Aggregation Trade Offs in Family Based Recommendations

Shlomo Berkovsky, Jill Freyne, and Mac Coombe

CSIRO Tasmanian ICT Center,
GPO Box 1538, Hobart, 7001, Australia
firstname.lastname@csiro.au

Abstract. Personalized information access tools are frequently based on collaborative filtering recommendation algorithms. Collaborative filtering recommender systems typically suffer from a data sparsity problem, where systems do not have sufficient user data to generate accurate and reliable predictions. Prior research suggested using group-based user data in the collaborative filtering recommendation process to generate group-based predictions and partially resolve the sparsity problem. Although group recommendations are less accurate than personalized recommendations, they are more accurate than general non-personalized recommendations, which are the natural fall back when personalized recommendations cannot be generated. In this work we present initial results of a study that exploits the browsing logs of real families of users gathered in an eHealth portal. The browsing logs allowed us to experimentally compare the accuracy of two group-based recommendation strategies: aggregated group models and aggregated predictions. Our results showed that aggregating individual models into group models resulted in more accurate predictions than aggregating individual predictions into group predictions.

1 Introduction

The quantity of potentially interesting information services available online has been growing rapidly and exceeds human processing capabilities. The vast amount of online information necessitates Web sites and portals to provide users with intelligent and personalized navigation support tools. These tools help users to identify the information items most relevant to them and filter out the rest by predicting the level of interest of users in particular information items. Collaborative filtering [9] is a statistical recommendation technique that can be applied to predict the interest level of a user in unvisited Web pages.

Collaborative filtering is commonly used in many online recommender systems to support users in selecting news items [2], courses [3], and many more [13]. The input for a collaborative filtering algorithm is a two dimensional matrix consisting of user models describing their preferences, interests, and information needs in the form of a feature vector. Collaborative filtering is based on the assumption that users with similar interests prefer similar information items [15]. In order

A. Nicholson and X. Li (Eds.): AI 2009, LNAI 5866, pp. 646–655, 2009.

to generate a recommendation, collaborative filtering initially compares the user models to identify users with the highest similarity to the current user and then generates predictions on items by calculating the normalized and weighted average of the opinions of the similar users[1].

One of the emerging practical problems of collaborative filtering recommender systems is the sparsity of user data [7], *i.e.*, the lack of sufficient information about users, which prevents the system from generating accurate and reliable predictions of interest in yet unseen information items. To partially resolve this problem and increase the accuracy of the generated recommendations, [8] proposed to aggregate the sparse individual user data into group-based data and then use the aggregated data in the collaborative recommendation process. Although in most conditions group-based recommendations cannot be as accurate as the personalized recommendations, they have the potential to be more accurate than general non-personalized recommendations, which are the natural fall back if the sparsity problem prevents the system from generating the personalized recommendations.

In this work we analyze family-based collaborative filtering recommendations – a particular case of group recommendations – using real life browsing data gathered in a study involving the users of an experimental eHealth family portal. We implemented several strategies that aggregated individual browsing logs into group-based data, generated collaborative filtering recommendations using the aggregated data, and then evaluated them against the observed browsing logs of the users.

The obtained experimental results demonstrate that group recommendations are superior to global and inferior to personalized recommendations. Also, we compared two aggregation strategies. The first aggregated the individual user models into group models and then applied collaborative filtering to the aggregated models. The second applied collaborative filtering algorithm to the individual user models and then aggregated the individual predictions into group predictions. The results show that aggregating the user models allows generating more accurate recommendations than aggregating the predictions.

Hence, the main contributions of this work are two-fold. Firstly, we evaluate the accuracy of collaborative filtering group recommendations and compare it to the accuracy of personalized and general recommendations. Secondly, we compare two strategies for the data aggregation: aggregation of browsing models and aggregation of predictions.

The rest of this paper is structured as follows. In section 2 we discuss related work on collaborative filtering and group recommendations. In section 3 we present and formulate the two aggregated group-based recommendation strategies. In section 4 we present our experimental settings, results and findings. Finally, in section 5 we conclude this work and present our future research directions.

---

[1] This presentation of collaborative filtering is narrowed down to user-to-user memory based approach. For a recent through survey of collaborative filtering algorithms the reader is referred to [14].

2 Collaborative Filtering and Group Recommendations

Collaborative filtering is one of the most popular and widely-used recommendation algorithms. It is based on the notion of *word of mouth* [15], which assumes that users who agreed in the past will agree in the future. In other words, it uses opinions of similar users to generate future predictions for a target user. The opinions of users on the items are expressed either as explicit ratings given by users according to a predefined scale or as implicit ratings accumulated and inferred through logging users' interactions with the system.

The main stages of the collaborative filtering recommendation generation process are: (1) recognizing commonalities between users by computing their inter-user similarities; (2) selecting the most similar users; and (3) generating recommendations by aggregating the opinions of the most similar users [9]. As it is being based on the similarities of users, the collaborative filtering process is sometimes referred to *people-to-people correlation*. In comparison with other recommendation algorithms, the main advantage of collaborative filtering over other algorithms is that it is not domain specific and independent of the representation of users and items. That is, a single collaborative filtering recommender systems can generate recommendations for any type of items (movies, images, or text) regardless of their content. As such, it is considered a universal technique applicable to a wide variety of domains and applications [13].

Collaborative filtering recommender systems suffer from the well-known sparsity problem [7]. It prevents the system from generating accurate predictions due to the insufficient data available about the users. Two particular cases of the sparsity problem can be differentiated: new user problem – the number of user ratings is insufficient for the identification of similar users and reliable generation of recommendations for that user [10], and new item problem – the number of item ratings is insufficient for a reliable generation of recommendations for that item [5].

In recent years the focus of collaborative filtering recommendation algorithms shifted from predictions for individuals to the more complex task of predictions for groups. To date, group recommendations were generated using one of the following three strategies: merging recommendations generated for individuals (very rare occurrence; will not be considered in this work), aggregating individual user models into group-based models, or aggregating them predictions for individual users into group-based recommendations [8].

The group modeling and aggregated predictions strategies differ in the timing of the aggregation of information in the recommendation process as illustrated in Figure 1. Specifically, group modeling strategy [4,16] aggregates individual user models of the group members *before* the prediction computation and then generates recommendations basing on the aggregated group model. Alternatively, aggregated predictions [11,12] treats group members as individuals for the prediction computation and *afterwards* aggregates the individual predictions to generate group recommendations.

As discussed in [8], the selection between the group modeling and aggregated prediction strategies depends on external factors, such as the ability to examine

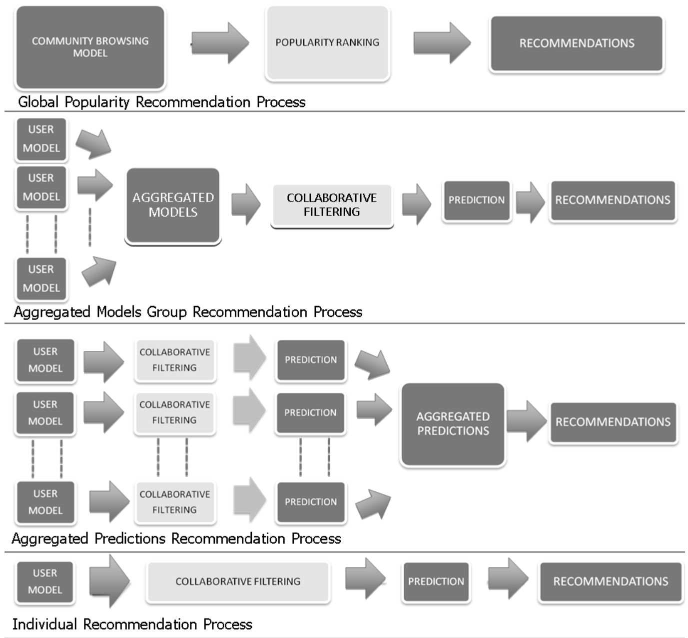

Fig. 1. Recommendation generation process

or negotiate group preferences, coverage of the recommendations, privacy considerations, and ability to explain the recommendations. However, to the best of our knowledge no prior work compared the *accuracy* of collaborative filtering recommendations generated using the above two strategies.

3 Prediction Strategies

The aim of this work is to determine which of the above two strategies for aggregating individual data and generating group-based recommendations is more appropriate when dealing with coherent groups consisting of individuals within a nuclear family structure. We concentrate on the following four recommendation strategies (see Figure 1). Our baseline strategy, *global popularity*, exploits the wisdom of the crowd at large and recommends the same most frequently visited items to all users. Our second and third strategies, *group modeling* and *aggregated predictions*, examine group-based recommendation algorithms and focus, respectively, on the group modeling and aggregated predictions strategies. Our fourth strategy is a standard *personalized collaborative filtering* recommendation algorithm. We will elaborately present these four strategies.

The *global popularity* strategy implements a simple social navigation mechanism [1], which guides users to areas of global interest. Each page p_i is assigned a predicted popularity score $pred(p_i)$ based on the number of times that it was visited across all users $u_x \epsilon U$ as shown in equation (1), where visit indicator $vis(u_x, p_i) = 1$ if u_x visited p_i and 0 otherwise.

$$pred(p_i) = \sum_{x \epsilon U} vis(u_x, p_i) \tag{1}$$

The *group modeling* strategy initially constructs a family based interest score $int(f_a, p_i)$ for family f_a and page p_i by aggregating the visit indicators $vis(u_x, p_i)$ of all family members u_x that belong to the family f_a according to their relative weight $\omega(u_x, f_a)$[2] as shown in equation (2).

$$int(f_a, p_i) = \frac{\sum_{u_x \epsilon f_a} \omega(u_x, f_a)\, vis(u_x, p_i)}{\sum_{u_x \epsilon f_a} \omega(u_x, f_a)} \tag{2}$$

Then, collaborative filtering is applied to the family model as shown in equation (3). Family based prediction is computed by assigning similarity degrees $sim(f_a, f_b)$ between the target family f_a and all other families $f_b \epsilon F$, and using these similarity degrees to aggregate the family based interest scores $int(f_b, p_i)$ in the target page p_i.

$$pred(f_a, p_i) = \frac{\sum_{f_b \epsilon F} sim(f_a, f_b)\, int(f_b, p_i)}{\sum_{f_b \epsilon F} sim(f_a, f_b)} \tag{3}$$

Finally, the computed family based prediction $pred(f_a, p_i)$ is assigned to all the users u_x that belong to the family f_a, *i.e.*, $pred(u_x, p_i \mid u_x \epsilon f_a) = pred(f_a, p_i)$.

The *aggregated prediction* strategy maintains an individual model for each user and initially generates individual predictions using the standard collaborative filtering recommendation algorithm as shown in equation (4). Prediction $pred(u_x, p_i)$ for user u_x and page p_i is computed by assigning similarity degrees $sim(u_x, u_y)$ between the target user u_x and all other users $u_y \epsilon U$, and using these similarity degrees to aggregate the individual visit indicators $vis(u_y, p_i)$ for the target page p_i.

$$pred(u_x, p_i) = \frac{\sum_{u_y \epsilon U} sim(u_x, u_y)\, vis(u_y, p_i)}{\sum_{u_y \epsilon U} sim(u_x, u_y)} \tag{4}$$

Then, the process becomes group focused. To generate a family based prediction $pred(f_a, p_i)$, the individual predictions $pred(u_x, p_i)$ for the family members are aggregated according to their relative weight $\omega(u_x, f_a)$ as shown in equation (5).

$$pred(f_a, p_i) = \frac{\sum_{u_x \epsilon f_a} \omega(u_x, f_a)\, pred(u_x, p_i)}{\sum_{u_x \epsilon f_a} \omega(u_x, f_a)} \tag{5}$$

Similarly to the previous strategy, the computed family based prediction $pred(f_a, p_i)$ is assigned to all the users u_x that belong to the family f_a.

[2] Uniform weighting is currently used to assign equal weight $\omega(u_x, f_a) = 1$ to all the users. Other weighting strategies will be investigated in the future.

The *personalized collaborative filtering* recommendation strategy examines the browsing patterns of individual users regardless of their membership in a family. For each user u_x, each page p_i is assigned a prediction score $pred(u_x, p_i)$ using the standard collaborative filtering algorithm as shown in equation (4) presented in the previous strategy.

In all four strategies we simplify the recommendation generation and recommend k unvisited pages having the highest prediction scores, *i.e.*, k pages that maximize the product $\prod_{i=1}^{k} pred(u_x, p_i)$. Note that the global popularity strategy generates one list of recommendations for all users, the two group-based strategies generate one list of recommendations for each family, and the personalized collaborative filtering produces one list for each family member.

4 Evaluation

The presented analysis was carried out through the browsing logs gathered as part of an eHealth family portal study. The aim of the analysis was to determine which strategy would be best to implement in a family based recommender in future versions of the portal. Specifically, we aimed to ascertain the differences (1) between the simple global popularity model, the aggregated family based models, and the personalized recommendation model, and (2) between the combined group model and the aggregated predictions strategies.

4.1 Experimental Setting

The data used was gathered over a two week period in March 2009. Members of the general public (families to be specific) were invited to take part in a study of family engagement with an eHealth application. The task for each family member was to visit the experimental eHealth portal, possibly browse the healthy living content, and submit suggestions for improving their lifestyles. A by product of the study was the capture of browsing activity for all the members of the involved families over the 23 portal pages.

In total, 64 users from 40 families took part in the trial. In 24 families only one person interacted with the portal, in 8 families two members interacted with the portal, in 2 families three members interacted with the portal, and in 6 families all four members interacted with the portal. In total 188 individual page visits and 151 aggregated family based interest levels were logged, yielding an individual matrix having 87.23% sparsity and a denser family based matrix having 83.59% sparsity[3]. Each user visited on average 2.94 pages (stdev=2.77) and each page was visited on average by 8.17 users (stdev=4.33).

The distribution of page visits across the users demonstrates a typical long tail distribution. Only 2 users visited more than 10 pages, 6 users visited between 5 and 10 pages, and 56 users visited less than 5 pages. Conversely, the distribution

[3] We disregard the families in which only one member interacted with the portal and exclude them from the evaluation. However, we do use these users' browsing logs as sources of recommendation content in the training set.

of page visits across the pages is more balanced. 7 pages were visited by more than 10 users[4], 7 pages were visited by between 5 and 10 users, and 9 pages were visited by less than 5 users.

For each user or family, a one off similarity matrix with other users or families was created using Pearson's Correlation similarity metric [9][5]. Using this similarity matrix, four recommendation lists were produced for each user using the four prediction strategies detailed in Section 3 (global, group modeling, aggregated prediction, and personalized collaborative filtering). A leave one out experimental evaluation was carried out to evaluate the performance of the algorithms. In particular, the accuracy of the recommendations was evaluated using the classification accuracy metrics of precision, recall, and F1 by comparing the recommendation lists with the actual logs of the users [6].

Let us denote by $\mathcal{V}$ the set of pages that were visited by a user (will be considered as the relevant pages) and by $\mathcal{R}$ the set of pages that were recommended by the system to the user. In this context, *precision* of the recommendations is computed by $\frac{|\mathcal{V}\cap\mathcal{R}|}{|\mathcal{R}|}$ and *recall* by $\frac{|\mathcal{V}\cap\mathcal{R}|}{|\mathcal{V}|}$. When the size of the recommended set $\mathcal{R}$ is limited to k, the computed precision metric is referred to as *precision@k*. Combining the two metrics of precision and recall yields a single metric, *F1 score*, which represents their harmonic mean assigning them equal weights. The F1 score is computed as

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

4.2 Experimental Results

The first question we posed related to the accuracy of recommendations based on the *global* strategy of all users versus smaller groups of users in *aggregated models* and *aggregated predictions* strategies versus individual activity in *personalized collaborative filtering* strategy. Table 1 shows the average precision, recall, and F1 scores obtained for each of the above recommendation strategies.

Table 1. Precision, Recall, and F-measure

	global	aggregated models	aggregated predictions	personalized
precision	0.219	0.300	0.235	0.534
recall	0.552	0.689	0.609	0.779
F1	0.314	0.418	0.339	0.633

It can be seen that, as expected, the personalized recommendation strategy outperformed all other strategies in terms of accuracy, returning the highest

[4] One of the pages was an outlier – it was visited by 20 users.
[5] Similar experimental results were obtained for the Cosine Similarity metric.

precision, recall, and F1 scores. Statistical significance tests[6] showed that the personalized strategy significantly outperformed both the group-based strategies. For precision we obtained $p = 4.36 \times 10^{-4}$ vs. aggregated models and $p = 1.44 \times 10^{-5}$ vs. aggregated predictions, and for recall we obtained $p = 2.32 \times 10^{-2}$ vs. aggregated models and $p = 2.72 \times 10^{-2}$ vs. aggregated predictions. Both group-based strategies outperformed the global strategy. For precision we obtained $p = 5.20 \times 10^{-2}$ vs. aggregated models and not statistically significant difference vs. aggregated predictions, and for recall we obtained $p = 5.45 \times 10^{-2}$ vs. aggregated models and not statistically significant difference vs. aggregated predictions.

Examining the whole recommendation lists and their accuracy is only one dimension of the recommendations' success. Precision@k measure analyzes the position of the visited pages within the recommendation lists. Figure 2 depicts the precision@k of the four recommendation strategies for gradually increasing from 1 to 7 values of k. Precision@k curves showed that the personalized strategy outperformed both the group-based strategies and the global strategy. For example, for $k = 1$ (the most strict metric focusing on the first recommended page) the personalized strategy achieved a precision of 74% in comparison to 38% and 44% for the two group-based strategies, and only 29% for the global non-personalized strategy. This observation was valid also for other values of k.

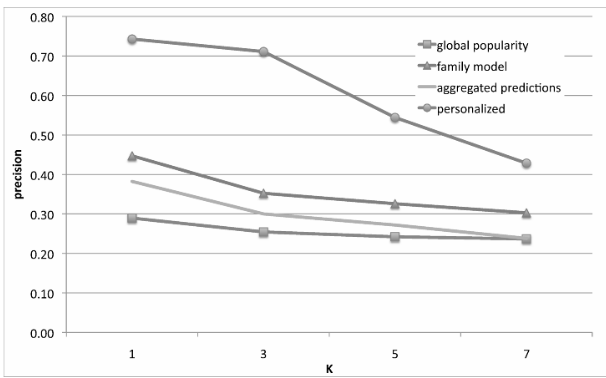

Fig. 2. Precision@k for various values of k

The second question we posed related to the comparative accuracy of the two group-based strategies. Both Table 1 and Figure 2 showed that the aggregated model strategy consistently outperformed the aggregated predictions strategy. For the overall precision and recall scores, the differences were statistically significant: $p = 1.79 \times 10^{-2}$ for precision and $p = 1.84 \times 10^{-2}$ for recall. We explain these findings by observing that aggregation of individual models yields a reasonably dense and accurate group model, which allows the system to generate

[6] All statistical significance results hereafter refer to a two-tailed t-test assuming equal variances.

reasonably accurate recommendations. Conversely, recommendations generated using the individual sparse models are inaccurate, such that their aggregation does not allow to improve their accuracy.

5 Conclusions and Future Work

The sparsity of user data is a well known problem of collaborative filtering recommender systems. To resolve it, the sparse individual user data can be aggregated into group-based data and these can then be used in the recommendation process. In this work we analyzed collaborative filtering family based recommendations using the browsing logs of the users of an experimental eHealth family portal. We implemented several strategies that aggregated individual data into group-based data, generated family based collaborative filtering recommendations using the aggregated data, and evaluated their accuracy against the observed browsing logs of the users.

Our empirical results showed that group recommendations were, as expected, more accurate than non-personalized one-size-fits-all recommendations at determining relevant Web pages. However, personalized collaborative filtering recommendations still outperformed group recommendations when comparing precision, recall, F1, and precision@k scores.

While previous works analyzed conditions when one group recommendation strategy would be preferred over another, this work experimentally compared the accuracy of two group-based aggregation strategies with real families of users of an eHealth portal. Our results consistently showed that aggregating individual browsing models into group models resulted in more accurate recommendations than aggregating the predicted interest levels. That is, generating recommendations using a dense group model was more accurate than aggregating the predictions generated for individual users.

In this work the users were assigned uniform weights when aggregating individual models into the family models. However, this is not reflective of the real setting, where different users may have different browsing patterns and frequencies. In the future we will evaluate the impact of various weighting heuristics on the accuracy of the recommendations. The results presented in this work are preliminary as they are supported by a reasonably small data set. In the future we will conduct a larger scale user study of online group recommendations, which will allow us to determine which strategies perform best under varying conditions such as richer user models, larger and more heterogeneous groups of users, and different content domains.

Acknowledgments

This research is jointly funded by the Australian Government through the Intelligent Island Program and CSIRO. The Intelligent Island Program is administered by the Tasmanian Department of Economic Development, Tourism and the Arts. The authors thank Nathalie Colineau, Cecile Paris, Dipak Bhandari, and Greg

Smith for their help with the development of the experimental eHealth family portal.

References

1. Brusilovsky, P., Chavan, G., Farzan, R.: Social adaptive navigation support for open corpus electronic textbooks. In: De Bra, P.M.E., Nejdl, W. (eds.) AH 2004. LNCS, vol. 3137, pp. 24–33. Springer, Heidelberg (2004)
2. Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., Sartin, M.: Combining content-based and collaborative filters in an online newspaper. In: Workshop on Recommender Systems: Algorithms and Evaluation (1999)
3. Farzan, R., Brusilovsky, P.: Social navigation support in a course recommendation system. In: Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 91–100 (2006)
4. Freyne, J., Smyth, B.: Cooperating search communities. In: Proceedings of the International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, pp. 101–110 (2006)
5. Gokhale, A., Claypool, M.: Correlation thresholds for more accurate collaborative filtering. In: Proceedings of the IASTED International Conference Artificial Intelligence and Soft Computing (1999)
6. Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems 22(1), 5–53 (2004)
7. Huang, Z., Chen, H., Zeng, D.: Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering. ACM Transactions on Information Systems 22(1), 116–142 (2004)
8. Jameson, A., Smyth, B.: Recommendation to groups. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 596–627. Springer, Heidelberg (2007)
9. Konstan, J.A., Miller, B.N., Maltz, D., Herlocker, J.L., Gordon, L.R., Riedl, J.: Grouplens: applying collaborative filtering to usenet news. Commun. ACM 40(3), 77–87 (1997)
10. Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering. IEEE Internet Computing 7(1), 76–80 (2003)
11. Masthoff, J.: Group modeling: Selecting a sequence of television items to suit a group of viewers. User Model. User-Adapt. Interact. 14(1), 37–85 (2004)
12. O'Connor, M., Cosley, D., Konstan, J.A., Riedl, J.: Polylens: a recommender system for groups of users. In: Proceedings of the European Conference on Computer Supported Cooperative Work, pp. 199–218 (2001)
13. Schafer, B.J., Konstan, J.A., Riedl, J.: E-commerce recommendation applications. Data Mining and Knowledge Discovery 5(1), 115–153 (2001)
14. Schafer, J., Frankowski, D., Herlocker, J., Sen, S.: Collaborative filtering recommender systems. In: Brusilovsky, P., Kobsa, A., Nejdl, W. (eds.) Adaptive Web 2007. LNCS, vol. 4321, pp. 291–324. Springer, Heidelberg (2007)
15. Shardanand, U., Maes, P.: Social information filtering: Algorithms for automating "word of mouth". In: Proceedings of the International Conference on Human Factors in Computing Systems, pp. 210–217 (1995)
16. Yu, Z., Zhou, X., Hao, Y., Gu, J.: Tv program recommendation for multiple viewers based on user profile merging. User Modeling and User-Adapted Interaction 16(1), 63–82 (2006)