

MBTI-Based Collaborative Recommendation System: A Case Study of Webtoon Contents

Myeong-Yeon Yi, O-Joun Lee^(✉), and Jason J. Jung

School of Computer Science and Engineering, Chung-Ang University,
Heunseok-Dong, Seoul 156-756, South Korea
{dalaetm, concert034, j3jung}@cau.ac.kr

Abstract. A large number of Webtoon contents has caused difficulties on finding relevant Webtoons for users. Thereby, an efficient recommendation services are needed. However, since the existing recommendation method (e.g. collaborative filtering) has two fundamental problems: (i.e., data sparsity and scalability problem), it has difficulties with reflecting users' personality. In this paper, we propose the MBTI-CF method to solve these problems and to involve users' personality by building personality-based neighborhood using MBTI. In order to verify the efficiency of the proposed method, we conducted statistical testing by user survey (anonymous users have rated set of the pre-selected Webtoon contents). Three experimental results have shown that MBTI-CF provides improvement in terms of the data sparsity problem and the scalability problem and offers more stable performance.

Keywords: Webtoon · Recommendation · MBTI (Myers-Briggs Type Indicator) · Collaborative filtering

1 Introduction

Webtoon¹ is digital comic contents published on the web. With the emergence of smart devices, the Webtoon has been the most popular digital contents in South Korea and actively reproduced in various media like movies, dramas and so on. The amount of the Webtoons is huge and increasing rapidly. For example, in 2014, the number of Webtoons in Naver recorded 520 while it is released in 2004, later than Yahoo! Korea's "Cartoon Sae-sang" (in 2002), and Daum's "Manhwa sok Sae-sang" (in 2003). Webtoon Platform corp. TapasMedia recorded about 21,500 Webtoons in 2014, its first year. It is getting more difficult for users to find relevant Webtoons since the number of the Webtoons is growing rapidly. Hence, an efficient recommendation service is needed.

There are three major approaches to recommend digital contents: content-based filtering, demographic filtering and collaborative filtering. The Content-based filtering and the demographic filtering require external data that are difficult to get. However, the collaborative filtering can work without external data. Moreover, it performs higher performance than others [1]. Therefore, we propose recommendation system for the webtoon using collaborative filtering.

¹ Webtoon is also known as web comics, online comics, internet comics.

Two major problems of collaborative filtering are data sparsity problem and scalability problem [2, 15]. The data Sparsity problem occurs when a rating matrix is sparse. The rating matrix can be sparse in many situations such as cold-start problem, new-user, new-item problem (when a new user or item has entered in the system). Since the user has not rated or purchased items or the item has not been rated yet, it is difficult to find a group of similar users or items. In other word, the lack of rating history causes the data sparsity [3].

The scalability problem occurs when the number of users or items grows. Dealing with the scalability problem is important because the numbers of the users and the items are extremely large in the real world. Since the existing CF algorithms have to estimate the similarities of every users and items, they suffer serious scalability problem.

Another issue of the existing recommendation systems is that they have difficulties with reflecting users' personalities. We need to develop a new approach to extract users' personalities and reflect them on the recommendation system. There were some studies suggesting personality-based recommendation system [4, 17]. They used *Big Five Factor personality model* (Big Five) to solve the data sparsity problem of the collaborative filtering. However, the scalability problem, one of the collaborative filtering's major problems, remained unsolved.

In this paper, we propose to use *Myers-Briggs Type Indicator* (MBTI) to understand users' personalities and to solve the major two problems of the collaborative filtering. The MBTI is a psychometric questionnaire to measure psychological preference. There are four bipolar discontinuous scales which are implied in Jung's theory that humans experience the world by the four major principles - sensation, intuition, feeling, and thinking [6]. The MBTI categorize people in 16 types by an abbreviation of the four initial letters of each of their four type of preferences. For example, people who are extraverted and prefer sensing, thinking and judgment will be categorized to ESTJ type.

The remainder of this paper is organized as follows. In Sect. 2, we provide a brief overview of related works. Section 3 explains the procedure of the proposed method. In Sect. 4, three experiments evaluate the improvement and the performance by comparing the result of the proposed method with that of the existing method. In Sect. 5, we draw a conclusion that provides a summary of the proposed method and future work.

2 Related Work

2.1 Existing Recommendation Systems for the Webtoon

There have been no precedent study of recommendation Systems for the Webtoon. The recommendation services that the existing webtoon service platforms are providing are limited to using content-based filtering and demographic filtering or so on. For example, there are platforms provided by Naver, Daum and so on [14]. They recommend the Webtoons which are preferred by users with common age and gender or which are in same genre that the user preferred.

However, the content-based filtering works based on a description of the content or the users' choices made in the past. The demographic filtering suppose that the users

with common personal attributes like age and gender will also have common preferences [12]. Therefore, the content-based filtering and the demographic filtering cannot give high-performance recommendations if the system has not enough external information [11]. The collaborative filtering, however, analyzes the users' ratings without much external information of a specific domain. Since the Webtoons are multimedia contents, we propose recommendation system using the collaborative filtering.

2.2 Recommendation System Based on Personality Information

Personality of human consists of two parts. One is consistent behavior patterns and the other is intrapersonal processes [7]. The previous researches have shown that the personality is highly related to the preferences and tastes [9]. In other words, people with similar personality would have similar behavior patterns and preferences. Hence, we can predict a person's preference of items by investigating users' purchased items or preferences whose personalities are similar to the person.

Recommendation System Based on the Big Five. There are studies on collaborative filtering recommendation systems that use personality information to group similar users [4, 17]. They apply the Big Five, the widely used personality model within psychology. The Big Five is a hierarchical model of personality traits with five factors [5]. However, since the Big Five measures personality traits on a dimensional scale, they had to estimate similarities between all users. Thus, the scalability problem had grown worse as a result of using the Big Five.

To solve the scalability problem, we propose to use the MBTI. The MBTI is leading academic model of personality with the Big Five. On the contrary to the Big Five, the MBTI is typological so that we can categorize users to 16 types. Since we make personality-based neighborhood group by using MBTI, we do not have to estimate all similarities between all the users. Thus, we can relieve the scalability problem.

Recommendation System Based on MBTI. Song et al. proposed recommendation system based on collaborative filtering using emotional word selection and the MBTI. They conducted experiments to show that the users with the same MBTI type will select similar emotional words and have similar movie preference [8]. In the experiments, they selected several movies and extracted emotional words for subjective evaluation on the movies. Then, they classified subjects by the MBTI and encouraged them to select the emotional words on each movie and to evaluate the rating for each movie.

As a result of the experiments, the subjects with the same MBTI type had choose similar emotional words and rated similarly except for extremely popular movies. Finally, they drew a conclusion that the users with the same MBTI type have similar emotions and preferences on the same movie. Thus, it is reasonable to use the MBTI to recommend the movies. Since the Webtoons are narrative contents like the movies, we propose a new approach to use the MBTI to recommend the Webtoons.

3 MBTI-Based Collaborative Filtering

The collaborative filtering using the Big Five Factor personality model has been already proposed. It has solved the data sparsity problem by building personality-based neighborhoods using the Big Five. However, the Big Five measures personality traits on a dimensional scale so recommendation system using the Big Five had to estimate similarities between all users. Thus, the scalability problem, one of the collaborative filtering's major problem, had grown worse. Therefore, we need to develop a new approach to relieve the collaborative filtering's fundamental problems.

The MBTI-based Collaborative Filtering (MBTI-CF) can solve the data sparsity by building neighborhoods using the MBTI. In contrast with the Big Five, the MBTI is typological so users can be categorized into 16 types. Since the MBTI-CF makes the personality-based neighborhood group by using the MBTI, we do not have to estimate all similarities between all users. Thus, we can relieve the scalability problem. And by using the psychometric questionnaire MBTI, we can make recommendation system reflecting users' personalities.

The proposed method consists of three major parts.

- (1) Normalizing ratings and Grouping users by the MBTI: we normalize ratings to unify the standard of ratings, and identify neighborhoods of the user using the MBTI.
- (2) Computing Similarities between users in a neighborhood: we compute similarities between users using vector cosine similarity.
- (3) Estimating Prediction of user-preference: we estimate prediction of user-preference by computing weighted average of all the ratings for the item.

3.1 Normalizing Ratings and Grouping Users by MBTI

Standard points and measure of the rating scores vary across users, we need to unify the standard of the ratings. Therefore, newly-arrived ratings are normalized by the Gaussian Probability Model. This step is formulated as

$$preR_{u,n} = \frac{F_{u,n} - \overrightarrow{F_u}}{\sigma(F_u)} \quad (1)$$

where $F_{u,n}$ represents a newly-arrived rating that user u rated for item n , $\overrightarrow{F_u}$ and $\sigma(F_u)$ are the average and standard deviation of the historical ratings rated by the user u respectively. $preR_{u,n}$ indicates the preprocessed rating of which standard is unified. After normalizing ratings, we identify a neighborhood of the new user by using the MBTI.

3.2 Computing Similarities Between Users in Neighborhood

We have to estimate all similarities between the users with the same neighborhood before we make a prediction of the preference.

In order to estimate the similarities, we adopted the vector cosine similarity, which compares two users' ratings by the cosine of the angle between the users' corresponding rating vectors [13]. The vector cosine similarity between user u_i and u_j is given by

$$w_{u_i, u_j} = \cos(\vec{u}_i, \vec{u}_j) = \frac{\vec{u}_i \cdot \vec{u}_j}{\|\vec{u}_i\| * \|\vec{u}_j\|} \quad (2)$$

where u_i is i -th user, u_j is j -th user, then \vec{u}_i is vectors consist of rating by the user u_i and \vec{u}_j is the vectors consist of rating by the user u_j . And “ \cdot ” denotes the dot-product of the two vectors [2]. If R is the $n \times m$ user-rating matrix, then the similarity between two users u_i and u_j is defined as the cosine of the n dimensional rating vectors corresponding to the i -th and j -th row of the matrix R .

3.3 Estimating Prediction of User-Preference

In Sect. 3.2, the similarity matrix between the users is created. Each component of the matrix has a similarity between the users, and the similarities are measured by using the vector cosine similarity.

To make a prediction of the preference on an item i for active user a , a weighted average of all the ratings on that item is given by

$$P_{a,i} = \bar{r}_a + \frac{\sum_{u \in U} (r_{u,i} - \bar{r}_u) \cdot w_{a,u}}{\sum_{u \in U} |w_{a,u}|} \quad (3)$$

where \bar{r}_a and \bar{r}_u indicate the average ratings of all the items that the user a and user u rated. $w_{a,u}$ is the weight between the user a and user u that we estimated in the Sect. 3.3. The summation is over all the users $u \in U$ who have rated the item i . Finally, we can predict $P_{a,i}$, which is the preference on the item for the user [2].

4 Experimental Results and Analysis

This section presents three experiments that aim to verify improvement of the proposed method on three different foci: performance, robustness for the data sparsity and robustness for the scalability [10]. In the first experiment, we compared performance of the MBTI-CF with that of the CF. In the second experiment, the improvement of the robustness for the data sparsity is verified. We verify the improvement of new-user problem which is one of the major situations in that the data sparsity occurs. In the third experiment, the improvement of the robustness for the scalability is verified by comparing a time complexity of MBTI-CF with that of CF.

4.1 Experimental Environment

In order to collect Webtoons ratings, we conducted a survey of anonymous users by Google docs. The questionnaire we distributed is shown in Table 1.

Table 1. Average, standard deviation and range of MAE of CF and MBTI-CF.

	Question	Choices
1	What is your MBTI type?	ISTJ, ISTP, ISFJ, ISFP, INFJ, INTJ, INFP, INTP, ESTP, ESFP, ESTJ, ESFJ, ENFP, ENTP, ENFJ, ENTJ
2	What is your gender?	Male, Female
3	What age group are you in?	~ 10, 10 ~ 19, 20 ~ 29, 30 ~ 39, 40 ~ 49, 50 ~
4	What method do you mostly use to read Webtoons?	Smartphone, PC, Tablet PC, Comic book,
5	What Webtoon service platform do you mainly use? (Pick two of them)	Naver, Daum, Lezhin Comics, Toptoon, Olleh Market, Nate, Kakao Page, Yahoo, etc.
6	Give ratings to these 20 Webtoons.	1 ~ 10

For question 6, we selected 20 Webtoons which are popular in different genre. We finally have obtained 90 of survey replies.

4.2 Performance Evaluation

In order to evaluate the performance, we compared Mean Absolute Error (MAE) of the MBTI-CF with that of the CF. The MAE is the most widely used measure of prediction error of the CF [2]. It is an average of the absolute deviation between a predicted rating and a real rating for the items. The MAE is formulated as

$$\text{MAE} = \frac{\sum_{i=1}^N |r_i - p_i|}{N}, \quad (4)$$

where N is the number of items, and p_i and r_i represent the prediction of preference and the real rating of i -th item. A lower MAE value means better prediction performance.

Evaluation of performance was conducted as followings. First, we predicted users' preferences of items using the MBTI-CF and the CF. Second, we measured MAE of the two methods and compared MAE of the two methods.

As shown in Table 2, the MBTI-CF presents more stable performance than the CF but shows a slight lower performance on average performance. The MBTI-CF shows an improvement of 70.88 % over the given standard deviation and 50.96 % over the given range. The MBTI-CF shows lower performance of 9.58 % over the given average.

The results show that the performance of the MBTI-CF is more stable but less accurate than that of the existing CF.

Table 2. Average, standard deviation and range of MAE of CF and MBTI-CF

	CF	MBTI-CF	Improvement
Average	0.678	0.743	-9.58 %
S.D.	0.261	0.076	70.88 %
Range	1.042	0.511	50.96 %

4.3 Dealing with Data Sparsity

To verify the improvement of the robustness against the data sparsity, we investigated the improvement of the new-user problem which is one of the major situation that the data sparsity occurs.

The CF cannot recommend items to a new user since there is not enough user’s historical rating. However, the MBTI-CF can estimate the new user’s prediction of preferences on the items by assuming that the new user has the same weight to other users in neighborhood.

To evaluate the performance of the MBTI-CF on the new-user problem, we compared the MAEs under two circumstances, when the data of the user exist and when the user is a new user. Figure 1 shows MAEs for each circumstance classified by MBTI types and averages of them. Table 3 shows the average, standard deviation, and range of the MAE for the two circumstances.

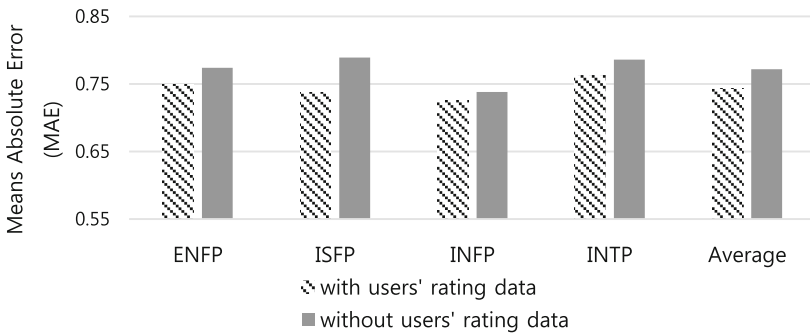


Fig. 1. MAE of each MBTI type and average

Table 3. Average, standard deviation and range of MAE when data exist and when the user is new user.

	With users' rating data	Without users' rating data
Average	0.743	0.772
S.D.	0.076	0.078
Range	0.511	0.523

The results in Fig. 1 indicate the MBTI-CF performs similarly whether the data of the user exist or not. The increase of average of MAE is 3.80 % when the user is the new user.

As shown in Table 3, the stability of the performance has slightly reduced. The standard deviation and range of MAEs when the user is new user had increased 2.63 % and 2.34 % in percentage, respectively.

Observing the results, we can claim that the performance of the MBTI-CF is stable whether the user is new or not

4.4 Improvement of the Scalability

To verify the improvement of the scalability, we compared time complexity of MBTI-CF with that of CF. Let U be a set of n users and I a set of m items and k a size of the neighborhood.

The collaborative filtering procedure is composed of 3 major parts: estimating similarity, grouping the similar users and predicting the preference of the item. The computation complexity of a user similarities to other users is $O(nm)$ as explained below:

```

 $u_a$  : a particular user
For each user  $u_i$ 
  For the set of  $m$  items that have been co-rated by
    user  $u_a$  and  $u_i$ 
      Compute similarity between  $u_a$  and  $u_i$ 

```

In order to group the similar users, we have to find k users who are most similar to the user to build a neighborhood. The computation complexity of grouping the similar users is $O(n)$ since we have to find k users which have maximum similarity to the user in n users.

The computation complexity of predicting the preference of the item is $O(k)$ since we have to compute weighted average of 1 items for k users.

The most expensive computation of the classic CF is the computation of the user-to-user similarities [16]. Since the MBTI-CF builds 16 neighborhoods by MBTI types, we don't have to estimate the user similarities of all the users to group the similar users. We have to estimate similarities between the user and the users in the neighborhood. The average neighborhood size will be $\frac{m}{16}$. Thus, the computation complexity of the user similarities to the other users reduced in 1 over 16. The computation complexity of predicting preference of the item is $O(k)$ since we have to compute weighted average of 1 items for k users.

Since the MBTI-CF does not have to estimate the user similarities of all the users, the computation of user-to-user similarities is reduced linearly. Therefore, building the neighborhood using the MBTI improves the scalability.

4.5 Result Analysis

The results of the experiments show that the scalability problem in the existing CF is relieved in the MBTI-CF. Also, the performance of the MBTI-CF is more stable than that of the existing CF. But the performance of the MBTI-CF is less accurate than that of the classic CF.

This is because there was a tradeoff between the accuracy and the scalability. Grouping users by the MBTI could exclude a user who rated most similarly to the user from the neighborhood. However, the scalability is improved since the number of similarities that has to be computed is reduced.

The result of the second experiment shows that the performance of the MBTI-CF is stable whether the user is the new user or not. Since we can identify the new user's neighborhood by his MBTI type, we can solve the data sparsity problem.

5 Conclusion

A large number of the Webtoon contents has caused difficulties to find relevant Webtoons for users. Thus, we need a systematic process to recommend the users a suitable Webtoon. We propose recommendation system using collaborative filtering since collaborative filtering works without external data and performs high performance. However the collaborative filtering has two fundamental problems: the data sparsity problem, the scalability problem. In addition, the existing recommendation systems for the Webtoons have difficulties with reflecting users' personality.

In this paper, we proposed the MBTI-CF to solve the data sparsity problem and the scalability problem by building personality-based neighborhood using the MBTI personality type. Also, the MBTI-CF reflects users' personalities.

In order to verify the improvement of the proposed method, we conducted survey of anonymous internet users to collect Webtoons ratings. The three experiments have shown that MBTI-CF provides improvement in terms of the data sparsity problem and the scalability problem. Also, the proposed method offers more stable performance.

In the future work, we will conduct testing of the proposed method with more datasets and understand further about the performance of the MBTI-CF. The MBTI-CF presents a more stable performance than the existing CF, but shows a slight lower performance on average performance. Therefore, we will find solutions to improve the performance of the MBTI-CF.

Acknowledgments. This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2015-H8501-15-1018) supervised by the IITP (Institute for Information & Communications Technology Promotion). Also, this work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (NRF-2014R1A2A2A05007154).

References

1. Adomavicius, G., Tuzhilin, A.: Toward the next generation of recommender systems: a survey of the state of-the-art and possible extensions. *IEEE Trans. Knowl. Data Eng.* **17**(6), 734–749 (2005)
2. Su, X., Khoshgoftaar, T.M.: A survey of collaborative filtering techniques. *Adv. Artif. Intell.* **2009**, 1–19 (2009). Article ID 421425, Hindawi
3. Billsus, D., Pazzani, M.J.: Learning collaborative information filters. In: Proceedings of the 15th International Conference on Machine Learning, vol. 98, pp. 46–54 (1998)
4. Tkalčič, M., Kunaver, M., Tasič, J., Košir, A.: Personality based user similarity measure for a collaborative recommender system. In: Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction-Real World Challenges, pp. 30–37 (2009)
5. Gosling, S.D., Rentfrow, P.J., Swann Jr., W.B.: A very brief measure of the Big-Five personality domains. *J. Res. Pers.* **37**(6), 504–528 (2003)
6. Harasym, P.H., Leong, E.J., Juschka, B.B., Lucier, G.E., Lorcheider, F.L.: Myers-briggs psychological type and achievement in anatomy and physiology. *Am. J. Physiol.* **268**(6 pt. 3), S61–S65 (1995)
7. Burger, J.M.: Personality, 7th edn. Thomson/Wadsworth, Belmont (2008)
8. Kim, H.-G., Namgoong, H., Eune, J., Song, M.: A proposed movie recommendation method using emotional word selection. In: Ozok, A., Zaphiris, P. (eds.) OCSC 2009. LNCS, vol. 5621, pp. 525–534. Springer, Heidelberg (2009)
9. Rentfrow, P.J., Gosling, S.D.: The do re mi’s of everyday life: the structure and personality correlates of music preferences. *J. Pers. Soc. Psychol.* **84**(6), 1236–1256 (2003)
10. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. *ACM Trans. Inf. Syst.* **22**(1), 5–53 (2004)
11. Pazzani, M.J., Billsus, D.: Content-based recommendation systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 325–341. Springer, Heidelberg (2007)
12. Pazzani, M.J.: A framework for collaborative, content-based and demographic filtering. *Artif. Intell. Rev.* **13**(5–6), 393–408 (1999)
13. Breese, J.S., Heckerman, D., Kadie, C.: Empirical analysis of predictive algorithms for collaborative filtering. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43–52 (1998)
14. Naver Webtoon. <http://comic.naver.com/webtoon/weekday.nhn>
15. Lee, O.-J., Jung, J.J., You, E.-S.: Predictive clustering for performance stability in collaborative filtering techniques. In: Proceedings of 2nd IEEE International Conference on Cybernetics, CYBCONF 2015, Gdynia, Poland, pp. 24–26 (2015)
16. Rousidis, I., Plexousakis, D., Theoharopoulos, E., Papagelis, M.: Incremental collaborative filtering for highly-scalable recommendation algorithms. In: Hacid, M.-S., Murray, N.V., Raš, Z.W., Tsumoto, S. (eds.) ISMIS 2005. LNCS (LNAI), vol. 3488, pp. 553–561. Springer, Heidelberg (2005)
17. Hu, R., Pu, P.: Enhancing collaborative filtering systems with personality information. In: RecSys 2011, Proceedings of the Fifth ACM Conference on Recommender Systems, pp. 197–204 (2011)