

# A User Personality-Similarity Model for Personalized Followee Recommendation in SINA Microblog

Pan Xiao<sup>(✉)</sup>, YongQuan Fan, YaJun Du, and Chun Yu

School of Computer and Software Engineering, Xihua University,  
Chengdu 610039, China  
1049434055@qq.com

**Abstract.** Followee recommendation plays an important role in information sharing over microblogging platforms. But as the popularity of microblogging sites increases, the difficulty of deciding who to follow also increases. To solve this problem, in this paper, we propose a User Personality-Similarity (UPS) model for followee recommendation, a novel personality followee recommendation scheme over microblogging systems based on user attributes and the big-five personality model. It quantitatively analyses the effects of user personality in followee selection by combining personality traits with the most commonly user-based predictive factors in microblog. We conduct comprehensive experiments on a large-scale dataset collected from Sina Weibo, the most popular microblogging system in China. The results show that our scheme greatly outperforms existing schemes in terms of precision and an accurate appreciation of this model tied to a quantitative analysis of personality is crucial for potential followees selection, and thus, enhance recommendation.

**Keywords:** Followee recommendation · Microblog  
Personality traits · User similarity

## 1 Introduction

Since the emergence of social networks and microblogging sites, such as Twitter and Sina Weibo, hundreds of millions of users have become to use the microblogging service as a tool to propagate and share information on the Internet [1]. Personality affects everyone's thinking, behavior and decision-making and maintain its long-term stability and affect its social relations [2]. The behavior of users in the online world and in the real world is the same [3]. However, there are few researches on personality theory in the previous studies.

---

This work is supported by the National Nature Science Foundation (Grant No. 61472329 and 61532009), the Key Natural Science Foundation of Xihua University (Z1412620) and the Innovation Fund of Postgraduate, Xihua University.

The researcher combines personality psychology and social network analysis to analyze and predict the user's personality through the user's behavior and other data. The potential value of personality prediction will be very helpful in solving these problems.

The rest of this paper is organised as follows. In Sect. 2, we discuss related work. Section 3 presents the followee recommendation method and the UPS model we propose. In Sect. 4, we evaluate the performance of our design experiment, present the proposed weights for quantitatively evaluating users' personality and how to combine it with other recommendation and compare to existing schemes. Section 5 summarises the conclusions obtained from the performed experiment evaluation.

## 2 Related Work

Personalized recommendation technology is the core and critical technology of E-commerce recommendation system. However, these traditional methods only focus on how to improve the accuracy of recommendation, while ignoring the inherent characteristics of the user behavior is determined by their personality characteristics [4].

Some more related worked on followee recommendation has also been proposed. Most of the existing followee recommendation systems on micro-blogging platforms rely on either topological or content-based factors. Hannon et al. [5] use bag-of-word model to exploit the content information created by a user and recommend him/her followees based on the content similarity between the candidate followees and the target user. Sun et al. [6] construct a diffusion graph to select a small subset of tweets as recommended emergency new feeds for regular users. There are a few people use personality trait in followee recommendation systems. Wu et al. [7] generate personalized tags for Twitter users to label their interest by extracting keywords from tweets they post. Armentano et al. [8] propose a followee recommender system using social relation features, including user popularity and number of common friends, to measure the relevance among users in Twitter. However, the recommendation effect of adding personality traits has improved compared with the traditional recommendation method, but there is much room for improvement.

## 3 Our Method

### 3.1 User-Based Factor

**The Analysis of User Attributes of Microblogging.** In this section, we discuss the representation of the micro-blogging user model. In this work, we consider 4 properties of user information, which are location information,

micro-blogging text, social information and interactive information. Thus, the user model of the user  $u$  is represented as follows:

$$Profile_{user}(u) = \{Place(u), Posts(u), Relation(u), Interaction(u)\} \quad (1)$$

where  $Place(u)$  is the location information of the user  $u$ , both of which are short text and can be represented as a string.  $Post(u)$  represents the long text that the user  $u$  has released into the micro-blogging, which is represented as a text vector.  $Relation(u)$  represents the social information of  $u$ , including two kinds of attribute information, which is concerned about attention information, fans information.  $Interaction(u)$  represents the interactive information of user  $u$ , including two kinds of attribute information, that is, forwarding information and comments.

Firstly, we define the user( $u, v$ ) in the various attribute similarity and then weighted and combined. The similarity is given by the following equation:

$$\begin{aligned} sim(u, v) = & \omega_1 sim(Place(u), Place(v)) + \omega_2 sim(Post(u), Post(v)) \\ & + \omega_3 sim(Relation(u), Relation(v)) \\ & + \omega_4 sim(Interaction(u), Interaction(v)) \end{aligned} \quad (2)$$

In essence,  $\omega_i$  is the weight of each attribute similarity, and  $\omega_1 + \omega_2 + \omega_3 + \omega_4 = 1$ . The  $sim(Place(u), Place(v)) = 1$  if the information of the provincial and city of the  $Place(u)$  and  $Place(v)$  are the same. The  $sim(Place(u), Place(v)) = 2/3$  if the information of provincial of the  $Place(u)$  and  $Place(v)$  is the same, but the city information is different. For two users( $u$ ) and user( $v$ ), their micro-blogging text can be represented as two text feature vectors:  $Post(u) = (w_{u1}, w_{u2}, \dots, w_{un})$ ,  $Post(v) = (w_{v1}, w_{v2}, \dots, w_{vn})$ . The text similarity  $sim(Post(u), Post(v))$  is calculated by cosine similarity, which used TF-IDF. The similarity of the relationship information between the two users are computed weighted averages of  $Followee(u)$  and  $Followee(v)$ . Also, we use forwarding and comment numbers to compute the similarity of  $sim(Interaction(u), Interaction(v))$ .

### 3.2 Personality-Based Factors

**The Big-Five Personality Model.** The ‘‘Big Five’’ model of personality dimensions has emerged as one of the most well-researched and well-regarded measures of personality structure in recent years. The model’s five domains of personality, Openness, Conscientiousness, Extroversion, Agreeableness, and Neuroticism, were conceived by Tupes and Christal as the fundamental traits that emerged from analyses of previous personality across age, gender, and cultural lines.

**The Matching Calculation of the Personality Traits.** TextMind is a Chinese language psychological analysis system developed by Computational Cyber-Psychology Lab, Institute of Psychology, Chinese Academy of Sciences. Through

the relationship between the function of these words and the text, in this paper, we can obtain the relationship between each word in TextMind Chinese psychological analysis system dictionary and each specific factor in the big five personality. The character factor values for the  $i$ th dimension in word  $w$  are defined as follows:

$$BFM(\omega_i) = \frac{\sum_{j=1}^n P_{ij}}{n} \quad (3)$$

In Eq. (3),  $\omega_i$  denotes the  $i$ th personality factor in word  $w$ .  $n$  indicates that word  $w$  has  $n$  functional nouns.  $P_{ij}$  represents the relationship between the  $j$ th word function and the  $i$ th personality factor. The functional word of TextMind and BFM corresponding a correspondence table can be divided into five groups.

In this paper, we calculate the personality score vector of each dimension of the big five personality according to the BFM correspondence table [9]. The user  $u$ 's personality score in the  $i$ th dimension is calculated as follows:

$$Score(u)_i = \frac{\sum_{j=1}^{N_i} k_{\omega_j} \cdot BFM(\omega_{ji})}{N_i} \quad (4)$$

Where  $BFM(\omega_{ji})$  denotes the personality factor value of the  $j$ th-class word in the  $i$ th dimension in the micro-blogging text published by the user  $u$ .  $k_{\omega_j}$  represents the word frequency of the  $j$ th-class word  $\omega$  of the microblogging text published by the user  $u$ .  $N_i$  is the total number of functional words that are statistically relevant under the  $i$ th personality dimension. We use the Eq. (4) to calculate the average score between five dimensions in the table, which were  $\mu_E$ ,  $\mu_A$ ,  $\mu_C$ ,  $\mu_N$ ,  $\mu_O$ , and then we put the user scores compared with the average correlation score.

**The User Personality-Similarity Model.** In this Section, we proposed the user personality-similarity model to make the final recommendation. The personality matching score between a user  $u$  and the potential blogger  $pf$  is expressed as follows:

$$TPM(u, pf) = \mu(\sum MS(u, pf, dim)) \quad (5)$$

In Eq. (5),  $TPM(u, pf)$  is the total personality matching (TPM) score between user  $u$  and potential followee  $pf$ .  $\mu$  is the average value of each dimension.  $MS(u, pf, dim)$  indicates the personality matching score of the user  $u$  and the potential recommendation followee  $pf$  in a certain dimension.

Next, the matching score calculation formula for each dimension is defined as follows [10]:

$$scoreAgreement(u, pf, dim) = \begin{cases} 0.5 & \text{both } u \text{ and } pf \text{ are dimension} \\ 0.25 & \text{either } u \text{ or } pf \text{ are dimension} \\ 0 & \text{None is dimension} \end{cases} \quad (6)$$

Finally, we formally define the evaluation of the formula in the UPS model as follows,

$$FScore = \gamma_1 sim(u, v) + \gamma_2 TPM(u, pf) \quad (7)$$

Where  $\gamma_i$  is the weight value of the formula, and  $\gamma_1 + \gamma_2 = 1$ . Finally, we use this formula as the final evaluation formula in followee recommendation.

## 4 Experiment

In this experiment, We use the microblogging users between the various attributes of information similarity calculation method for user recommendation experiment. In this paper, two evaluation indicators are used to evaluate the user similarity algorithm which are p@N and average precision.

We use the large-scale dataset crawled from Sina Weibo, the most popular microblogging system in China. The dataset contains 256.7 million users' social link information and 550 million tweets. The dataset include tweets, user relations and user background information.

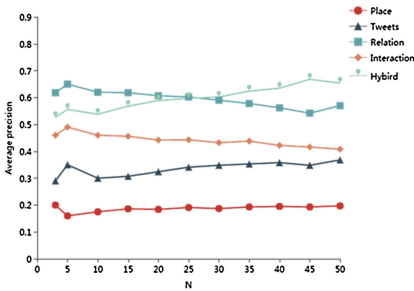


Fig. 1. Recommendation results based on user similarity

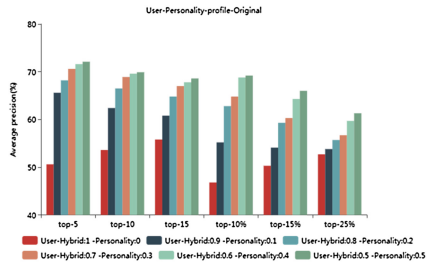


Fig. 2. Personality-aware of compared methods

### 4.1 Results

In the experiment, the weight of the attribute similarity calculation formula between the user(u) and user(v) is calculated by Analytic Hierarchy Process (AHP).

Figure 1 shows the experimental results of the above similarity on P@N, which reflects the motivation of the experiment is to examine the user information and its four kinds of attribute information (place, tweets, relation, interaction) in the calculation of user similarity performance. The experiment results show that compared with the four attribute information of user information, the P@N results of the similarity of relation are best in the previous P@25 and the overall similarity of the users is the second.

Unlike when computing the personality profiles in which ever term appearing in tweets was considered, terms in tweets were filtered according to a processing approach. The approach considered the full-text of tweets (named Original). The effect of adding personality as a factor in user-based followee recommendation

is shown in Fig. 2. The figure summarises the average precision for each of the predefined  $N$ , including results for six linear combinations of factor's weights. As a result, it can be stated that considering a quantitative analysis of personality in combination with user-based factors could help to correctly place the most important or interesting users in the first positions of the ranking of suggested users.

## 5 Conclusions

In this paper, we propose a User Personality-Similarity (UPS) model for followee recommendation, a novel personality followee recommendation scheme over microblogging systems based on user attribute and the big-five personality model. This paper analysed how user personality conditions the followee selection process by combining a quantitative analysis of personality traits with the most commonly used predictive factors for followee recommendation. The combined attributes were insert into a recommendation algorithm that computed the similarity among target users and potential followees. We conduct experiments using large-scale traces form Sina Weibo to evaluate our design. Results show that UPS model greatly outperforms existing recommendation schemes.

## References

1. Dong, Y., Tang, J., Wu, S., et al.: Link prediction and recommendation across heterogeneous social networks. In: IEEE International Conference on Data Mining, pp. 181–190. IEEE (2013)
2. Selfhout, M., Burk, W., Branje, S., et al.: Emerging late adolescent friendship networks and Big Five personality traits: a social network approach. *J. Personal.* **78**(2), 509–538 (2010)
3. Barrick, M.R., Mount, M.K.: The Big Five personality dimensions and job performance: a meta-analysis. *Pers. Psychol.* **51**(1), 849–857 (1991)
4. Gao, R., Hao, B., Bai, S., et al.: Improving user profile with personality traits predicted from social media content. In: ACM Conference on Recommender Systems, pp. 355–358. ACM (2013)
5. Hannon, J., Mccarthy, K., Smyth, B.: The pursuit of happiness: searching for worthy followees on Twitter. In: Irish Conference on Artificial Intelligence and Cognitive Science, pp. 1761–1762 (2011)
6. Sun, A.R., Cheng, J., Zeng, D.D.: A novel recommendation framework for microblogging based on information diffusion. *Social Science Electronic Publishing*, vol. 12(7) (2010)
7. Wu, W., Zhang, B., Ostendorf, M.: Automatic generation of personalized annotation tags for twitter users. In: Proceedings of the Human Language Technologies: Conference of the North American Chapter of the Association of Computational Linguistics, DBLP 2010, Los Angeles, California, USA, 2–4 June 2010, pp. 689–692 (2010)
8. Armentano, M.G., Godoy, D.L., Amandi, A.: A topology-based approach for followees recommendation in Twitter. In: International Workshop on Intelligent Techniques for Web Personalization & Recommendation, pp. 22–29 (2013)

9. Mairesse, F., Walker, M.A., Mehl, M.R., et al.: Using linguistic cues for the automatic recognition of personality in conversation and text. *J. Artif. Intell. Res.* **30**(1), 457–500 (2007)
10. Tommasel, A., Corbellini, A., Godoy, D., et al.: Personality-aware followee recommendation algorithms: an empirical analysis. *Eng. Appl. Artif. Intell.* **51**(C), 24–36 (2016)