Addressing the New User Problem with a Personality Based User Similarity Measure

Marko Tkalčič, Matevž Kunaver, Andrej Košir and Jurij Tasič

University of Ljubljana Faculty of electrical engineering
Tržaška 25, Ljubljana, Slovenia
marko.tkalcic@fe.uni-lj.si

Abstract

The new user problem is a recurring problem in memory based collaborative recommender systems (MBCR). It occurs when a new user is added to the system and there are not enough information to make a good selection of the user's neighbours. As a consequence, the recommended items have poor correlation with the user's interests. We addressed the new user problem by observing the user similarity measure (USM) employed. In this paper we present two novelties that address the new user problem : (i) the usage of a personality based USM to alleviate the new user problem and (ii) a method for establishing the boundary of the CSP. We successfully used a personality based USM that yielded significantly better recommender performance in the period where the new user problem occurs. Furthermore we presented a new methodology for assessing the boundary of the period where the new user problem occurs.

1 Introduction

The new user problem is an important issue in memory based collaborative recommender systems [1, 6]. It occurs when a new user joins the system and there are no (or there are too few) overlapping ratings to calculate good estimates of user similarities with with rating based user similarity measures (USM). We will denote this period as the cold start period (CSP). The consequences of being in the CSP are bad rating predictions for unseen items and thus poor quality of the recommender system. Usually, the new user problem has been addressed by introducing content based approaches which resulted in hybrid systems [3, 7, 4, 2]. Once the system has enough overlapping items it is not in the CSP and rating based USM can be used.

In this paper we present two novelties that address the new user problem : (i) the usage of a personality based USM to alleviate the new user problem and (ii) a method for establishing the boundary of the CSP.

We propose to use the personality based USM developed by [12] in the CSP to alleviate the issue of poor recommendations. In this USM each user is modeled with a five-tuple of scalar values which represent the big five personality factors as described by [5]. The drawback of this approach is a potentially annoying filling of the questionnaire. The benefit however is big as it allows us to calculate user similarities immediately, without waiting for the user to rate several items. The underlying assumption for choosing personality as the basis for the proposed user similarity measure is that people with similar personalities have similar tastes. By using personality for the calculation of user similarity we have taken into account a substantial part of between-users variance. We base the latter statement on the assumption that personality accounts for individual differences in emotive, interpersonal, experiential and attitudinal styles as claimed by [8].

The second novelty is a statistical method for determining at which point the new user problem stops occurring. A review of literature showed that authors either (i) did not set limits for the CSP [11] or (ii) provided limits without further argumentation, e.g. [10] defined *cold start users as the users who have expressed less than 5 ratings*. We propose to determine the boundary of the CSP with a statistical approach, as the number of ratings where the recommender's performance stops being significantly lower than the performance with higher number of ratings given by the user.

We used the LDOS-PerAff-1 dataset [13] to validate the proposed solutions. We conducted a simulation of the new user problem by removing ratings from the rating matrix and thus controlling the sparsity level of the rating matrix. User similarities were calculated using the rating based USM proposed by [9] and the proposed personality based USM. We conducted a collaborative filtering experiment based on both USM that calculated the predicted ratings for the missing values in the rating matrix. We assessed the performance of such a recommender system with the confusion matrix based scalar F-measure F.

2 The new user problem

Most commonly the new user problem in collaborative filtering recommenders is described as the period from the moment when a user joins the system to the moment when there are enough ratings to yield stable lists of neighbours. We rewrote this description from various sources [1, 11, 2, 6]. To the best of the authors' knowledge no formal definition of the new user problem period is available.

In this section we define the boundary of the CSP. Let us have a user u joining the system. The user starts using the system and gives ratings $r(u,h)$ to items $h \in H$ where $H \subset \{h_1 \dots h_J\}$, a set of J items. At any given moment the user has given n ratings to n different items which yields the set

$$R_u^n = \{r_1^u \dots r_n^u\} \quad (1)$$

The boundary of the new user problem period (the CSP) for the selected user is the number of ratings N_u^{CS} after which the system starts to yield stable sets of users. The consequence of a stable set of users is a stable confusion matrix of recommended items. We define that the confusion matrix is stable if a sequence of F measure values (as defined in [6]), has statistically equivalent means at different n.

We choose the F measure as a scalar measure of the confusion matrix. We denote the F measure when n ratings have been used to calculate neighbours as F^n. We define the CSP boundary as the point where the means of F values of the sets

$$R_u^{NJ} = \{F1^N \dots F^J\} \qquad R_u^{(N-1)J} = \{F^{(N-1)} \dots F^J\} \quad (2)$$

are significantly different.

After we defined the boundary of the CSP we evaluated the performance of the standard rating based USM as defined in [9] with the personality based USM as defined in [12].

3 Materials and Methods

We conducted two experiments: (i) one with the personality based USM and (ii) one with the rating based USM.

3.1 Personality based USM recommender

We used the LDOS-PerAff-1 [13] dataset which contained all data necessary to carry out our experiments. We used the personality based USM as defined in [12] to calculate the distances between users and took the $k = 7$ nearest neighbours. We calculated the predicted ratings based on the neighbours' ratings using the adjusted Pearson's coefficient as defined in [9]. We then compared the predicted ratings with the real ones provided by the dataset and got the confusion matrix for the recommender with the personality based USM.

3.2 Simulation of the new user problem

The dataset used in our experiments had a full ratings-items table without missing values with I users and K ratings. To simulate the new user problem we determined a usage history path, in the form of a random sequence of ratings, for each user separately. We iterated through cold start stages s from one (the user has given only one rating) to K (the user has rated all items) for each user separately. At each stage $1 \leq s \leq K$ we performed the recommender procedure and calculated the confusion matrix for the observed user u at the observed stage s. We chose the F measure, as defined in [6], as the performance measure of the recommender system. The experimental procedure thus yielded a table of F values at different stages $s \in \{1..K\}$ and for each user $u \in U$.

3.3 Evaluation methodology

We compared the performance of both user similarity measures by testing the hypothesis $H_0 : \mu_R = \mu_{B5}$ at different cold start stages using the t test. The value μ_R represents the mean F values using the rating based USM and μ_{B5} represents the mean F values using the personality based USM.

We determined the position of the CSP boundary by testing the hypothesis $H_0 : \mu_s = \mu_{s-K}$ where μ_s represents the mean F value at stage s and μ_{s-K} represents the mean F value from stages $s+1$ to K, where K is the last observed cold start stage.

4 Results

We analyzed the CSP by graphing the quality rate of the recommender(F) versus the number of ratings used. At each cold start stage s the F measures for each user were calculated. We grouped the F values for each cold start stage and visualized the distribution of F over all users as a boxplot with the cold start stage fixed. Fig. 1 shows the boxplots for cold start stages from 1 to 70 for the rating based USM. The distribution of F measures has a lower mean at low cold start stages and starts to converge toward a stable mean F value as the cold start stage increases. The leftmost boxplot in Fig. 1 represents the results of the personality based USM and are independent of the stage s.

The results of the t test showed that, on the dataset used, the personality based USM yields a significantly higher mean of F values than the rating based USM when the number of ratings taken in account for the calculation of the neighbours is lower than 50 (see Fig. 3). When the number of ratings is higher than 50 the means of F values for both similarity measures are not significantly different at $\alpha = 0.05$.

When seeking for the CSP boundary we calculated the p values which are shown in Fig. 2. On the dataset used we observed that $p < 0.05$ occurs when the cold start stage is $s < 6$.

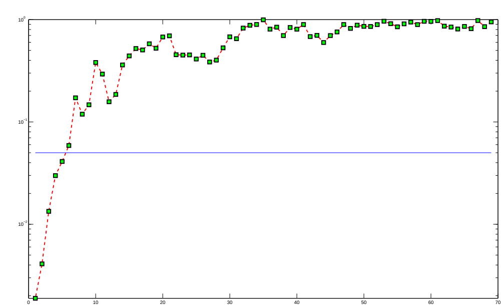

Figure 2: p values of the t test for the CSP boundary. On the dataset used the CSP occurs when $s < 6$.

5 Discussion and conclusion

Experimental results showed that the personality based USM performs significantly better than the rating based

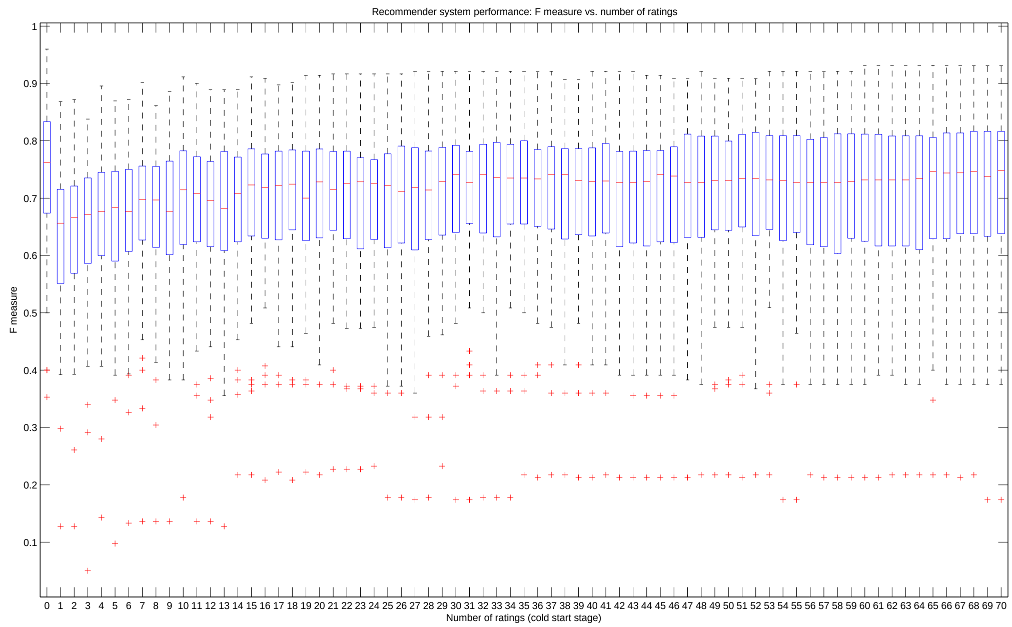

Figure 1: Boxplots of the distributions of F values at different cold start stages s. Note that stage $s = 0$ (the leftmost boxplot) represents the results of the personality based USM, which was independent of the number of ratings. Boxplots at stages $1 \leq s \leq 70$ represent the results of the rating based USM.

USM in cold start conditions. A positive outcome is also the fact that the personality USM is statistically equivalent to the rating based USM which makes it a good candidate for a complete replacement of the rating based USM.

The main drawback of the personality based USM is the difficulty of acquisition of end users' personality parameters. There are two main obstacles in this: (i) it is annoying for the end user to fill in questionnaires and (ii) the acquisition of personality data raises ethical and privacy issues that need to be addressed first. In our case, a controlled experiment has been conducted with users signing approval forms for the usage of their personality data for research purposes. The progress beyond the state of the art here is the knowledge that personality does account for between-users variance in entertainment applications.

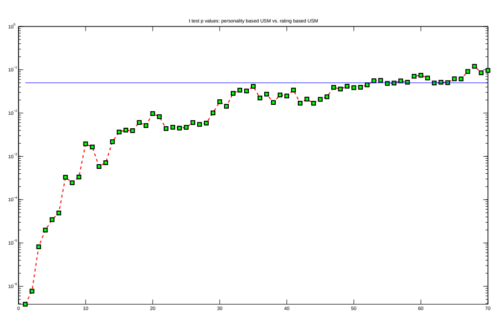

Figure 3: p values of the t test of the comparison of the personality based USM vs. rating based USM.

In the lack of existing methodologies for assessing the boundaries of the new user problem we chose a statistical approach. We acknowledge that further investigations should be conducted to determine how to test for the CSP boundary and that these investigations might conclude that a different approach is more suitable.

The number of nearest neighbours chosen to calculate the predicted ratings ($k = 7$) was chosen based on the number of users in the dataset ($J = 52$) and literature suggestions [9, 11]. As no methodology for the assessment of the correct k value was found, further investigations should be carried out to explain the impact of different k values on the results.

We provided a methodology for the assessment of the new user boundary. The results presented should not be taken for granted and several repetitions of the procedure should be carried out on different datasets.

In this paper we have evaluated a personality based USM under cold start conditions. The results showed that the personality based USM performed significantly better than the rating based USM. Furthermore we described a methodology for the assessment of the CSP border. Both novelties are important in the field of memory based collaborative filtering recommender systems and should be further explored.

References

[1] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems: A survey of

the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6):734–749, 2005.

[2] H.J. Ahn. A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178(1):37–51, 2008.

[3] I.A. Al Mamunur Rashid, D. Cosley, S.K. Lam, S.M. McNee, J.A. Konstan, and J. Riedl. Getting to know you: learning new user preferences in recommender systems. In *Proceedings of the 7th international conference on Intelligent user interfaces, January*, pages 13–16. Citeseer, 2002.

[4] K.W. Cheung and L.F. Tian. Learning user similarity and rating style for collaborative recommendation. *Information Retrieval*, 7(3):395–410, 2004.

[5] Lewis R. Goldberg. *Personality Psychology in Europe*, volume 7, chapter A Broad-Bandwidth, Public-Domain, Personality Inventory Measuring the Lower-Level Facets of Several Five-Factor Models, pages 7–28. Tilburg University Press, 1999.

[6] Jonathan L. Herlocker, Joseph A. Konnstan, Loren G. Terveen, and John T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactionson Information Systems*, 22(1), January 2004.

[7] Z. Huang, H. Chen, and D. Zeng. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. *ACM Transactions on Information Systems (TOIS)*, 22(1):116–142, 2004.

[8] Oliver P. John and Sanjay Srivastava. The big five trait taxonomy: History, measurement, and theoretical perspectives. In Lawrence A. Pervin and Oliver P. John, editors, *Handbook of Personality: Theory and Research*, pages 102–138. Guilford Press, New York, second edition, 1999.

[9] Matevž Kunaver, Tomaž Požrl, Matevž Pogačnik, and Jurij Tasič. Optimisation of combined collaborative recommended systems. *International Journal of Electronic Communications*, 61:433–443, 2007.

[10] P. Massa and B. Bhattacharjee. Using Trust in Recommender Systems: An Experimental Analysis. In *Trust management: second international conference, iTrust 2004, Oxford, UK, March 29-April 1, 2004: proceedings*, page 221. Springer-Verlag New York Inc, 2004.

[11] A.I. Schein, A. Popescul, L.H. Ungar, and D.M. Pennock. Methods and metrics for cold-start recommendations. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 253–260. ACM, 2002.

[12] Marko Tkalčič, Matevž Kunaver, Jurij Tasič, and Andrej Košir. Personality based user similarity measure for a collaborative recommender system. In C. Peter, E. Crane, L. Axelrod, H. Agius, S. Afzal, and M. Balaam, editors, *Proceedings of the 5th Workshop on Emotion in Human-Computer Interaction - Real world challenges*, pages 30–37. Fraunhofer Verlag, September 2009.

[13] Marko Tkalčič, Jurij Tasič, and Andrej Košir. The ldos-peraff-1 corpus of face video clips with affective and personality metadata. In *To appear in: Proceedings of the LREC 2010 Workshop on Multimodal Corpora: Advances in Capturing, Coding and Analyzing Multimodality*, 2010.