Enhancing Collaborative Filtering with Demographic Data: The case of Item-based Filtering

Manolis Vozalis and Konstantinos G. Margaritis
Parallel Distributed Processing Laboratory
Department of Applied Informatics, University of Macedonia
Egnatia 156, P.O. 1591, 54006, Thessaloniki, Greece
E-mail: {mans,kmarg}@uom.gr
URL: http://macedonia.uom.gr/~{mans,kmarg}

***Abstract*— Demographic data regarding users and items exist in most available recommender systems data sets. Still, there has been limited research involving such data. This work sets the foundations for a novel filtering technique which relies on information of that kind. It starts by providing a general, step-by-step description of an approach which combines demographic information with existing filtering algorithms, via a weighted sum, in order to generate more accurate predictions. I-Demog is presented as an application of that general approach specifically on Item-based Collaborative Filtering. Several experiments involving different settings of the proposed approach support its utility and prove that it shows enough promise in generating predictions of improved quality.**

keywords: collaborative filtering, memory-based filtering, demographic data, personalization, prediction, recommender systems

I. Introduction

Recommender Systems were introduced as a computer-based intelligent technique whose purpose was to assist with the problem of information and product overload. They can be utilized to efficiently provide personalized services in most e-business domains, benefiting both the customer and the merchant.

Two basic entities are featured in all Recommender Systems: the *user* and the *item*, where m is the number of users, and n is the number of items. A user who wishes to take advantage of the Recommender System, called *active user*, should provide his opinion about a variety of items. The *goal* of the Recommender System would be to generate suggestions about new items for that particular user. The process is based on the *input* provided, and the *filtering algorithm*, which is applied on that input. In the majority of cases, the input for recommender systems consists of ratings on past items. An alternative kind of input is the demographic data regarding the user or the item. *Demographic data*, which will be the focal part of this paper, refers to information such as the age, the gender and the education of the user, or to product information regarding the item in mind. This kind of data is usually difficult to obtain and is normally collected explicitly by asking the user or manually from product catalogues.

Recommender systems are usually distinguished as belonging to one of two wide categories: *Memory-based* or *Model-based Systems*. Memory-based Systems are more efficient as a result of their producing recommendations without a need for any preprocessing. Still, they suffer from serious scalability problems. User-based Collaborative Filtering [1] [2] and Content-based Filtering [3] are both examples of Memory-based filtering. A different approach is taken by Model-based Systems [4] [5]. These algorithms develop a model of user ratings in order to produce their predictions. The construction of that model requires time, but once created, it speeds up considerably the generation of the recommendations. Model-based algorithms often approach the filtering process from a probabilistic perspective [6] [7]. A recent representative of this class, utilized in this paper, is Item-based Collaborative Filtering [8]. This method bases its prediction on the construction of neighborhoods of similar items.

Demographic Recommender Systems rely on user or item attributes, classified as demographic data, in order to produce their recommendations. Their functionality is sometimes enhanced by the utilization of pre-generated demographic clusters. Demographic Generalization, the approach developed by Krulwich [9] and applied in his Lifestyle Finder, suggested items and web pages to users, after assigning them to the appropriate clusters. The classification was made possible with the help of a commercially available database of demographic data and the input provided by the user, which constituted his profile. Pazzani [10] attempted to find regularities among users, by applying on them the Winnow algorithm, an algorithm originally designed for text classification. The user profiles utilized for similarity calculations had the form of user demographic vectors.

Hybrid systems [11] combine different filtering techniques in order to produce improved recommendations. Fab [12] utilizes collaborative filtering techniques to compare user profiles, which were generated by content analysis. Content-Boosted Collaborative Filtering [13] improves on user-based collaborative filtering predictions by enhancing the initial matrix of ratings through the application of a content-based predictor. P-Tango [14] combines different filtering methods, by first relating each of them to a distinct component and then basing its predictions on the *weighted average* of the predictions generated by those components. GroupLens introduce the use of filterbots [15] [16], which are automated rating agents that evaluate new items based on the rating algorithm that they carry. Smyth and Cotter implement their PTV system [17] as a mixed hybrid, by presenting together recommendations, generated by a collaborative filtering and a content-based component. Basu, Hirsh and Cohen present Ripper [18] as an inductive learning system that learns rules by utilizing both content and collaborative features. Condliff et al. propose a two stage mixed-effects Bayesian Recommender [19], which first fits a Bayes classifier to each user, utilizing item features, and then links those classifiers across different users through a regression model.

In this paper we present the *Demog algorithm*, a novel filtering approach which can utilize a filtering algorithm as its base and extend it by exploring the usefulness of demographic data as an enhancing factor. According to the Burke's taxonomy [11], our approach can be classified as a *Feature combination* hybrid, since features from different recommendation data sources are blended into a single recommendation algorithm. The base filtering algorithm selected for this work is Item-based Collaborative Filtering. The item neighborhoods, as generated by this algorithm, are further refined, by the help of demographic correlations, before final predictions are produced.

This paper can be outlined as follows. Section II. provides information about the utilized data set and the evaluation metrics, two issues closely related to the success of the implemented experiments. Section III. presents an overview of Item-based Collaborative Filtering, the base filtering algorithm involved in our experiments. Section IV. starts by describing the general structure of the Demog Algorithm. It does so in a formal, step-by-step manner. Section IV. proceeds by applying the Demog Algorithm on the selected base method. The utility of the resulting algorithm, *I-Demog*, is supported by the execution of several experiments and a presentation of their results. The paper is concluded in Section V. where an overall evaluation of the Demographic algorithm is attempted, while pointers for future research are outlined.

II. Experimental Methodology

For the execution of our subsequent experiments we utilized the data publicly available from the GroupLens movie recommender system. The MovieLens data set, used by several researchers [20] [21] [22], consists of 100.000 ratings which were assigned by 943 users on 1682 movies. Users should have stated their opinions for at least 20 movies in order to be included. Ratings follow the 1(bad)-5(excellent) numerical scale. Starting from the initial data set five distinct splits of training and test data were generated. For each data split, 80% of the original set was included in the training and 20% of it was included in the test data. The test sets in all cases were disjoint.

Several techniques have been used to evaluate Recommender Systems. Those techniques have been divided by Herlocker et al. [22] into three categories. The first category includes *Predictive Accuracy Metrics*, such as mean absolute error, mean squared error and normalized mean absolute error. These metrics measure how close the recommender's predictions are to the true user ratings. The second category, *Classification Accuracy Metrics*, includes methods such as ROC curves and F1, and measures how often a recommender system can decide correctly whether an item is beneficial for the user and therefore should be suggested to him. Those metrics require a binary classification of items into useful and not useful. The last category of metrics, called *Rank Accuracy Metrics*, measure the proximity of a predicted ordering of items, as generated by a recommender system, to the actual user ordering of the same items.

The choice among those metrics should be based on the selected user tasks and the nature of the data sets. P-Tango [14] calculates its efficiency via the deviation of generated predictions from the user-specified ratings. As a result the selected evaluation metric is *inaccuracy*, an alternative name for mean absolute error. On the other hand, Breese et al. [4] prefer to evaluate a ranked list of recommended items and therefore they calculate the expected utility of that list to the user and use that amount as their metric. We wanted our proposed algorithms to derive a predicted score for already rated items rather than generate a top-N recommendation list. Based on that specific task we proceeded in the selection of the initial evaluation metric for our experiments. That metric was *Mean Absolute Error (MAE)* [2]. It is a statistical accuracy metric which measures the deviation of predictions, generated by the Recommender System, from the true rating values, as they were specified by the user. MAE is measured only for those items, for which a user has expressed his opinion. The second metric utilized was *Coverage* [23]. It measures the percentage of items for which a filtering algorithm can generate predictions. It is not a rare occurrence that a Recommender System will not be able to provide a prediction for specific items because of the sparsity in the data of the initial user-item matrix, R, or because of other restrictions, which are set during the Recommender System's execution. Coverage and MAE complement each other.

III. The Base Algorithm

In this section we will briefly discuss the filtering algorithm which will be utilized by the proposed algorithm.

A. Item-based Collaborative Filtering

Similarly to User-based Collaborative Filtering, Item-based Filtering is based on the creation of neighborhoods. Yet, unlike the User-based Collaborative Filtering approach, those neighbors consist of similar items rather than similar users [24].

The execution steps of the algorithm are (a) *Data Representation* of the ratings provided by m users on n items. This step is based on the construction of an mxn user-item matrix, R. (b) *Neighborhood Formation*, which concludes by the construction of the active item's neighborhood. Similarities for all possible pairs of items existing in the data set are computed by the application of the preferred similarity metrics. Items most similar to the active item, which refers to the item for which predictions should be generated, are selected for its neighborhood. (c) *Prediction Generation*, where final predictions are calculated as a weighted sum of ratings given by a user on all items included in the active item's neighborhood.

IV. The Demographic Algorithm

We will present a hybrid algorithm that keeps the core ideas of existing recommender systems and enhances them with relevant information extracted from demographic data. In the subsequent experiments, Item-based Collaborative Filtering will be the base filtering algorithm of choice. Still, initial experiments have proven that our technique can be successfully applied on User-based Collaborative Filtering.

In the approach taken by Pazzani [10], user profiles were expressed as vectors constructed solely from demographic data and similarities among those user profiles were calculated in order for final predictions to be generated. In other words, the only source of user data utilized for recommendations was the demographic information available for them, while the ratings awarded by the same users for past items were disregarded.

In our proposal, similarly to what is known from plain User-based [25] and Item-based Collaborative Filtering [24], user and item correlations, based exclusively on past ratings, lead to the construction of neighborhoods. Still, before these neighborhoods should be utilized for the generation of final predictions, the correlations between neighborhood members and active users or items are re-evaluated, this time by also taking into account existing demographic correlations. An outline of our approach is the following:

- *Step 1*: Construct demographic vectors for the m users and n items that participate in the recommendation process. The information required for those vectors are usually collected explicitly and can be found in most collaborative filtering data sets, like MovieLens and EachMovie.
- *Step 2*: Execute the first two steps of the selected filtering algorithm, that is Data Representation and Neighborhood Formation. Neighborhood Formation will utilize ratings-based correlation metrics and generate a neighborhood of users/items similar to the active user/item, ui_a.
- *Step 3*: Calculate the demographic correlation between the active user/item, ui_a, and each of the members of its neighborhood. Demographic correlation is defined by the similarity of the vectors which represent specific users or items. So, demographic correlation, dem_cor_{ai}, between active user/item, ui_a and a member of its neighborhood, ui_i, can be calculated as follows:

  $$dem\_cor_{ai} = vect\_sim(\overrightarrow{ui_a}, \overrightarrow{ui_i}) = \frac{\overrightarrow{ui_a} \cdot \overrightarrow{ui_i}}{||\overrightarrow{ui_a}||_2 * ||\overrightarrow{ui_i}||_2} \quad (1)$$

  where "$\cdot$" denotes the dot-product of the two vectors.
- *Step 4*: Calculate the **Enhanced Correlation**, enh_cor_{ai}, for every pair of the form $\{ui_a, ui_i\}$, where ui_a is the active user/item and ui_i is a member of its neighborhood. Enhanced Correlation can be thought as incorporating the contributions of the ratings-based correlation and the newly acquired demographic correlation. It can be calculated via a weighted sum of the following form:

  $$\begin{aligned} enh\_cor_{ai} = & \\ & \alpha * rat\_cor_{ai} + \beta * dem\_cor_{ai} \\ & + \gamma * (rat\_cor_{ai} * dem\_cor_{ai}) \end{aligned} \quad (2)$$

  where rat_cor_{ai} and dem_cor_{ai} represent the ratings-based and the demographic correlation between active user/item ui_a and neighborhood member ui_i, while α, β and γ are flags that define the participation of each of the three components.
- *Step 5*: Proceed with the final step of the recommendation procedure, which is Prediction Generation. Regardless of the selected filtering algorithm (User-based or Item-based Collaborative Filtering), which is reflected by the prediction generation formula to be applied, our demographic approach differs from the one set by the base algorithms: Specifically, in the prediction generation formula it replaces the ratings-based correlation, rat_cor_{ai}, between the active user/item, ui_i, and any of the members of its neighborhood, ui_i, by their enhanced correlation, enh_cor_{ai}.

In the following paragraphs we will describe how this general approach can be applied specifically in the case of Item-based Collaborative Filtering, enhancing its predictions, and, depending on the various parameter settings, lead to more accurate recommendations.

A. I-Demog: Enhancing Item-based Collaborative Filtering with Demographic Correlations

We will now give a detailed description of how our demographic approach can be applied on Item-based Collaborative Filtering. The execution steps of *I-Demog* can be outlined as follows:

- *Step 1*: Construct the demographic vectors for the n items which participate in the recommendation process.

  For this step we are taking into account the demographic information available from the MovieLens data set for items, which in our case correspond to films rated by users. The MovieLens data set distinguishes 18 distinct film genres, ranging from Children's to Horror. This results to a vector of 19 features, $itdemog[19]$, if we utilize an additional slot, rightfully called "unknown", especially for films that cannot be categorized under any of the existing genres. It is important to point out that a film can belong to more than one genres at the same time. For example, it can be a Comedy, a Children's flick and a Musical. In that case, the slots which correspond to each of these categories should take a value of 1 (True), with the rest staying fixed at 0 (False).
- *Step 2*: Execute the first two steps of Item-based Collaborative Filtering, that is Data Representation and Neighborhood Formation.

  The Neighborhood Formation step of the Item-based

Filtering algorithm concludes with the construction of the active item's neighborhood. This neighborhood includes the l items, i_k, with $k = 1, 2, ..., l$, which are most similar to the active item, i_j, according to the selected ratings-based correlation metric. The correlation metric of our choice is the *Adjusted Cosine Similarity*, which performs better than other proposed metrics, based on previously described experiments [24] [26].

- *Step 3*: Calculate the demographic correlation between the active item, i_j, and each of the members of its neighborhood.

  For the execution of this step we are required to isolate the demographic vectors only for those items included in the active item's neighborhood. We can now compute the correlation between the active item, i_j, and its neighborhood items, i_k, with $k = 1, 2, ..., l$. Their demographic correlation, dem_cor_{jk}, should be calculated by applying the vector similarity formula on the corresponding vectors.

- *Step 4*: Calculate the Enhanced Correlation, enh_cor_{jk}, for every pair of the form $\{i_j, i_k\}$, where i_j is the active item and i_k is a member of its neighborhood.

  The Enhanced Correlation for Item-based Collaborative Filtering, which unites the outcomes from Adjusted Cosine Similarity and Demographic Correlation for each $\{i_j, i_k\}$ pair, can be calculated from the following formula:

$$enh\_cor_{i,jk} = \alpha * adj\_cor_{jk} + \beta * dem\_cor_{jk} + \gamma * (adj\_cor_{jk} * dem\_cor_{jk}) \quad (3)$$

  where adj_cor_{jk} is the Adjusted Cosine Similarity for items i_j and i_k, and dem_cor_{jk} is their demographic correlation. α, β and γ are flags that define the participation of each of the three components in the final result. In the sections that follow, we report the results from testing 4 distinct combinations of values for these flags, which correspond to different ways of combining ratings-based and demographic correlations.

- *Step 5*: Proceed with the final step of the recommendation procedure, which is Prediction Generation.

  Plain Item-based Filtering computes the predicted rating of user u_a on item i_j as follows:

$$pr_{aj} = \frac{\sum_{k=1}^{l} r_{ak} * adj\_cor_{jk}}{\sum_{k=1}^{l} |adj\_cor_{jk}|} \quad (4)$$

  r_{ak} represents the documented ratings of active user u_a for items i_k, with $k = 1, 2, ..., l$, while adj_cor_{jk} is the Adjusted Cosine Similarity, based on past ratings, between the active item, i_j, and the items from its neighborhood, i_k, with $k = 1, 2, ..., l$.
  Our approach modifies the prediction generation formula as follows:

$$idem\_pr_{aj} = \frac{\sum_{k=1}^{l} r_{ak} * enh\_cor_{jk}}{\sum_{k=1}^{l} |enh\_cor_{jk}|} \quad (5)$$

  Clearly, the only difference between the two prediction generation formulas lies in the use of the enhanced correlation, instead of the ratings-based correlation which is utilized in the original algorithm. The enhanced correlation, as defined in Step 4, incorporates the contributions from both ratings-based and demographic correlations.

1) Comparing 4 implementations of I-Demog with Item-based Collaborative Filtering: In the following paragraphs we will discuss 4 alternative I-Demog implementations. Each implementation tests different values for the flags (α, β and γ) of the enhanced correlation formula 3. By these experiments we intend to evaluate distinct combinations of the ratings-based and demographic correlations, while documenting the impact of the demographic correlations in the final predictions. The basic attributes of the implementations involved in our experiments, as expressed by the values of the flags and the corresponding enhanced correlation formulas, are summarized in Table I.

Fig. 1 compares the Mean Absolute Errors (MAE) collected from our I-Demog implementations (*i-demog1,2,3,4*) and Item-based Collaborative Filtering (ib), for neighborhoods with varying sizes of 20-140 items. In both cases, we tested the behavior of the algorithms for a single threshold value, *ut= 10*, meaning that at least 10 users should have rated the active item and a random item, in order for the latter to be included in the active item's neighborhood. The results reported here were averaged over all 5 data splits of the MovieLens data set.

Table II collects the lowest MAE values, among all neighborhood sizes tested, for the 4 I-Demog implementations and for the Item-based Collaborative Filtering approach.

A careful look at Fig. 1 and the best accuracy values achieved by the tested methods leads us to the following conclusions:

- The error values from the tested approaches can be thought as belonging to 3 distinct clusters whose members display similar accuracy behavior. Specifically: The first cluster includes *I-Demog3* and *I-Demog4*, the two implementations with the best overall accuracy results. The second cluster includes *I-Demog2* and Item-based Filtering, with MAE values worse than those of the previous cluster. Within the cluster, *I-Demog2* displays a slightly better behavior when compared to Item-based Filtering. The final cluster includes only *I-Demog1*, which generated the worse accuracy results in our experiments.
- The poor performance of *I-Demog1* can be attributed to the decisive role assigned to demographic correlation by the specific approach. Regarding the significant accuracy difference between *I-Demog1* and the remaining I-Demog implementations we have to make two related notes:
  1) Demographic vectors, constructed by the distinct film genres that apply to a specific item, are difficult

TABLE I

BRIEF DESCRIPTION OF I-DEMOG IMPLEMENTATIONS

	flags	enhanced correlation
I-Demog1	$\alpha = 0$, $\beta = 0$, $\gamma = 1$	$enh\_cor = adj\_cor * dem\_cor$
I-Demog2	$\alpha = 1$, $\beta = 0$, $\gamma = 1$	$enh\_cor = adj\_cor + adj\_cor * dem\_cor$
I-Demog3	$\alpha = 1$, $\beta = 1$, $\gamma = 1$	$enh\_cor = adj\_cor + dem\_cor + adj\_cor * dem\_cor$
I-Demog4	$\alpha = 1$, $\beta = 1$, $\gamma = 0$	$enh\_cor = adj\_cor + dem\_cor$

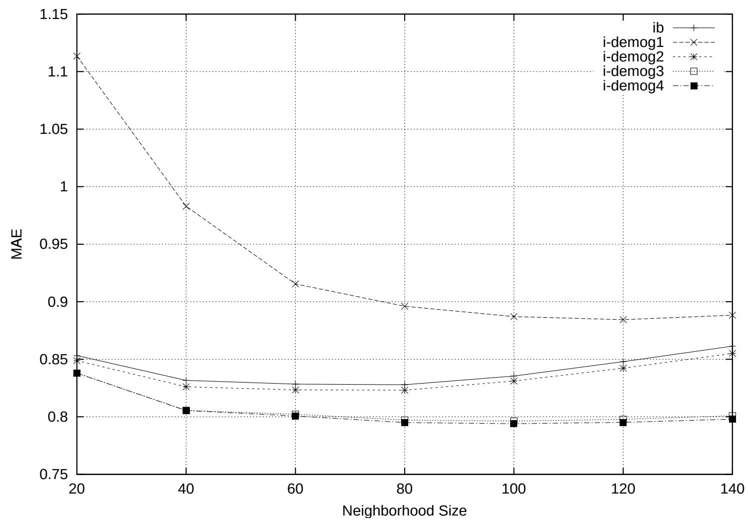

Fig. 1. I-Demog1: Average MAE for various neighborhood sizes

TABLE II

BEST MAE VALUES OF I-DEMOG IMPLEMENTATIONS AND ITEM-BASED FILTERING

	ib	i-demog1	i-demog2	i-demog3	i-demog4
MAE	0,8279	0,8844	0,8231	0,7964	0,7939

to coincide, making it common for the corresponding items to have a demographic correlation of 0. Based on the definition of enhanced correlation for *I-Demog1*, we can easily conclude that demographic correlations equal to 0 will result to a considerable shrinking of the participating item neighborhood.

2) Our experiments with I-Demog implementations start with a neighborhood that includes 20 items. This neighborhood size is small and will become even smaller if we take into account all the couples of items with demographic correlation equal to 0. This leads to the poor performance of *I-Demog1* for small neighborhoods and the time it takes for demographic correlations to have an actual impact on the final predictions.

- The remaining I-Demog implementations (*I-Demog2*, *I-Demog3* and *I-Demog4*) assign to demographic correlation a complementary role, reducing its impact in the prediction generation process. The improvement in the accuracy results proves that such a role fits perfectly to demographic correlation. Furthermore, while the improvement of *I-Demog2* over Item-based Filtering is small, the MAE values achieved by *I-Demog3* and *I-Demog4* are significantly better than those of Item-based Filtering. Based on the experimental results for *I-Demog3* and *I-Demog4* we can attribute the performance improvement observed in both cases to the existence of the component $adj\_cor + dem\_cor$ in the enhanced correlation prediction formula.
- As documented by the performance comparison between I-Demog implementations and plain Item-based Collaborative Filtering, the involvement of demographic information about items in the recommendation process can possibly lead to a measurable accuracy improvement. Still, as proven by the implementations we experimented with, there should be a careful selection regarding the role of demographic correlation, which in all cases should merely complement ratings-based correlation.

V. CONCLUSIONS

In this work we have presented a unique filtering approach which draws ideas from existing algorithms and combines them with demographic information available in recommender systems data sets. That general filtering approach was used to demographically enhance Item-based Collaborative Filtering, leading to I-Demog. We tested several degrees of demographic data involvement in the recommendation process, expressed by varying parameter settings for I-Demog.

The experimental results supported the ability of our demographically enhanced approach to generate predictions which are more accurate than those of the base filtering algorithm. There was only one case where our technique showed a performance degradation when compared to Item-based Filtering. In this case, demographic information did assume a major role in the recommendation process. In all other implementations, with demographic information assuming just a complementary role, the accuracy of our proposed approach displayed an improvement over the base filtering algorithm. The extent of this improvement was significant for specific parameter settings of the I-Demog algorithm.

These results prove that the demographic information existing in the MovieLens data set do not hold enough data about the items in order to capture their distinguishing features and generate accurate predictions, when utilized just by themselves. Still, when combined appropriately with other forms of filtering, such as collaborative filtering, they can enhance the recommendation process and lead to improved predictions.

The MovieLens data set includes demographic data regarding the users who participated in the rating procedure. Consequently, the next step of testing our Demog approach would be to apply it on User-based Collaborative Filtering in order to see how the resulting demographically enhanced method will perform.

REFERENCES

[1] P. Resnick, N. Iacovou, M. Sushak, P. Bergstrom, and J. Riedl, "Grouplens: An open architecture for collaborative filtering of netnews," in *ACM 1994 Conference on Computer Supported Cooperative Work*, New York, NY, 1994, pp. 175–186.

[2] U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating 'word of mouth'," in *Proceedings of Computer Human Interaction*, 1995, pp. 210–217.

[3] G. Salton and C. Buckley, "Term weighting approaches in automatic text retrieval," *Information Processing and Management*, vol. 24, no. 5, pp. 513–523, 1988.

[4] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical analysis of predictive algorithms for collaborative filtering," in *Fourteenth Conference on Uncertainty in Artificial Intelligence*, Madison, WI, 1998.

[5] D. Billsus and M. J. Pazzani, "Learning collaborative information filters," in *15th International Conference on Machine Learning*, Madison, WI, 1998.

[6] Y.-H. Chen and E. I. George, "A bayesian model for collaborative filtering," in *Proceedings of the Seventh International Workshop on Artificial Intelligence and Statistics*, 1999.

[7] T. Hofmann and J. Puzicha, "Latent class models for collaborative filtering," in *Proceedings of the International Joint Conference in Artificial Intelligence*, 1999.

[8] M. Deshpande and G. Karypis, "Item-based top-n recommendation algorithms," *ACM Transactions on Information Systems*, vol. 22, pp. 143–177, 2004.

[9] B. Krulwich, "Lifestyle finder: Intelligent user profiling using large-scale demographic data," *Artificial Intelligence Magazine*, vol. 18, pp. 37–45, 1997.

[10] M. Pazzani, "A framework for collaborative, content-based and demographic filtering," *Artificial Intelligence Review*, vol. 13, pp. 393–408, 1999.

[11] R. Burke, "Hybrid recommender systems: Survey and experiments," *User Modeling and User-Adapted Interaction*, vol. 12, pp. 331–370, 2002.

[12] M. Balabanovic and Y. Shoham, "Fab: Content-based, collaborative recommendation," *Communications of the ACM*, vol. 40, 1997.

[13] P. Melville, R. J. Mooney, and R. Nagarajan, "Content-boosted collaborative filtering," in *ACM SIGIR Workshop on Recommender Systems*, New Orleans, LA, 2001.

[14] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes, and M. Sartin, "Combining content-based and collaborative filters in an online newspaper," in *ACM SIGIR Workshop on Recommender Systems-Implementation and Evaluation*, Berkeley, CA, 1999.

[15] B. M. Sarwar, J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. T. Riedl, "Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system," in *Conference on Computer Supported Cooperative Work*, 1998.

[16] N. Good, J. B. Schafer, J. A. Konstan, A. Borchers, B. M. Sarwar, J. Herlocker, and J. T. Riedl, "Combining collaborative filtering with personal agents for better recommendations," in *Conference of the American Association of Artifical Intelligence (AAAI-99)*, 1999, pp. 439–446.

[17] B. Smyth and P. Cotter, "Surfing the digital wave: Generation personalized tv listings using collaborative, case-based recommendation," in *Third International Conferece on Case-based Reasoning*, Munich, Germany, 1999.

[18] C. Basu, H. Hirsh, and W. Cohen, "Recommendation as classification: Using social and content-based information in recommendation," in *Proceedings of the 15th National Conference on Artificial Intelligence*, Madison, WI, 1998.

[19] M. K. Condliff, D. D. Lewis, D. Madigan, and C. Posse, "Bayesian mixed-effects models for recommender systems," in *ACM SIGIR '99 Workshop on Recommender Systems: Algorithms and Evaluation*, Berkeley, CA, 1999.

[20] A. I. Schein, A. Popescul, L. H. Ungar, and D. M. Pennock, "Methods and metrics for cold-start recommendations," in *ACM SIGIR-2002*, Tampere, Finland, 2002.

[21] S. Ujjin and P. J. Bentley, "Particle swarm optimization recommender system," in *Proceedings of the IEEE Swarm Intelligence Sympoisum 2003*, Indianapolis, 2003.

[22] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, pp. 5–53, 2004.

[23] J. L. Herlocker, "Understanding and improving automated collaborative filtering systems," Ph.D. dissertation, University of Minnesota, 2000.

[24] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. T. Riedl, "Item-based collaborative filtering recommendation algorithms," in *10th International World Wide Web Conference (WWW10)*, Hong Kong, 2001.

[25] J. Herlocker, J. A. Konstan, A. Borchers, and J. T. Riedl, "An algorithmic frameworkd for performing collaborative filtering," in *The 1999 Conference on Research and Development in Information Retrieval*, 1999.

[26] E. G. Vozalis and K. G. Margaritis, "Recommender systems: An experimental comparison of two filtering algorithms," in *Proceedings of the 9th Panhellenic Conference in Informatics - PCI 2003*, 2003.