

# Enhancement of Collaborative Filtering Using Myers-Briggs Type Indicator (MBTI) Applied in Recommendation System

Arvin Jay C. Nadal<sup>a</sup>, Angelo S. Reyes<sup>a</sup>, Mark Christopher R. Blanco<sup>a</sup>, Antolin J. Alipio<sup>a</sup>, Dan Michael A. Cortez<sup>a</sup>

<sup>a</sup> Computer Science Department, Pamantasan ng Lungsod ng Maynila, Manila, Philippines

## Corresponding author.

Correspondence: Arvin Jay C. Nadal

Email: ajcnadal2018@plm.edu.ph

## Article info

Received 14th January 2021

Received in revised form 23 February

2022 Accepted 26 March 2022

## Keywords

Cold-Start Problem, Collaborative Filtering, Myers-Briggs Type Indicator, Recommender System

## Abstract

Collaborative filtering is one of the most popular recommender systems being used today. Collaborative filtering algorithm depends on the association of one client's activity with another client's activity to discover his nearest neighbors. Related items are expected to be recommended according to his neighbor's similar interests or inclinations. Collaborative filtering algorithm deals with a major problem called the new user challenge or also known as the 'cold-start' problem that arises due to the lack of enough information about the new-coming user. The authors have employed an enhanced collaborative filtering algorithm by incorporating Myers-Briggs Type Indicator or MBTI. With the means of identifying each users MBTI personality types to create neighbourhoods, the researchers have alleviated the problem on the lack of similarities between inexperienced users to existing users. In addition, the system can predict new user ratings for each item using the average rating of users in the same neighbourhood. After predicting the rating, the item with the highest rating is recommended to inexperienced users, which provides a solution to the 'cold-start' problem.

## I. INTRODUCTION

### A. Background of the Study

Recommender system is quite possibly one of the most pervasive techniques to offer customized dynamic assistance to users, through analysis of user preferences, data filtering and artificial intelligence. There are two regularly utilized recommender systems: one is content-based recommender system, which depends on detailed portrayal or descriptions of the traits and attributes of the items being recommended. In contrast to content-based recommender system, another type of recommender system called collaborative filtering recommender system, does not require definite description of item but it is based on the evaluations or ratings of the item from a collective set of users, which shapes users' preference. The researchers will enhance collaborative filtering algorithm in this study.

The collaborative filtering algorithm is an algorithm considering the following three assumptions: individuals have identical inclinations and interests, their inclinations and interests are steady, and we can conclude their decision by implying to their past choices. In view of the above assumptions, collaborative filtering algorithm depends on the association of one client's activity with another client's activity to discover his nearest neighbors. Related items are expected to be recommended according to his neighbor's similar interests or inclinations.

The Myers-Briggs Type Indicator (MBTI®) is an instrument published by CPP, Inc. designed to determine a respondent's preferences in how they see the world and make decisions. The MBTI is designed to determine respondents' preferences in four opposite pairs known as "dichotomies." Each dichotomy is a division of two mutually exclusive groups, in this case, type preferences, typically referred to with a letter abbreviation: Extraversion (E) – Introversion (I), Sensing (S) – Intuition (N), Thinking (T) – Feeling (F), and Judging (J) – Perceiving (P). Based on these preferences patterns, the instrument categorizes a person into one of 16 personality types

## **B. Statement of the Problem**

Recommender systems face some issues and challenges. The two major challenges they deal with are: 'new user' and 'rating sparsity.' The new user challenge or also known as the 'cold-start' problem arises due to the lack of enough information about the new-coming user. The data sparsity challenge refers to the insufficiency of data about users' preferences and priorities. Studies show that there is a significant relationship between users' personality traits and their interests (Yusefi Hafshejani, 2018).

## **C. Objective of the Study**

The general objective of the study is to enhance collaborative filtering algorithm by using Myers-Briggs Type Indicator to be applied in Recommendation System. Specifically, this study aims to provide a solution for the insufficiency of data about users' preferences and the lack of information about the new-coming user known as the 'cold-start' problem.

# **II. LITERATURE REVIEW**

## **A. Recommender Systems**

Recommender systems use information filtering to recommend information of interest to a user and are defined as the system which recommends an appropriate product or service after learning the customers' preferences and desires. Most of the recommender systems research has been focused on the accuracy improvement of recommendation algorithms. Specifically, they have background data, input data and an algorithm that combines background and input data to arrive at its suggestions; therefore, the customers save time and unnecessary efforts to search items and they will receive recommendations for the right products to purchase. (Alyari and Jafari Navimipour, 2018).

The collaborative filtering algorithm is an algorithm based on the following three assumptions: people have comparable preferences and interests, their preferences and interests are stable, and we can conclude their choice by referring to their past preferences. Because of the above expectations, the collaborative algorithm is based on the connection of one user's behavior with another user's behavior to find his immediate neighbors and according to his neighbor's interests or preferences to predict his interests or inclination. Amazon, one of the most famous e-commerce sites, applied collaborative filtering to recommend products to users (Jiang et al., 2018).

## **B. Personality Based Recommender Systems**

A study by Fernandez-Tobias et al. (2016), presents three novel methods to alleviate the new user problem in CF: (a) personality-based CF, which directly improves the recommendation prediction model by incorporating user personality information, (b) personality-based active learning (AL), which utilizes personality information for identifying additional and useful user preference data to be elicited in a target domain, and (c) personality based cross-domain recommendation, which exploits personality information to better exploit user preference data from auxiliary domains in order to compensate the lack of user preference data in the target domain (Jeya Selvi M et al., 2022).

Personality has been increasingly incorporated into recommender systems in recent years. It is reasonable to believe that personality-based recommender systems can provide more personalized information or services, because they can better understand users from the psychological perspective and better explain why a user prefers one option to the other. For instance, as people with similar personality characteristics are more likely to have similar interests and preferences, personality has been adopted to enhance the nearest neighbor measure in collaborative filtering (CF) based recommender systems (Wu et al., 2018, Ibrahim, S. Jafar Ali et al., 2018).

## **C. Cold Start Problem**

Recommender systems face some issues and challenges. The two major challenges they deal with are: 'new user' and 'rating sparsity.' The new user challenge arises due to the lack of enough information about the new-coming user. The data sparsity challenge refers to the insufficiency of data about users' preferences and priorities. Studies show that there is a significant relationship between users' personality traits and their interests. To alleviate the sparsity and new user problems, a collaborative filtering system must cluster users based on their 'personality traits' using the big 5 personality model (Yusefi Hafshejani, 2018).

The Collaborative filtering algorithm also has several limitations such as low scalability when dealing with large amounts of data, and the problem of a cold start. Further, the traditional Collaborative Filtering algorithm needs to compute the similarities of increasing number users to all other users, and it requires higher computation efficiency. It is a significant challenge to improve computation speed for an online recommender system. A Category Preferred Canopy-K-means based Collaborative Filtering (CPCCKCF) algorithm was

proposed that addresses the challenges mentioned above and optimizes recommender systems regarding computation performance and prediction accuracy (Li et al., 2018).

### III. RESEARCH METHODOLOGY

Collaborative Filtering (CF) is a popular method for developing recommender systems. This method is focused on locating the most relevant k users from whom we can extract items to recommend based on their rating history (Duricic et. al., 2018).

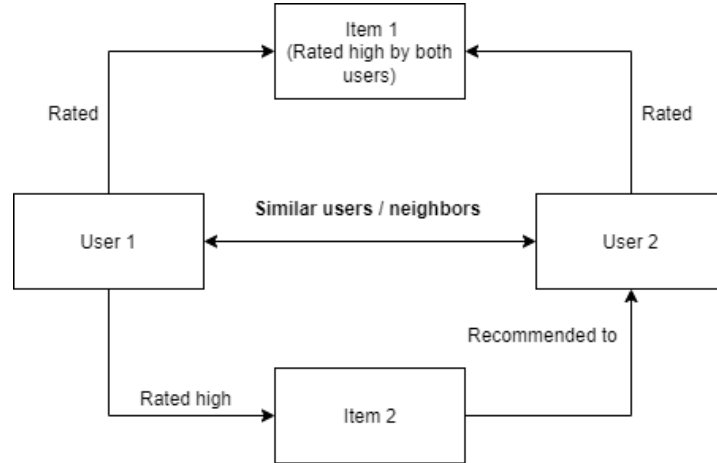


Figure 1. Representation of collaborative filtering.

In figure 1, User 1 and User 2 rated item 1 which makes them similar users or neighbors. Based on User 1 rating to other items, we can recommend items rated high to User 2 which in this figure, is Item 2.

#### A. Equation

Table 1.

Collaborative Filtering Parameters

Symbol	Definition
$U$	The set of all users.
$I$	The set of all items.
$I_{uv}$	The set of co-rated items of user u and user v.
$I_u, I_v$	The set of items which are rated by user u, user v.
$U_{ij}$	The set of co-rated users on item i and item j.
$R_{uj}, R_{vj}$	The rating of user u on item j, the rating of user v on item, respectively.
$\bar{R}_u, \bar{R}_v$	The average rating of user u and user v respectively.
$N(u), N(i)$	The neighbor set of user u and item i respectively.

The first step is to calculate the similarity between users by using the Pearson Correlation Coefficient (PCC). The PCC formula is shown below.

$$\text{sim}(u, v) = \frac{\sum_{j \in I_{uv}} (R_{uj} - \bar{R}_u) (R_{vj} - \bar{R}_v)}{\sqrt{\sum_{j \in I_{uv}} (R_{uj} - \bar{R}_u)^2} \sqrt{\sum_{j \in I_{uv}} (R_{vj} - \bar{R}_v)^2}} \quad (1)$$

We use the average score for users u and v to eliminate any difference in different users' scoring scales and to ensure similarity accuracy.

Find the target user's neighborhood set. After Step one, we obtain the user similarity matrix and sort this matrix according to the degree of similarity. Then, we can determine the top k-neighbor users who are

most similar to the target user, which we label as  $N(u)$ . After that, we can now predict the unknown rating of target user  $u$  for item  $i$ . Using UserCF, we can predict the rating of user  $u$  for item  $i$ , which can be expressed as shown below.

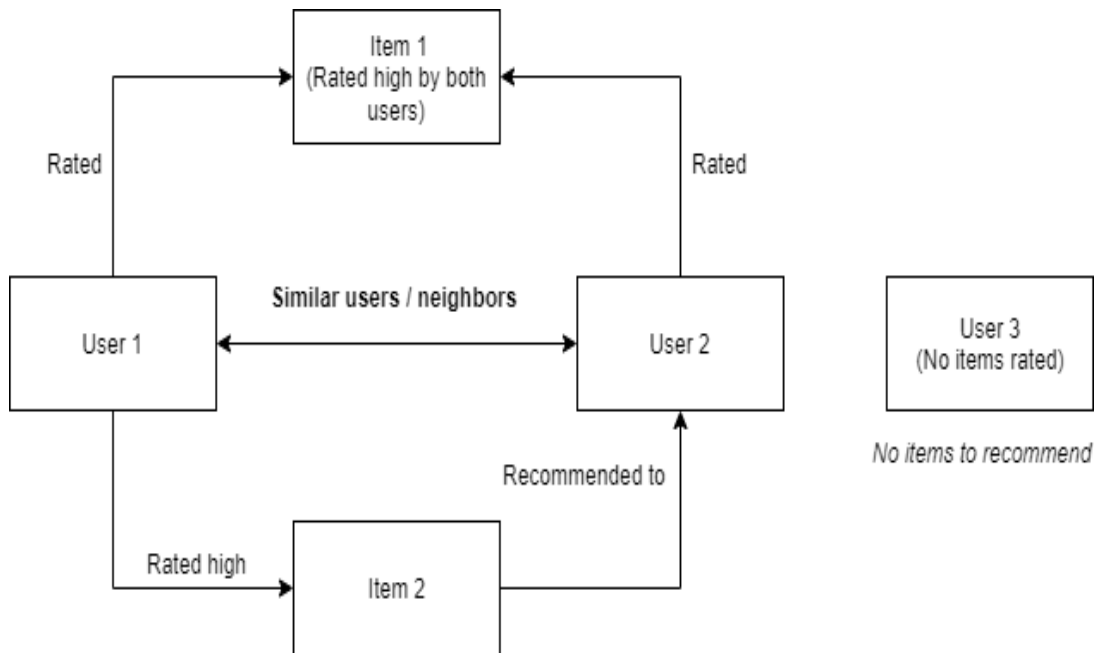
$$P_{UserCF}(R_{ui}) = \bar{R} + \frac{\sum_{v \in N(u)} \text{sim}(u,v)(R_{vi} - \bar{R})}{\sum_{v \in N(u)} \text{sim}(u,v)} \quad (2)$$

The above equation indicates that we can determine the unknown rating of the target user  $u$  for item  $i$  based on the weighted sum of the known rating of the user's neighbor set for item  $i$ . We calculate  $\text{sim}(u,v)$  using the equation in step one.

**B. Existing Algorithm**

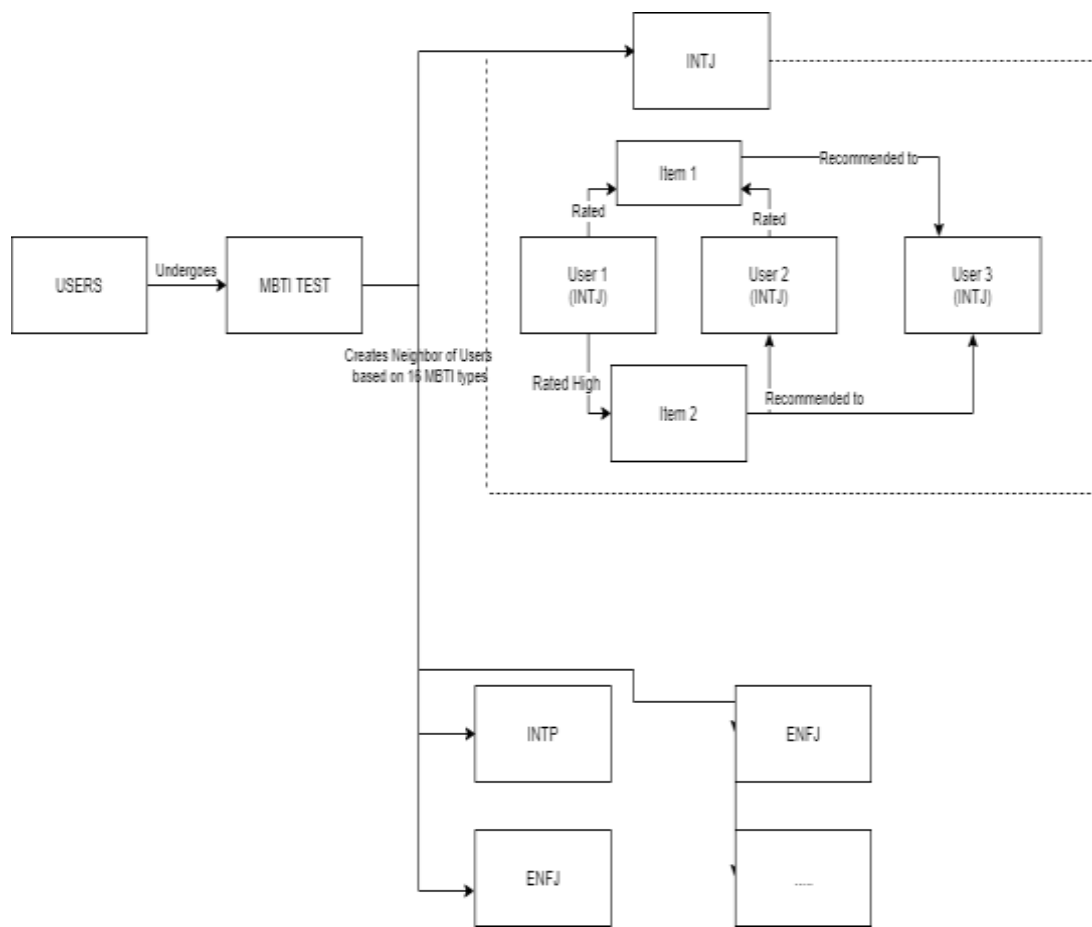
*Note: Same symbol parameters are used as shown in Table 1*

1. Identify set of all users and items
2. Collaborative Filtering Algorithm
  - 2.1 Calculate the similarity between the target user (user whom the prediction is for) and every other users with Pearson Correlation Coefficient
    - 2.1.1 Add all of the co-rated items of user  $u$  (target user) and user  $v$ . Summation of the product of co-rated items of user  $u$  and  $v$ 
      - 2.1.1.1 Subtract the average ratings of user  $u$  for every co-rated item. Multiply to the difference of the average ratings of user  $v$  for every co-rated item.
    - 2.1.2 Calculate the deviation ratings for user  $u$  and user  $v$  individually
    - 2.1.3 divide
    - 2.1.4 if  $\text{sim}(u,v) > 0$  then user  $u$  and  $v$  are similar or neighbors
    - 2.1.5 else if  $\text{sim}(u,v) < 0$  then user  $u$  and  $v$  is not similar and are different
  - 2.2 Predict the Rating of user  $u$  for item  $i$ 
    - 2.2.1 Summation of the product of pearson correlation of target user and user  $v$  rating for ever user  $v$  in target neighbor; divided by
    - 2.2.2 Summation of pearson correlation coefficient for every user  $v$  in nearest neighbor.
    - 2.2.3 Add the quotient to target user's average ratings



**Figure 2. Example of cold-start problem**

In figure 2, A new user or User 3, was introduced. User 3 does not have any history of rating items. This means that User 3 has no neighbors with similar interest. Since collaborative filtering utilizes ratings from similar users, there will be no items that can be recommended to User 3 which causes a cold-start problem.



**Figure 3. The proposed solution for the cold-start problem is presented.**

In figure 3, all users must determine their MBTI personality type. These users will have neighbors with the same personality type. New users, for example, User 3, will now have items recommended to them. All items rated by their neighbors will have the possibility to be recommended to them thus providing a solution to the cold-start problem.

**C. Existing Algorithm (MBTI Based Collaborative Filtering Recommendation System)**

*Note: Same symbol parameters are used as shown in Table 1*

1. Identify set of all users and items
2. Identify personality type of users based on MBTI's 16 personality types
3. Calculate the similarity between the target user (user whom the prediction is for) and every other users of the same personality type using Pearson Correlation Coefficient
  - 3.1 IF target user is a new user
    - 3.1.1 Go to step 4
  - 3.2 ELSE
    - 3.2.1 Add all of the co-rated items of user u target user) and user v. Summation of the product of co-rated items of user u and v
      - 3.2.1.1 Subtract the average ratings of user u for every co-rated item. Multiply to the difference of the average ratings of user v for every co-rated item.
    - 3.2.2 Calculate the deviation ratings for user u and user v individually
    - 3.2.3 divide
    - 3.2.4 if  $sim(u,v) > 0$  then user u and v are similar or neighbors
    - 3.2.5 else if  $sim(u,v) < 0$  then user u and v is not similar and are different
4. Predict the Rating of user u for item i
  - 4.1 IF target user is a new user
    - 4.1.1 Predicted rating is averaged ratings of users with the same neighborhood or personality type.
  - 4.2 ELSE Summation of the product of pearson correlation of target user and user v rating for ever user v in target neighbor; divided by
  - 4.3 Summation of pearson correlation coefficient for every user v in nearest neighbor.

4.4 Add the quotient to target user's average ratings.

In the proposed enhanced collaborative filtering algorithm above, the researchers applied the use of the MBTI personality types as shown in step 2 to solve the 'cold-start' problem. After identifying the personality type of the users, the system will place the users in the neighborhood that corresponds to their personality type. With this setup, the system can now recommend items more accurately even to new users based on the ratings of other users on the same neighborhood (personality type).

**IV. FINDING AND DISCUSSION**

The following table shows an example simulation of collaborative filtering algorithm to display how the cold-start problem occurs.

**A. Cold-Start Problem**

**Table 2.**  
**Example Simulation for Existing Collaborative Filtering Algorithm**

	ITEM 1	ITEM 2	ITEM 3	ITEM 4	ITEM 5
USER 1	2		5		4
USER 2	1	4			
USER 3	5		1	5	1
USER 4		5	5		3
NEW USER 5	?	?	?	?	?

The first step is to calculate the similarities of other users to New User 5. In order to compute the similarities of new user to other users, the set of co-rated items of the new user and other users must be determined. Since the new user doesn't have any existing ratings, then the set of co-rated items will be 0, thus there are no similarities between the new user and other users which causes the 'cold-start' problem of the collaborative filtering algorithm.

**B. Cold-Start Problem Solution**

**Table 3.**  
**Example Simulation for Enhanced Collaborative Filtering Algorithm using MBTI (Users with ISTJ Personality type)**

MBTI	ITEM 1	ITEM 2	ITEM 3	ITEM 4	ITEM 5
USER 1	2		5		4
USER 4	1	5	5		3
NEW USER 5 (predicted ratings)	1.5	5	5		3.5

The first step is to determine the MBTI Personality Type of all users. In order to compute the similarities of new user to other users, the set of co-rated items of the new user and other users must be determined. Since user 5 is a new user, then it doesn't have any existing ratings then it will not have similar neighbors for computing the recommendation of items. People with similar personality characteristics are more likely to have similar interests and preferences (Wu et al., 2017). Therefore, the researchers used the new user's personality to recommend items. Users with the same personality type can be considered as neighbors. Using the averaged ratings of the users within the same neighborhood, the system can predict the ratings of the new user to each item. After predicting the ratings, items with high ratings will be recommended to the new user therefore providing a solution for the 'cold-start' problem.

## V. CONCLUSION AND FURTHER RESEARCH

The researchers have employed an enhanced collaborative filtering algorithm by incorporating Myers-Briggs Type Indicator or MBTI. The existing collaborative filtering algorithm suffers from the 'cold-start' problem that arises due to the lack of enough information about the new-coming user. With the means of identifying each users MBTI personality types to create neighbourhoods, the researchers have alleviated the problem on the lack of similarities between new users to existing users. Additionally, using the averaged ratings of the users within the same neighborhood, the algorithm can predict the ratings of the new user to items rated within the same neighborhood. After predicting the ratings, items with high ratings will be recommended to the new user therefore providing a solution for the 'cold-start' problem.

## REFERENCES

1. Alyari, F. and Navimipour, N.J., 2018. Recommender systems: a systematic review of the state of the art literature and suggestions for future research. *Kybernetes*.
2. Duricic, T., Lacic, E., Kowald, D. and Lex, E., 2018, September. Trust-based collaborative filtering: Tackling the cold start problem using regular equivalence. In *Proceedings of the 12th ACM Conference on Recommender Systems* (pp. 446-450).
3. Fernández-Tobías, I., Braunhofer, M., Elahi, M., Ricci, F. and Cantador, I., 2016. Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction*, 26(2), pp.221-255.
4. Jiang, L., Cheng, Y., Yang, L., Li, J., Yan, H. and Wang, X., 2019. A trust-based collaborative filtering algorithm for E-commerce recommendation system. *Journal of ambient intelligence and humanized computing*, 10(8), pp.3023-3034.
5. Li, J., Zhang, K., Yang, X., Wei, P., Wang, J., Mitra, K. and Ranjan, R., 2019. Category preferred canopy-K-means based collaborative filtering algorithm. *Future Generation Computer Systems*, 93, pp.1046-1054.
6. Wu, W., Chen, L. and Zhao, Y., 2018. Personalizing recommendation diversity based on user personality. *User Modeling and User-Adapted Interaction*, 28(3), pp.237-276.
7. Yusefi Hafshejani, Z., Kaedi, M. and Fatemi, A., 2018. Improving sparsity and new user problems in collaborative filtering by clustering the personality factors. *Electronic Commerce Research*, 18(4), pp.813-836.
8. Ibrahim, S. Jafar Ali, and M. Thangamani. "Momentous Innovations in the Prospective Method of Drug Development." In *Proceedings of the 2018 International Conference on Digital Medicine and Image Processing*, pp. 37-41. 2018.
9. Jeyaselvi, M., M. Sathya, S. Suchitra, S. Jafar Ali Ibrahim, and N. S. Kalyan Chakravarthy. "SVM-Based Cloning and Jamming Attack Detection in IoT Sensor Networks." In *Advances in Information Communication Technology and Computing*, pp. 461-471. Springer, Singapore, 2022.