

A Collaborative Filtering Systems based on Personality Information

Jianwang Wang^{1, a}

¹ Beihang Sino-French Engineer School, Beihang University, Beijing 100191, China

^aemail: jwwang_ecpkn@163.com

Keywords: Recommender System, Collaborative Filtering, Personality, Cold Start

Abstract. As one of the most successful recommendation approaches, collaborative filtering (CF) has been attracting an increasing amount of interest in academia. However, the cold-start problem, where historical data is too sparse since new users have not rated enough items, limits the success of collaborative filtering in certain application domains. In this paper, an innovative item-personality-based collaborative filtering system based on personality information is proposed to resolve the cold-start problem. By incorporating human personality into the collaborative filtering framework, we improve the conventional user-based collaborative filtering method. By a linear combination of both user-based CF and item-personality-based CF, the predicted rating can be obtained. An experiment is conducted to evaluate the effectiveness. The results show that the proposed item-personality-based CF can considerably alleviate cold-start problem.

Introduction

Among various recommendation systems, the most successfully utilized is collaborative filtering (CF) system [1], which mimics word-of-mouth. Collaborative filtering methods [1], first proposed for predicting ratings of single items, have the capability of working in domains where items' attribute contents are difficult to obtain or cannot easily be parsed by automatic processes. Typically in a CF system, user-based approach is widely used to make recommendations. Namely predictions of the unknown rating for user u on a non-rated target item i are made by known ratings of the similar users.

However, apart from its widespread adoption, CF suffers from several major limitations including cold-start problems, system scalability, and synonymy. Among these problems, the cold-start issue is the problem that researchers most frequently confront. It refers to difficulties encountered by recommender algorithms when a new item or user is added to a CF system.

It has been presented that personality influences human decision making process and can reveal a person's long-term tastes [2]. Furthermore, there are studies in the personalization process about user modeling, human factors, such as personality and cognitive/learning style [3]. In this paper, considering the importance of personality, we show a novel approach to resolve the cold-start problem. Specifically, our approach incorporates human personality characteristics into traditional user-based collaborative filtering framework in a linear way. An experiment is conducted to compare the performance of the proposed personality-based CF system. The results present that our method can effectively alleviate the cold-start problem in conventional user-based CF recommender system.

Related Work

Cold-start problem

The cold-start problem includes both data scarcity and new user problems. For the first aspect, recommendation faces up with a problem of lack of historical data, such as what have been purchased or rated. As a result, it is highly probable that the similarity between any two given users is nearly zero, so that many measures are unreliable to use. In the other aspect, the problem is baneful in effectively identifying similar users, which leads to a malfunction of the collaborative filtering. Therefore, the cold-start problem considerably inhibits the practical applications of the CF methods.

Due to its practical importance, cold start has attracted numerous researches in recent years, and many methods have been proposed to resolve cold-start problem. Pazzani [4] presented a hybrid recommendation approach to combines content-based and CF techniques. Based on the readily available user data, Nguyen [5] predicted unobserved ratings and favorite genres for a new user by a rule-based induction process to automatically associate a better initial profile for a new user.

User-based collaborative filtering

User-based collaborative filtering, shown in Figure1, has been widely used in recommendation systems. User-based CF can be divided into two steps: the choice of neighborhood and rating prediction. In order to choose neighbors, CF systems identify k most similar users for a target user. To accurately measure the similarities between users, two similarity measures have been proposed in the literature [6], named Pearson and Cosine similarity coefficients.

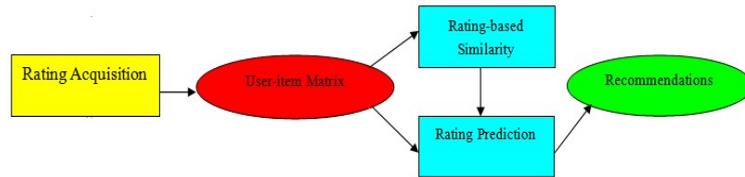


Fig.1. User-based collaborative filtering

The prediction is computed by aggregating neighbor’s ratings. More specially, the predicted rating of user u for item i , $\hat{r}_{u,i}$ can be calculated by the formula below:

$$\hat{r}_{u,i} = \bar{r}_u + \kappa \sum_{v \in N(u)} sim(u, v) \times (r_{v,i} - \bar{r}_v) \quad (1)$$

Where the $r_{v,i}$ is the rating of user v for item i , \bar{r}_u , \bar{r}_v is the mean rating of user u and user v respectively. $N(u)$ is the set of users similar to user u (who have rated item i). κ is the normalization factor. $sim(u, v)$ is the correlation coefficient. The Pearson correlation coefficient is one of the most commonly adopted similarity measure. By using it, the proximity between user u and user v is formalized in equation (2). Another widely-used method of calculating correlation coefficient is called Cosine similarity, it can be calculated as equation (3):

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_{uv}} (r_{v,i} - \bar{r}_v)^2}} \quad (2)$$

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} r_{u,i} r_{v,i}}{\sqrt{\sum_{i \in I_{uv}} r_{u,i}^2} \sqrt{\sum_{i \in I_{uv}} r_{v,i}^2}} \quad (3)$$

Where I is the set of items rated by both users u and v , $r_{u,i}$ is the rating of user u for item i .

Because different users use the ratings scale differently, the normalize ratings by the mean rating in the second step. So, the more similar with u a user v is, the higher the weight of his rating.

Personality

Personality, regarded as a consistent behavior pattern and intrapersonal processes originating within the individual [7], has great stability and foreseeability, which makes personality an enduring and primary factor influencing human behaviors. So there is a significant connection between personality and people’s tastes. It has been proved that people with similar personality will have similar interests and similar behavioral patterns [8], which provides basis for application of personality to recommender systems.

As the most successful model to measure the personality, the Big-Five framework [9] is a hierarchical model of personality traits with five broad factors, which represent personality at the broadest level of abstraction. It categorizes human personality traits into five bipolar dimensions: Openness to Experience, Conscientiousness, Extroversion, Agreeableness, and Neuroticism (also refers to as emotion stability). The Big-Five framework presents that most individual differences in human personality can be classified into five broad, empirically derived domains.

Personality-based approaches

Despite of the success of user-based CF, the cold-start problem is its inevitable obstacle. Considering the stability and foreseeability of personality, we integrate personality into the user-based CF and propose a novel kind of personality-based CF – item-personality-based CF.

Item-personality

Unlike the approach proposed by Hu in [10], our approach never depends on employing a personality-based neighborhood to distinguish users' personalities or find their similar behaviors. Instead, we concentrate on the item in network structure where different users are correlated by rating the same item. Because people with different personalities give different rating results, the item could indirectly reflect personality. Furthermore, the item reflects more personality on people who gave it high rate. So, we propose the new concept "item-personality", as its characteristic, which presents the personality best relevant to the item. To measure item-personality, we put a user's personality characteristics in a vector. Specifically, the personality descriptor $p_u = (p_u^1, p_u^2, \dots, p_u^n)^T$, is a n-dimension vector, and each dimension represents one characteristic of the personality. In our study, we adopt Big Five Factor personality model, the most widely used and extensively researched personality models. As the embodiment of human nature on item, the item-personality vector IP can also be a 5-dimension vector, expressed as:

$$IP = \frac{\sum_{i \in I} r_i P_i}{\left\| \sum_{i \in I} r_i P_i \right\|} \quad (4)$$

In this formula, the item-personality vector IP is a 5-dimension normalized vector for a certain item, and p_i is the normalized personality vector in net, while r_i is the rating of user i on item.

Item-personality-based CF

In order to alleviate cold-start problem, we combine rating from item-personality and that from traditional user-based CF in order to obtain item-personality-based CF. More specifically, the rating of item-personality-based CF can be presented as:

$$c\hat{r}_{u,i} = \alpha \hat{r}_{u,i} + (1 - \alpha) IP^T \cdot IP \quad (5)$$

$c\hat{r}_{u,i}$ is the predicted rating from item-personality-based CF, and $\hat{r}_{u,i}$ is predicted rating from user-based CF. α is the coefficient, and by simulation, we obtain $\alpha=0.7$ is the optimal value. By integrating item-personality, the item-personality-based CF has the advantage of user-based CF and can alleviate the cold start problem as the same time.

Experiment

In order to investigate the performance of our method, we conducted an experiment by comparing it with the traditional user-based CF method in both scenarios of sparse data and new users. Through our study, we proved our approach can effectively work on making predictions in cold-start scenario

At present, most available test datasets only contain user ratings, item attribute contents, or user demographic information. However, it is hard to freely get dataset containing both users' personality information and ratings data. In this study, according to the method in [11], we chose 2013 users in the MovieLens dataset, and acquired their Big-Five personality vectors by their rating on a certain group item. We added a new user, whose personality has been known, into the net. After a few behaviors, we recommend items for the new user by both user-based CF and item-personality-based CF. Then, we repeat the experiment on ten other new users. The result is shown in table 1:

Table.1. Number of movies recommended in both user-based CF and item-personality-based CF

	N_{\max}	N_{\min}	\bar{N}
User-based CF	3	0	1.2
Item-personality-based CF	11	5	7.6

where N_{\max} , N_{\min} , \bar{N} are respectively the largest number, the smallest number and the average number of movies recommended for all new user. Evidently, the item-personality-based CF can give a more effective recommendation for new users and alleviate cold-start problem.

Conclusion

In this paper, in order to deal with the cold-start problem, a new concept named item-personality is presented so that it can be integrated into the widely-used user-based collaborative filtering method to improve its performance. After a linear combination of the predicted rating gained by user-based CF and item-personality-based CF with a specific calculated parameter, a more accurate predicted rating is achieved and then a novel collaborative filtering system based on personality information is established. The effectiveness has been proved by the experiment we conducted comparing the proposed method with the traditional user-based CF in a cold-start scenario. The results show that the new approach which incorporates the item-personality-based CF can considerably improve the performance of the traditional user-based CF, particularly in the cold-start problem.

References

- [1] Resnick. P., Varian, H.R: Recommender Systems. Comm. of the ACM Vol. 40 (1997), p.56–58
- [2] P.J. Rentfrow: the structure and personality correlates of music preferences. J Pers Soc Psychol. Vol. 84 (2003), p.1236-1256.
- [3] Nunes, M.A.S.N: Recommender Systems based on Personality Traits: Could human psychological aspects influence the computer decision-making process? VDM 2009
- [4] M.J. Pazzani: A Framework for Collaborative, Content-Based and Demographic Filtering, Artif. Intell. Rev. Vol.13, (1999), p.393-408.
- [5] A. Nguyen: Improving new user recommendations with rule-based induction on cold user data, ACM conference on Recommender systems, 1297251,(2007), p.121-128.
- [6] H. J. Ahn: A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. Inf. Sci. Vol.178, (2008), p.37-51.
- [7] Burger, J.M: Personality. Wadsworth Publishing, Belmont, CA. (2010).
- [8] Kemp, A.E: The Music Temperament: Psychology and Personality of Musicians. Oxford University Press, New York. (1996).
- [9] S. Gosling, P. Rentfrow and W. Swann: A very brief measure of the Big-Five personality domains. Journal of Research in Personality. Vol.37, (2003), p.504-528.
- [10] R.Hu and P. Pu: Enhancing Collaborative Filtering Systems with Personality Information, ACM, (2011), p.197-204
- [11] S. D. Gosling, and P. J. Rentfrow: A very brief measure of the Big-Five personality domains, Journal of Research in Personality Vol.37, (2003), p.504–528