Offering Collaborative-Like Recommendations When Data Is Sparse: The Case of Attraction-Weighted Information Filtering

Arnaud De Bruyn[1], C. Lee Giles[2], and David M. Pennock[3]

[1] ESSEC Business School, Avenue Bernard Hirsch B.P. 105,
95021 Cergy-Pontoise Cedex, France
debruyn@essec.fr
[2] School of Information Sciences and Technology, Penn State University,
001 Thomas Building, University Park, PA 16802
giles@ist.psu.edu
[3] Yahoo! Labs, 74 N. Pasadena Ave., 3rd floor, Pasadena, CA 91103
david.pennock@overture.com

Abstract. We propose a low-dimensional weighting scheme to map information filtering recommendations into more relevant, collaborative filtering-like recommendations. Similarly to content-based systems, the closest (most similar) items are recommended, but distances between items are weighted by attraction indexes representing existing customers' preferences. Hence, the most preferred items are closer to all the other points in the space, and consequently more likely to be recommended. The approach is especially suitable when data is sparse, since attraction weights need only be computed across items, rather than for all user-item pairs. A first study conducted with consumers within an online bookseller context, indicates that our approach has merits: recommendations made by our attraction-weighted information filtering recommender system significantly outperform pure information filtering recommendations, and favorably compare to data-hungry collaborative filtering systems.

1 Introduction

Today, the ability to offer relevant recommendations to online visitors is a critical feature for most commercial websites, and many efforts have been dedicated to develop recommender systems suitable for this task.

Among those methods, *collaborative filtering* draws on an extensive database of consumers' ratings, preferences or past purchases to predict a visitor's affinity for items, based on comparisons to other consumers with similar tastes [1]. Another approach, *information filtering*, relies on item similarities and distance metrics across items to make recommendations.

The literature usually acknowledges the superiority of collaborative filtering over information filtering, while recognizing its vulnerability to information scarcity. Several algorithms have been proposed to combine the two [2-4], or to balance recom-

mendations made by both systems using statistical inference [5] or by measuring the relative data availability and effectiveness of the two methods [6]. Rarely, however, has the literature reported how collaborative and information filtering algorithms might yield fundamentally similar recommendations, and how this interdependency might be leveraged to improve the quality of the recommendations made when data is sparse.

Our goals in this paper are (a) to integrate information filtering and collaborative filtering into a unique, conceptual framework of spatial preferences, and (b) to draw on this framework to propose and test with real users a mapping procedure that is able to provide collaborative-like recommendations using an adapted version of an information filtering algorithm.

2   Information and Collaborative Recommendation Spaces

When one browses books at Amazon.com, for instance, recommendations appear to be closely related in terms of content to the current item; although the underlying recommender system is built on a collaborative filtering algorithm, it is more likely to recommend similar than dissimilar items, even though the algorithm does not draw on similarity metrics to make recommendations. However, anecdotal evidence also suggests that, regardless of items' similarities, collaborative filtering recommends certain items (i.e., bestsellers) more often.

Let's note Ω the space of information filtering recommendations and Ω' the space of collaborative recommendations. Each point in the space represents an item. Items' coordinates are similar in Ω and Ω', and similarly to Ω, closer items in Ω' are more likely to be recommended. However, we represent distance metrics in Ω' (noted d') as follows:

$$d'(X_i, X_j) = \frac{d(X_i, X_j)}{a_j}, a_j > 0 : \forall j, \tag{1}$$

where d represents the *distance* in Ω, and a_j represents an *attraction* index, an abstract measure of marketplace's preferences for X_j.

The measure d is not a proper distance metric anymore; if $a_i > a_j$, then $d'(X_i, X_j) > d'(X_j, X_i)$, and those items that have high attraction indexes are closer to all other items in the space, hence are more likely to be recommended.

If this model were an appropriate approximation of the truth, it would be possible to map an information filtering recommendation space into a collaborative recommendation space by choosing adequate measures of $a_i : \forall i$, while being less sensitive to information scarcity.

3   Empirical Investigation

We built a list of 6,000 business-related bestseller books, downloaded collaborative filtering recommendations made by Amazon.com, and built our own information fil-

tering recommender system, based on the *TFIDF* scheme [7]. We used the popular *cosine* distances to compute similarities among items.

For the purpose of this work, we computed the *attraction* indexes a_i as follows:

$$a_i = 1 + \ln\left(\frac{1 + \frac{c_i}{\bar{c}}}{1 + \frac{s_i}{\bar{s}}} \right), \tag{2}$$

where c_i is the number of times X_i was recommended by collaborative filtering; s_i is the number of items that are very similar to X_i, i.e., that pass a certain similarity threshold (≥ 0.6); and $\bar{c}$ and $\bar{s}$ are the average values of c_i and s_i in the database.

It follows that, everything being equal, the more an item is recommended by collaborative filtering, the higher its attraction index. The denominator is a correction for uniqueness.

We offered book recommendations to 28 undergraduate business students, based on an initial book selection, and asked them to rate each recommendation on a seven-point scale.

Recommendations included collaborative filtering, information filtering, attraction-weighted information filtering, and random recommendations (included as a control group), representing a total of 737 book recommendations. Fig. 1 shows the average ratings obtained for each recommender systems as a function of respondents' self-reported expertise.
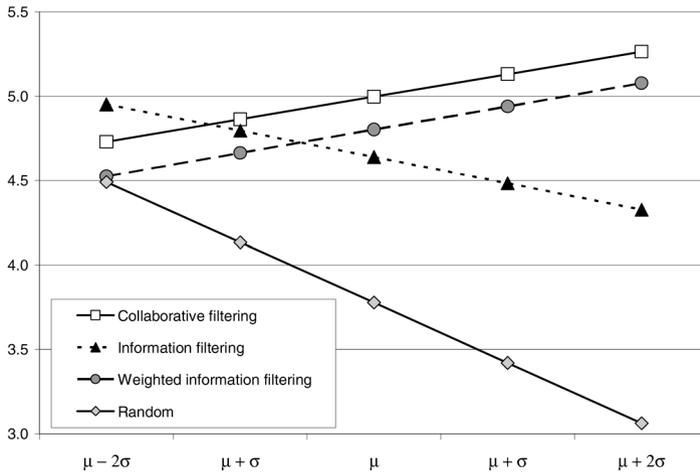
Fig. 1. Average ratings (regression lines) obtained by different recommender systems, as a function of respondents' expertise. μ=average expertise; σ=standard deviation

Respondents with very low expertise (i.e., μ-2σ, average expertise minus two standard deviations) give similar ratings to all methods: they cannot discriminate between thoughtful and random recommendations. Respondents with very high expertise (μ+2σ), on the other hand, strongly prefer collaborative and weighted information filtering recommendations, value pure information filtering recommendations significantly less, and give very low ratings to random recommendations. The fact that at-

traction-weighted information filtering is appreciated by expert respondents and achieves only marginally-lower ratings than collaborative filtering, is a very promising result, and seems to underline the added value of the mapping procedure we developed.

4   Discussion and Conclusions

We proposed a mapping procedure to conceptually link information filtering and collaborative filtering within a spatial model of preferences. Our assumption was that, by taking a spatial representation of similarity-based preferences (where the closest points represent the best recommendations), and by weighting each item by an appropriate *attraction* index to make the most preferred items 'closer' to all the others items in the recommendation space, it would be possible to improve information filtering recommendations. The results of a study conducted with real users showed promising results.

References

1. Resnick, Paul and Hal R. Varian (1997), "Recommender systems," Communications of the ACM, 40(3): 56–58.
2. Balabanovic, Marko and Yoav Shoham (1997), "Fab: Content-Based and Collaborative Recommendations," Communications of the ACM, 40 (3), 66-72.
3. Basu, Chumki, Haym Hirsh, and William Cohen (1998), "Recommendation as Classification: Combining Social and Content-Based Information in Recommendation," in Proceedings of the Fifteenth National Conference on Artificial Intelligence: AAAI.
4. Good, Nathaniel, J. Ben Schafer, Joseph A. Konstan, Al Borchers, Badrul Sarwar, Jon Herlocker, and John Riedl (1999), "Combining Collaborative Filtering with Personal Agents for Better Recommendations,".
5. Popescul, Alexandrin, Lyle H. Ungar, David M. Pennock, and Steve Lawrence (2001), "Probabilistic Models for Unified Collaborative and Content-Based Recommendation in Sparse-Data Environments," in Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann (Ed.). San Francisco.
6. Claypool, Mark, Anuja Gokhale, Tim Miranda, Pavel Murnikov, Dmitry Netes, and Mathew Sartin (1999), "Combining Content-Based and Collaborative Filters in an Online Newspaper," in Proceedings of ACM SIGIR Workshop on Recommender.
7. Salton, G. (1991), "Developments in Automatic Text Retrieval," Science, 253, 974-79.