Combining Sources of Information in a Collaborative Filtering Domain

1st Author	2nd Author	3rd Author
1st author's affiliation	2nd author's affiliation	3rd author's affiliation
1st line of address	1st line of address	1st line of address
2nd line of address	2nd line of address	2nd line of address
Telephone number, incl. country code	Telephone number, incl. country code	Telephone number, incl. country code
1st author's email address	2nd E-mail	3rd E-mail

ABSTRACT

In this paper, we describe a collaborative filtering approach that uses features of users and items to better represent the problem space and to provide better recommendations to users. Features of the collaborative filtering dataset are found and incorporated into a network representation of the collaborative filtering space where users and items are represented by nodes and where the nodes are connected by weighted edges. A spreading activation approach to collaborative filtering, using this representation, is compared with a traditional collaborative filtering approach.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *information filtering*.

General Terms

Design, Experimentation

Keywords

Collaborative Filtering, Spreading Activation

1. INTRODUCTION

There is much evidence from various domains within Information Retrieval that the combination of sources of evidence leads to more effective retrieval [1]. Although the types of information combined and the techniques used vary substantially across the domains it has been shown that there are advantages to be gained from considering more than one source of information. This paper considers additional information that can be found from the collaborative filtering data set itself and shows how, via a graph representation and a spreading activation approach, the incorporation of this information can aid in the collaborative filtering task.

2. METHODOLOGY

The collaborative filtering problem space is often viewed as a matrix consisting of the ratings given by each user for items in a collection. Using this matrix, the aim of collaborative filtering is to predict the ratings of a particular user for one or more items previously not rated by that user. Several researchers have adopted graph representations in order to develop recommendation algorithms [5], [6]. A variety of graphs have been used (e.g. directed and two-layer) and a number of graph algorithm approaches have been adopted (e.g. spreading activation [3]).

In this paper, a graph representation is used consisting of a set of user nodes and a set of item nodes (with user nodes and item nodes not directly connected to each other). Each user node and item node has an associated activity, output and threshold. User and item nodes are connected via weighted edges where the weights on the edges represent ratings given to items by users. The activity of a user or item node *a,* for *N* nodes connected to the node *a* with non-zero weight, is calculated by:

$$activity_a = \sum_{i=1}^{N} x_i w_i$$

where x_i is the output of the node *i* that is connected to node *a* and w_i is the weight on the edge connecting node *i* to node *a*. The output of a user or item node is calculated based on that node's activity and a threshold value. Spreading activation involves moving activation from one set of nodes to a second set of nodes.

Much implicit information about users and groups can be extracted from the collaborative filtering data set and can be represented in this graph model. The initial experimental work in this paper only considers two of the simpler user and item features: *rated* - the number of items rated by some user in comparison to the maximum and minimum number of items rated by users; and *avg-item-popularity* - the number of ratings received by some item in comparison to the maximum and minimum ratings received by items. Based on the values of these features certain user and item nodes are identified prior to filtering and these nodes are used to constrain the spreading activation approach thus resulting in only some of the edges being considered and only portions of the graph being traversed.

3. EXPERIMENTS

The experiment involves the comparison of two collaborative filtering approaches: a constrained spreading activation approach

and a traditional memory-based collaborative filtering approach. A standard subset of the Movie Lens dataset is considered that contains the ratings of 943 users for 1682 movies. Weights on the network edges indicate the strength of like/dislike for an item where 'dislike' can be viewed as an inhibitory or negative rating and 'like' can be viewed as an excitory or positive rating. Given that the original rating values in the Movie Lens dataset are all positive numbers the approach adopted maps the ratings to positive and negative values to indicate positive and negative influences. The mapping chosen is to subtract 2.5 from all non-zero values, giving:

$$\{0, 1, 2, 3, 4, 5\} \rightarrow \{0, -1.5, -0.5, 0.5, 1.5, 2.5\}$$

A proportion of the data set is removed for testing and the metric of precision is used to compare the performance of the two approaches at different recall points. Precision is used because the spreading activation approach returns a ranking of recommended items, not prediction values that can be compared with actual values, and also due to motivations presented in [2].

The traditional collaborative filtering approach uses Pearson Correlation to find correlated (similar) users. An adjustment is used in the Pearson Correlation calculation based on the number of items that users have rated in common (co-rated items) [4]. The "nearest" neighbours of a user are selected using a low neighbour selection threshold, with any correlation value greater than 0.01 being considered.

In the spreading activation approach to collaborative filtering, three stages corresponding to the three stages in the traditional memory-based collaborative filtering approach are used. Neighbours of some active user are found after activation has spread from the active user node to some set of item nodes (activating those items that the active user has rated) and from these item nodes to user nodes. At this stage the user nodes that have non-zero activity are the neighbours of the active user. When activation is spread again, from user nodes to item nodes, items not rated by the active user will be highlighted. These items are recommended to the user if the activity is sufficiently high. At any stage, a node may be constrained in spreading its activation if it has been previously identified based on the analysis of user and item features.

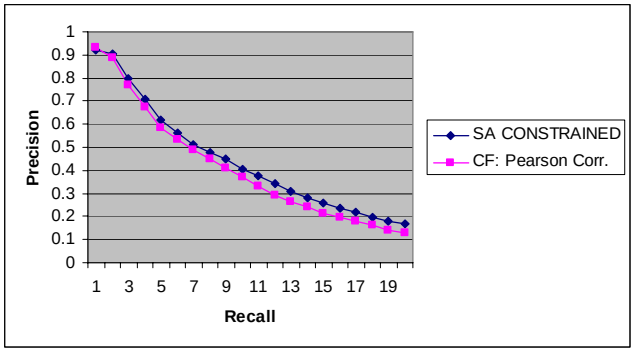

Figure 1: Constrained Spreading Activation and Pearson Correlation Approaches

4. RESULTS

Figure 1 illustrates the precision recall graph for the two approaches and precision values are given in Table 1. Results were averaged over 30 runs. It can be seen that even with the limited sources of additional information included in this representation, the constrained spreading activation approach outperforms the traditional memory-based approach at all recall points other than at the first returned recommendation. These results were shown to be statistically significant using a 2-tailed paired T-Test at p-value < 0.05.

Table 1. Constrained Spreading Activation and Pearson Correlation Approach (bold face shows statistically significant results at p-value < 0.05)

Recall Point	Constrained SA	CF: Pearson Correlation
1	0.9213	**0.933**
5	**0.6194**	0.585
10	**0.4022**	0.3688
15	**0.2579**	0.2136
20	**0.1668**	0.1273

5. CONCLUSIONS

We have presented a spreading activation approach to collaborative filtering using features of the dataset to constrain the activation approach. Results show that even incorporating information on simple features shows improved performance over a traditional memory based approach. Future work will consider more complex user features, as well as group features, and will show that the graph representation and algorithm outlined and demonstrated in this paper can easily be extended to incorporate additional sources of evidence.

6. REFERENCES

[1] Croft, W.B. *Combining Approaches to Information Retrieval*. Advances in Information Retrieval, Kluwer Academic Publishers, 2000.

[2] Herlocker, J.L., Konstan, J.A., Terveen, L.G. and Riedl, J.T. Evaluating Collaborative Filtering Recommender Systems, *ACM Transactions on Information Systems (TOIS), 22,(1),* 2004, 5-53.

[3] Huang, Z., Chung, W. and Chen, H. A Graph Model for E-Commerce Recommender Systems, *Journal of the American Society for Information Science and Technology, 55(3),* 2004, 259-274.

[4] McLaughlin, M.R. and Herlocker, J.L. *A Collaborative Filtering Algorithm and Evaluation Metric that Accurately Model the User Experience*, Proceedings of the 27th International *ACM SIGIR* Conference on Research and Development in Information Retrieval, 2004, 329-336.

[5] Mirza, B., Keller, B. and Ramakrishnan, N. Studying Recommendation Algorithms by Graph Analysis, *Journal of Intelligent Information Systems, 20(2)*, 2003, 131-160.

[6] Schwartz, M.F. and Wood, C. M. Discovering Shared Interests using Graph Analysis, *Communications of the ACM, 36(8)*, (August 1993), 78-89.