

Getting Cold Start Users Connected in a Recommender System's Trust Network

P. VICTOR, M. DE COCK and C. CORNELIS

Dept of Appl Math and CS, Ghent University, Gent, Belgium
E-mail: {Patricia.Victor,Martine.DeCock,Chris.Cornelis}@UGent.be

A. M. TEREDESAI

Institute of Technology, University of Washington, Tacoma, USA
E-mail: ankurt@u.washington.edu

Generating personalized recommendations for new users is particularly challenging, because in this case, the recommender system has little or no user record of previously rated items. Connecting the newcomer to an underlying trust network among the users of the recommender system alleviates this so-called cold start problem. In this paper, we study the effect of guiding the new user through the connection process, and in particular the influence this has on the amount of generated recommendations. Experiments on a dataset from Epinions.com support the claim that it is more beneficial for a newcomer to connect to an identified key figure instead of to a random user.

Keywords: trust network; recommender system; cold start problem

1. Introduction

Recommender systems (RSs) suggest items to users who might be interested in them, typically based on their user profiles or their previous actions.¹ One of the main challenges for RSs is the *cold start (CS) problem*: as new users have not yet rated a sufficient number of items, it is difficult to generate adequately personalized recommendations for them. Some RSs therefore actively encourage users to rate more items. The interface of the online DVD rental service Netflix^a for example explicitly hides two movie recommendations, and promises to reveal these after the user rates his most recent rentals. An alternative, complementary solution, is the use of a *trust network* in which the users of the RS are connected by trust scores indi-

^awww.netflix.com

cating how much they trust each other. Empirical evidence suggests that such a network can significantly help alleviate the CS problem, primarily because the trust information can be propagated:² in a *trust-enhanced RS*, users receive recommendations for items rated by people in their web of trust (WOT), or even by people who are trusted by WOT members, etc. Hence, by making a few clever connections in the trust network, newcomers can immediately gain access to a wide range of recommendations.

To get a recommendation, e.g. about which movie to rent, in real life it makes sense to consult people who have seen a lot of movies and have formed an opinion about them. In the context of a RS we call such users *frequent raters*, referring to the fact that they have rated a lot of items. Frequent raters are important suppliers for the RS, and connecting to them has a direct positive impact on the *coverage*, i.e. the amount of recommendations that can be generated. The main drawback is that they are not particularly visible to newcomers or, for that matter, to users of the RS in general: even if the system automatically identifies these *key figures* and suggests them as good connection points, it might still be hard for the user to decide whether or not to issue a trust statement. This situation is comparable with someone telling you which movies she liked, and to what extent, without explaining why. If you have not seen any of the movies, then this information is not particularly useful as you might very well have a very dissimilar taste.

The Epinions dataset³ used in Sect. 3 however contains another, more visible group of key figures. Epinions is a prominent e-commerce site where users can write reviews about products and assign a rating to them. We use the term *mavens* to refer to people who write a lot of such reviews. They are more visible than frequent raters, and are hence easier to evaluate: the more reviews someone writes, the better a new user can form an opinion on him and put him in his personal WOT or not.

It comes natural to assume that it is more beneficial for a new user to connect to one of these key figures as opposed to connecting to a random user. To quantify how much more beneficial, in Sect. 2 we introduce two local measures that reflect the impact of a user on the coverage for a newcomer. Sect. 3 contains the results from an experiment carried out on the Epinions dataset. The most important findings are summarized in Sect. 4.

2. Measuring the coverage impact

The web of trust of a user a , denoted by $WOT(a)$, is the set of users for which a issued a trust statement, i.e. $b \in WOT(a)$ iff there is an edge from

a to b in the trust network^b. Furthermore, let $Acc_0(a)$ denote the set of items rated by a , i.e., that are accessible from a in zero propagation steps. Through propagation, more items can become accessible; we use $Acc_1(a)$ to denote the items that are accessible in one step from a but not less. E.g., in the first scenario of Fig. 1, $WOT(a) = \{b_1, b_2\}$, $Acc_0(a) = \emptyset$, and $Acc_1(a) = \{i_1, i_2, i_3, i_4\}$. In terms of coverage, user b_1 is more important than user b_2 . Indeed, if we remove b_1 , then access to items i_1 and i_4 is lost, while with the omission of b_2 only item i_3 would not be reachable anymore. In other words, removing b_1 causes a higher loss of cohesion, resulting in a fragmented network where more items appear in isolated fragments. Hence we propose the following measure, which is a modification of the traditional fragmentation measure,⁴ for the impact that user b has on the coverage for user a .

Definition 2.1. (Fragmentation) Let a be a user and b a member of $WOT(a)$. The fragmentation of b for a is defined as $(F(b, a) \in [0, 1])$

$$F(b, a) = 1 - \frac{\sum_{j=1}^k s_j(s_j - 1)}{|Acc_1(a)| \cdot (|Acc_1(a)| - 1)}$$

in which k is the number of fragments after removing b from $WOT(a)$, and s_j is the number of items in the j th fragment.

The highest fragmentation score is reached when all items are in isolated fragments. Indeed, when $s_j = 1$, it holds that $s_j(s_j - 1) = 0$, hence there is no penalty in Def. 2.1. In the first scenario of Fig. 1, three fragments arise when removing b_1 , viz. $\{i_1\}$, $\{i_4\}$, and $\{i_2, i_3\}$, with $s_1 = 1$, $s_2 = 1$, and $s_3 = 2$ respectively. Hence $F(b_1, a) = 5/6$. On the other hand, $F(b_2, a) = 1/2$, which reflects that b_1 has more impact than b_2 .

The fragmentation score is however less suitable to compare between members of different WOT's. E.g., in the second scenario of Fig. 1, b_3 has rated one item that has not been rated by other WOT members. In this sense b_3 is as strong as user b_2 from the first scenario ($F(b_3, a) = 1/2$). In all fairness however, one should note that b_3 in the second scenario stands out among its peers, because it is the *only* user to have rated an item that is not rated by any other WOT user. In the first scenario on the other hand, user b_2 is outperformed by user b_1 in this respect. To reflect more truthfully how b compares relative to other members of a 's WOT, we therefore decrease $F(b, a)$ with the fragmentation score of the best WOT member besides b .

^bEpinions does not allow a graded trust score. A user can add other users either to his web of trust or to his blacklist. There are however far less distrust statements than trust statements (15% versus 85%), hence we focus on the latter.



Fig. 1. Fragmentation examples

Furthermore, although in theory $F(b, a)$ can range from 0 to 1, in practice its upper bound is determined by the behaviour of all users in $WOT(a)$, more in particular by the number of items that they rated in common: the more items are rated by more than one user, the harder it is for individual users to achieve a high fragmentation score. We call this practical upper bound the room for original contribution. It is defined by ($F^{max}(a) \in [0, 1]$)

$$F^{max}(a) = 1 - \frac{|Com(a)| \cdot (|Com(a)| - 1)}{|Acc_1(a)| \cdot (|Acc_1(a)| - 1)}$$

in which $Com(a)$ represents the set of items in $Acc_1(a)$ that are accessible through more than one user of a 's WOT. The following measure takes both considerations into account.

Definition 2.2. (Fragmentation utility) The fragmentation utility of user b for user a is defined as ($FU(b, a) \in [-1, 1]$)

$$FU(b, a) = \left(F(b, a) - \max_{u \in WOT_{-b}(a)} F(u, a) \right) / F^{max}(a)$$

in which $WOT_{-b}(a) = WOT(a) \setminus \{b\}$.

We already established that in scenario 2 of Fig. 1, $F(b_3, a) = 1/2$. Similarly, in this scenario, $F(b_1, a) = F(b_2, a) = 0$. Furthermore, $Com(a) = \{i_1, i_2, i_3\}$, hence $F^{max}(a) = 1/2$, and we obtain $FU(b_3, a) = 1$. In scenario 1 on the other hand, $Com(a) = \{i_2\}$. In this case $F^{max}(a) = 1$, which indicates that there is more room for original contribution than in scenario 2. Even though in absolute terms $F(b_1, a)$ from scenario 1 is higher than $F(b_3, a)$ from scenario 2, user b_3 exhibits a stronger behaviour as it filled the room for original contribution maximally while user b_1 from scenario 1 only managed to fill part of it, and received competition from b_2 . This is reflected in the lower score $FU(b_1, a) = 1/3$.

3. Results and discussion

To evaluate our approach, we use a large dataset³ from Epinions.com. One way of measuring the appropriateness of a key figure suggestion is to analyse the helpfulness scores that the CS user has given for the reviews written

by that key figure: high scores denote a successful recommendation, while low scores indicate that the key figure was not a good suggestion for that particular user. However, such an evaluation requires a dataset with temporal information. Instead, we analyse our technique by investigating the influence of the key figures on the coverage of CS recommendations. Therefore, we focus on the users who are not connected to any key figure and who evaluated from 1 (CS1) up to 4 reviews (CS4). We define a maven as someone who wrote at least 100 reviews (M-100+), and a frequent rater as someone who evaluated at least 2500 reviews (F-2500+). With these definitions, the community contains 1925 mavens, 1891 frequent raters, and 89 229 CS users. These CS users constitute a sizeable portion of the community, namely about 55% of all review raters.

For each user a in a CS group we calculate the difference $DFU(a)$ between the utilities $FU(b_1, a)$ and $FU(b_2, a)$, in which b_1 represents a randomly added key figure and b_2 a randomly added active user. Active users are those who rated at least one user or one item, hence this set contains key figures as well. In other words, DFU measures the extra utility benefit when connecting to a key figure instead of to a random user.

Fig. 2 depicts the average DFU when a specific key figure is added to the WOT. We experimented with several thresholds, a M-500 e.g. denotes a maven who wrote at least 500 and maximum 999 reviews. Connecting to a key figure is clearly more beneficial than connecting to a random user; on a scale from -1 to 1 , the FU of an added key figure is on average 0.41 higher than that of a randomly added user. This was expected for the frequent raters because without them there would only be few ratings and consequently far less recommendations. More surprising is the fact that mavens score very well too. This is especially useful for new users, because it is much easier to decide whether to trust the (more visible) mavens.

Fig. 2 also shows that, in general, the more active the key figure is, the more advantageous it is to have such a user in a WOT. For instance, users who are connected with a F-50000+ have a larger DFU than users connected with a F-2500. Note that the differences become larger for more

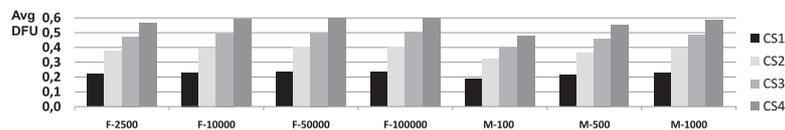


Fig. 2. Advantage of key figures over random users

active CS users. This is not because the utility of the key figure increases, but because the utility of randomly added users decreases more rapidly than the utility of key figures when the CS user rates more items.

4. Conclusions

Each key figure has its own characteristics; mavens are easy to evaluate and frequent raters provide a lot of ratings. We have shown that both have a large positive impact on the amount of the recommendations for CS users, and demonstrated that connecting to an identified key figure is more beneficial as opposed to connecting to a random user. However, to give a more complete picture, these results must be interpreted together with the accuracy of the recommendations. Preliminary results⁵ show that the absence or presence of a key figure in a WOT does not significantly affect the accuracy, but due to space restrictions we could not discuss this any further. Another research path is the investigation of the potential of connectors,⁵ i.e. users with many in- and outlinks in the trust network. Their strength lies not in their rating capacity or visibility, but in their ability to reach a large group of users through trust propagation, who, in turn, make more recommendations accessible. In this paper we only considered direct trust links; we refer to Ref. 5 for results obtained by one step trust propagation.

Acknowledgements

Patricia Victor would like to thank the Institute for the Promotion of Innovation through Science and Technology in Flanders, and Chris Cornelis the Research Foundation-Flanders, for funding their research. We thank Epinions for making the data available, in particular the authors of Ref. 3.

References

1. P. Resnick and H. Varian, Recommender systems, *Commun ACM* **40**, 56–58 (1997).
2. P. Massa and P. Avesani, Trust-aware collaborative filtering for recommender systems, *LNCS* **3290**, 492–508 (2004).
3. R. Guha, R. Kumar, P. Raghavan and A. Tomkins, Propagation of trust and distrust, in *Proc. of WWW2004*, 403–412 (2004).
4. S. P. Borgatti, Identifying sets of key players in social networks, *Comput Math Organiz Theor* **12**, 21–34 (2006).
5. P. Victor, C. Cornelis, A. Teredesai, and M. De Cock, Whom should I trust? The impact of key figures on cold start recommendations, in *Proc. of SAC2008*, 2014–2018 (2008).