# An Effective Approach for Providing Diverse and Serendipitous Recommendations

**Ivy Jain and Hitesh Hasija**

**Abstract** Over the years, recommendation systems successfully suggest relevant items to its users using one of its popular methods of collaborative filtering. But, the current state of recommender system fails to suggest relevant items that are unknown (novel) and surprising (serendipitous) for its users. Therefore, we proposed a new approach that takes as input the positive ratings of the user, positive ratings of the similar users and negative ratings of the dissimilar users to construct a hybrid system capable of providing all possible information about its users. The major contribution of this paper is to diversify the suggestions of items, a user is provided with. The result obtained shows that as compared to general collaborative filtering, our algorithm achieves better catalogue coverage. The novelty and serendipity results also proved the success of the proposed algorithm.

**Keywords** Recommender systems · Collaborative filtering · Weighted catalog coverage · Novelty · Serendipity

## 1 Introduction

Recommendation systems or recommender systems are used very frequently now days to provide ratings or preferences to a given item like movie, restaurants, etc. [1]. Recommender systems are basically classified into three types [2] and they are:

I. Jain (✉)
Vivekananda Institute of Professional Studies, Delhi, India
e-mail: ivy.jain@yahoo.in

H. Hasija
Delhi Technological University, Delhi, India
e-mail: hitoo.hasija@gmail.com

1. Content filtering approach
2. Collaborative filtering approach
3. Hybrid filtering approach

In collaborative filtering approach [3], the recommendations are provided by determining similar kinds of users and then analysing their preferences, considering that similar kinds of users have same likings and same ratings for different items. Last.fm is an example of it. Collaborative recommender is further classified into model and memory based approaches [4]. In memory based approach the rating is based on aggregate of similar user ratings for the same item while in model base approach, ratings are used to train a model which is further used to provide ratings.

In content filtering approach [5], past ratings provided by a user are used to provide new recommendations with various algorithms like Bayesian classifier, decision trees, genetic algorithm etc. [6–8]. For example, based on the previous news browsing habits of a user, it could be determined whether particular news would be liked by user or not. In hybrid based approach [9–13], the characteristics of both content and collaborative are combined into a single model to overcome the shortcomings of both simultaneously. Netflix is the best example of it.

This paper consists of 5 sections, Sect. 2 consists of background details, methodology used to solve the problem is described in Sect. 3, results and comparison with previous approaches has been covered in Sect. 4, finally conclusion and future work are described in Sect. 5.

## 2 Background

Collaborative filtering works on the opinion of people that have interest like you as seen in the past, to predict the items that may interest you now. While providing recommendation, in collaborative filtering approach similar users are determined by various approaches like k-nearest neighbour classifier in such a way that the thinking of all these users is same. Hence, the recommendations provided by these users would also be similar. As one of the approaches to determine similar thinking users in collaborative filtering is Pearson's correlation coefficient. So for our work, we would be using Pearson's correlation coefficient [14, 15] to find the similarity of a user with other users. Pearson's correlation coefficient is calculated by the formula:

$$\rho = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})(x - \bar{x})}\sqrt{\sum (y - \bar{y})(y - \bar{y})}} \tag{1}$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} xi \tag{2}$$

As stated in Eq. (1), Pearson's correlation coefficient signifies the linear relationship between the two variables. As stated in Eq. (2), x (bar) is the calculation of sample mean. According to Cohen [16], the value of Pearson's correlation coefficient ranges from −1 to +1. The zero value of coefficient signifies that, there is no dependency between the two variables. The negative value implies inverse relationship i.e. as the value of one variable increases the value of another variable decreases. The positive value implies positive relationship that is, as the value of one variable increases so as the value of another variable. Pearson's correlation coefficient is symmetric in nature. It is also invariant to the scaling of two variables. Thus, it can also be used for determining correlation between two different variables like the height of a player and his goal taking capability in a basketball match. While determining the value of Pearson's correlation coefficient between two different variables, their units are not changed. Because, the method of determining this coefficient is such that it do not depends on the units of those variables. But, collaborative filtering suffers from some major problems as stated in [17]. Sparse data problem deals with the few or none ratings provided by user for some of the input data, and there are various approaches as justified by [18–20] to deal with it, in different situations. Hence, opting different approaches for different problem domains is a cumbersome task [21].

## 3 Methodology Used to Solve the Problem

If the value of Pearson's correlation coefficient is determined between two users by comparing their recommendations for two different movies, then that value will provide us the information that whether these two users have similar liking or not. If a high positive value which is greater than 0.5 is obtained for any two users then they could be considered as friends of each other. Similarly, friends of friends could be determined. Now, if the value obtained is less than −0.5 then, those two users would be considered as enemies of each other. But, enemies of enemies have similar likings as that of its friends, just like the opposite of opposite would be equivalent to the similar one. In this way, first of all friends as well as friends of friends are determined and then enemies of enemies are determined. In a nutshell, it will provide us with those users whose thinking's are similar. This approach is just like an alternative to the various approaches used in collaborative filtering. Finally, five recommendations are obtained based on friends of a particular user i.e. based on collaborative filtering approach. Then, using positive ratings of similar users i.e. likes of likes for a user, three novel recommendations are obtained. At the end, taking negative ratings of dissimilar users i.e. taking dislikes of dislikes for a user, two serendipitous recommendations are obtained.

## 4 Results and Comparison

**A. Study of the system**

To evaluate the performance of our system we created browser based software and conducted a user study for 100 users and 200 movies in IT laboratory. The group selected for study included IT professional and the movies selected were from "movie lens" dataset of varying genres. The screenshot of the system is shown in Fig. 1.

The system does not provide any recommendation till the user provide ratings of at least 5 movies. To help user's in selecting a movie to watch, this system suggests some movies based on their popularity following a non personalized recommender approach. After the user provided some ratings, the system suggested some normal recommendations based on the popular collaborative filtering, some novel recommendations based on the positive ratings of similar user and some serendipitous recommendations based on the negative ratings of the dissimilar users.

The user then evaluates the recommendations provided by giving a rating on scale 1–5. The user is also asked a question- "are you familiar with the movie" to know the novelty content provided by the system. The user is also provided with an explanation facility, using a question mark beside the novel and serendipitous recommendation in the software. This explanation facility is to let the user
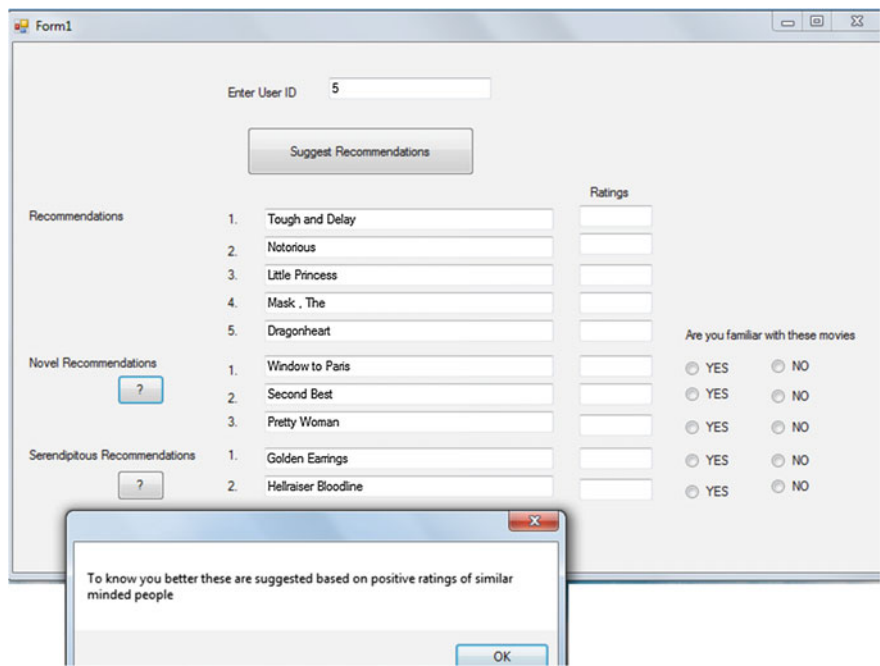


**Fig. 1** Snapshot of the system designed on proposed approach

understand the reason behind separating the recommendations in three different categories and help them evaluate the alternative recommendations better without affecting the interest, user have in the recommender system.

## B. Performance

For the evaluation of system performance, we calculated the metrics based on novelty and serendipity factor only. As, the system has also been provided with recommendations from the collaborative filtering approach the accuracy of the system is not affected and need no attention. The following measures are used for evaluation.

4.1 Precision for less popular items—Precision in general, refers to the number of movies rated positive (above 3) by a user to the total number of movies recommended. We calculated precision for movies recommended as novel and serendipitous. As a result, even low value achieved for such precision would be quite beneficial for the system.

4.2 Weighted catalogue coverage—it is the proportion of useful movies that is recommended by the system for a particular user. It is given by Eq. (3)

$$Weighted\_catalogue\_coverage = \frac{(S_r) \cap (S_s)}{S_s} \tag{3}$$

where $S_r$ the set of movies is recommended to a user and $S_s$ is the set of movies considered useful for that user.

4.3 Serendipity—to measure serendipity we used concept similar in. It is based on the distance applied to movie labels (genre). The comparison between movies is given by Eq. (4)

$$Dist(i,j) = 1 - \frac{(G_i) \cap (G_j)}{(G_i) \cup (G_j)} \tag{4}$$

where $G_i$ is the set of genres describing positively rated serendipitous movie and $G_j$ is the set of genres describing positively rated movie by collaborative filtering.Then to measure surprise of movie i we take,
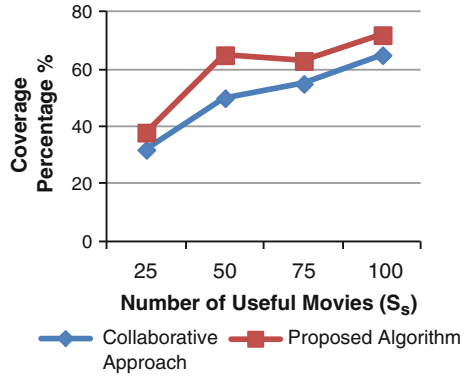
$$S_{count}(i) = \min dist(i, j) \tag{5}$$

In addition, to analyze the novelty of items suggested we consider the answer to the question—"Are you familiar with the movie" as given by the users.

## C. Results

The system was operational for a period of 4 weeks and the ratings and feedback from the users were stored in the system. Precision value of about 60 % was achieved for less popular items. The figures for the precision value obtained are low

**Fig. 2** Graph representing catalogue coverage percentage



but considering it only for the less popular items was significant. For novelty we found that users were not familiar with the movies suggested under the novelty and serendipity section and due to the explanation provided to them, they were interested to try them (Fig. 2).

The major achievement of the system was in case of the weighted catalogue coverage. The result in figure shows that our algorithm was able to provide a better coverage of the movies that was useful to a user. Thus, the proposed algorithm was able to suggest movies rated positive by a user and not included in the list provided by collaborative filtering.

Finally, for label-based surprise metric our algorithm was able to provide good results with an average $S_{count}$ of all items rated positive in serendipitous list = 0.653.

## 5    Conclusion and Future Work

In summary, we proposed an effective recommender system that focussed on providing diverse recommendations to its users that are new and surprising in addition to being relevant. In our method, a system has been designed using positive rating of item given by users (collaborative) to find relevant item, then using the positive rating of the similar users and negative rating of the dissimilar users to suggest the novel and serendipitous recommendations respectively. The system suggests serendipitous movies as a separate list so that even if these movies do not match with user interest, then user should not stop using the recommender system as a whole. In addition to this, user is provided with an explanation of as to why the list is separated thus helping the user to evaluate these movies better.

Result showed that the method helps the system to know user choices better and recommend some diverse movies completely unknown to its users. The precision values calculated in the result section are low but considering it for unknown items is quite high. However, the overall precision for the complete system remains unaffected. The weighted catalogue coverage for a particular user is found to be

better than collaborative filtering as a user is provided with variety of options that matches the user interest. Based on calculation of serendipitous factor, it was showed that a movie rated positive by a user from the serendipitous list is far from the profile of movies in the normal recommendation list. This helps the users to understand the system better.

As a part of future work we would like to study the trade off between coverage obtained and the overall accuracy of the system. We would also like to validate the performance of the proposed algorithm by obtaining user opinions in live study to understand the reason for trying out any recommendation that is completely unknown to them.

# References

1. Francesco Ricci and Lior Rokach and Bracha Shapira, Introduction to Recommender Systems Handbook, Recommender Systems Handbook, Springer, (2011), pp. 1–35.
2. Gediminas Adomavicious et al, "Incorporating Contexual information in recommender systems using a multidimensional approach", ACM transactions on Information Systems, vol. 23, Issue1, (2005).
3. Hosein Jafarkarimi; A.T.H. Sim and R. Saadatdoost A Naïve Recommendation Model for Large Databases, International Journal of Information and Education Technology, June (2012).
4. Juan F. Huete, J.M. Fernandez-Luna, Luis M. de Campos, Miguel A. Rueda-Morales, "using past-prediction accuracy in recommender systems", Journal of Information Sciences 199(12), pp. 78–92, (2012).
5. Prem Melville and Vikas Sindhwani, Recommender Systems, Encyclopedia of Machine Learning, (2010).
6. R. Agrawal, T. Imielinski, and A. Swami, Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD Int'l Conf. Management of Data, pp. 207–216, (1993).
7. J. S. Breese, D. Heckerman, and C. Kadie, Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the 14th Annual Conference on Uncertainty in Artificial in Intelligence, pp. 43–52, (1998).
8. J.-S. Kim, Y.-A Do, J.-W. Ryu, and M.-W. Kim. A collaborative recommendation system using neural networks for increment of performance. Journal of Korean Brain Society, 1 (2):234–244, (2001).
9. R. J. Mooney and L. Roy (1999). Content-based book recommendation using learning for text categorization. In Workshop Recom. Sys.: Algo. and Evaluation.
10. Walter Carrer-Neto, Maria Luisa Hernandez-Alcaraz, Rafael Valencia-Garcia, Francisco Garcia-Sanchez, "Social knowledge-based recommendation system. Application to the movies domain", Journal of Expert Systems with Applications 39(12), pp. 10990–11000, (2012).
11. Duen-Ren Liu, Chin-Hui Lai, Wang-Jung Lee, "A hybrid of sequential rules and collaborative filtering for product recommendation", Journal of Information Sciences 179(09), pp. 3505–3519, (2009).
12. Lingling Zhang, Caifeng Hu, Quan Chen, Yibing Chen, Young Shi, "Domain Knowledge Based Personalized Recommendation Model and Its Application in Cross-selling", Journal of Procedia Computer Science 9(12), pp. 1314–1323, (2012).
13. Tian Chen, Liang He, "Collaborative Filtering based on Demographic Attribute Vector", In Proceedings of International conference on future Computer and Communication, IEEE, (2009).

14. N. Tintarev and J. Masthoff, "The effectiveness of personalized movie explanations: An experiment using commercial meta-data," in Proc. International Conference on Adaptive Hypermedia, (2008), pp. 204–213.
15. Juan Pablo Timpanaro, Isabelle Chrisment, Olivier Festor, "Group-Based Characterisation for the I2P Anonymous File-Sharing Environment".
16. Jacob Cohen. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). Routledge Academic, 2 edition, January (1988).
17. Sanghack Lee and Jihoon Yang and Sung-Yong Park, Discovery of Hidden Similarity on Collaborative Filtering to Overcome Sparsity Problem, Discovery Science, (2007).
18. George Lekakos, George M. Giaglis, Improving the prediction accuracy of recommendation algorithms: Approaches anchored on human factors, Interacting with Computers 18 (2006) 410–431.
19. Tae Hyup Roh, Kyong Joo Oh, Ingoo Han, The collaborative filtering recommendation based on SOM cluster-indexing CBR, Expert System with Application 25(2003) 413–423.
20. Shih, Y.-Y., & Liu, D.-R., Product recommendation approaches: Collaborative filtering via customer lifetime value and customer demands, Expert Systems with Applications (2007),
21. DanEr CHEN, The Collaborative Filtering Recommendation Algorithm Based on BP Neural Networks, (2009) International Symposium on Intelligent Ubiquitous Computing and Education.