to evaluate serendipity, and a comparison was made among conventional methods, existing methods from previous studies, and six proposed methods. The experimental results revealed that the optimal recommendation method differs depending on the number of data points evaluated by the user. Specifically, for users with a small number of data, recommendation of items similar to items with a large difference between the target user's actual measured value and the average of all users' actual measured values was effective, and for users with a medium or larger number of data, recommendation of items similar to items that other users do not prefer but have high actual measured values was shown to be effective The results of this study are as follows. We confirmed that these methods increase unexpectedness, diversity, and serendipity without sacrificing usefulness. In other words, these methods are expected to alleviate user boredom caused by the recommendation of only similar content, provide new information unknown to the user, and increase user satisfaction.

In future research, it will be necessary to evaluate the generality and effectiveness of this method through further validation on various data sets and different domains. In addition, the development of a dynamic recommendation system that reflects real-time user feedback can be considered to further improve satisfaction through recommendations.

References

[1] D.T. Hawkins, L.R. Levy, and K.L. Montgomery, “Knowledge gateways: The building blocks,” Information Processing & Management, vol. 24, no. 4, 1988, pp. 459–468.

[2] M.J. Bates, “The design of browsing and berrypicking techniques for the online search interface,” Online Review, vol. 13, no. 5, 1989, pp. 407–424.

[3] D. Roy and M. Dutta, “A systematic review and research perspective on recommender systems,” Journal of Big Data, vol. 9, no. 59, 2022.

[4] R. Mu, X. Zeng, and L. Han, “A survey of recommender systems based on deep learning,” IEEE Access, vol. 6, 2018, pp. 69009–69022.

[5] K. Najmani, E.H. Benlahmar, N. Sael, and A. Zellou, “Collaborative filtering approach: A review of recent research,” in Advanced Intelligent Systems for Sustainable Development (AI2SD ’2020), Advances in Intelligent Systems and Computing, vol. 1418, Springer, Cham, 2022, pp. 151–163.

[6] P. Lops, D. Jannach, C. Musto, T. Bogers, and M. Koolen, “Trends in content-based recommendation,” User Modeling and User-Adapted Interaction, vol. 29, no. 2, 2019, p. 239–249.

[7] G. Figa` Talamanca and S. Arfini, “Through the newsfeed glass: Rethinking filter bubbles and echo chambers,” Philosophy & Technology, vol. 35, no. 20, 2022.

[8] M.A. Qaz, M. Nadeem, S.S. Sohail, R. Imam, F. Doctor, Y. Himeur, A. Hussain, A. Amira, “Filter bubbles in recommender systems: Fact or fallacy—a systematic review,” WIREs Data Mining and Knowledge Discovery, vol. 13, no. 6, 2023.

[9] S. Copeland, W. Ross, and M. Sand, “Introduction—a science of serendipity?,” in Serendipity Science: An Emerging Field and its Methods, Springer, Cham, 2023, pp. 1–16.

[10] Y. Ping, Y. Li, and J. Zhu, “Beyond accuracy measures: The effect of diversity, novelty, and serendipity in recommender systems on user engagement,” Electronic Commerce Research, 2024.

[11] S. Karimi, H.A. Rahmani, M. Naghiaei, and L. Safari, “Provider fairness and beyond- accuracy trade-offs in recommender systems,” in 6th FAccTRec Workshop: Responsible Recommendation, RecSys 2023, arXiv preprint arXiv:2309.04250, 2023.

[12] H. Domoto, T. Uchiya, and I. Takumi, "Development of a hybrid information recommendation system considering serendipity," in 2023 IEEE 12th Global Conference on Consumer Electronics (GCCE 2023), IEEE, Kyoto, 2023, pp. 1065–1066.

[13] W. Ren, L. Wang, K. Liu, R. Guo, E.P. Lim, and Y. Fu, "Mitigating popularity bias in recommendation with unbalanced interactions: A gradient perspective," in Proc. 22nd IEEE Int'l Conf. Data Mining (ICDM), IEEE, USA, 2022, pp. 1198–1203.

[14] "Movielens dataset," [Online]. Available: https://grouplens.org/datasets/movielens/. [Accessed: June 21, 2024].

[15] A. Antiopi and B. Boutsinas, "Addressing the cold-start problem in recommender systems based on frequent patterns," Algorithms, vol. 16, no. 4, 2023, p. 182. DOI: 10.3390/a16040182. License CC BY 4.0.

[16] B. Lika, K. Kolomvatsos, and S. Hadjiefthymiades, "Facing the Cold Start Problem in Recommender Systems," Expert Systems with Applications, vol. 41, no. 4, part 2, 2014, pp. 2065–2073.

[17] I. Dokmanic´, R. Parhizkar, J. Ranieri, and M. Vetterli, "Euclidean distance matrices: Essential theory, algorithms, and applications," IEEE Signal Processing Magazine, vol. 32, no. 6, 2015, pp. 12–30.

[18] R.M. Roberts, Serendipity: Accidental discoveries in science, 2nd ed., John Wiley & Sons, New York, 1989.

[19] R.J. Ziarani and R. Ravanmehr, "Serendipity in recommender systems: A systematic literature review," Journal of Computer Science and Technology, vol. 36, no. 2, 2021.

[20] T. Murakami, K. Mori, and R. Orihara, "Metrics for evaluating the serendipity of recommendation lists," Lecture Notes in Computer Science (JSAI 2007), vol. 4914, Springer, 2008, pp. 40–46.

[21] A. Bellog´ın, I. Cantador, and P. Castells, "A study of heterogeneity in recommendations for a social music service," in Proc. 2010 Int'l Conf. Web Intelligence and Intelligent Agent Technology (WI-IAT), vol. 1, IEEE, Toronto, 2010, pp. 150–157.

[22] X. Ning, C. Desrosiers and G. Karypis, "A Comprehensive Survey of Neighborhood-Based Recommendation Methods," in Recommender Systems Handbook, vol. 2, Springer, 2015, pp. 37–76.