Personalized advertising services through hybrid recommendation methods: the case of digital interactive television

George Lekakos
Department of Informatics
Cyprus University
glekakos@cs.ucy.ac.cy

Abstract

Recommender systems are a special class of personalized systems that aim at predicting a user's interest on available products and services by relying on previously rated items or item features. Human factors associated with a user's personality or lifestyle, although potential determinants of user behavior, are rarely considered in the personalization process. In this paper we demonstrate how the concept of lifestyle can be incorporated in the recommendation process to improve the prediction accuracy by efficiently managing the problem of limited data availability. We propose two approaches: one relying on lifestyle alone and another integrating lifestyle within the nearest neighbor approach. Both approaches are empirically tested in the domain of recommendations for personalized television advertisements and are shown to outperform existing nearest neighborhood approaches in critical cases.

Keywords: Personalization, Recommenders systems, Collaborative filtering, Lifestyle.

1. Introduction

The vast amount of information available over digital platforms, coupled with the diversity of user information needs, have urged the development of personalized systems that are capable of distinguishing one user from another to provide content, services, and information tailored to individuals. Although personalization has long drawn the attention of academics and practitioners, methods and techniques for implementing personalization remain a topical issue of research. Early typical user modeling approaches, although successful at the academic level, have proven of little significance in real life applications (Strachan et al., 2000). Conversely, recommender systems, which typically follow low-complexity modeling approaches, have demonstrated remarkable success in actual conditions (Konstan, 2001).

Recommendation approaches typically rely upon implicitly or explicitly acquired behavioral data denoting users' interest but are rarely concerned with factors that characterize the users themselves. Human factors, such as personality and cognitive/learning style, can play an important role in the personalization process and have been studied in user modeling research (Picard, 1997; Hudlicka, 2001). However, lifestyle remains understudied despite being one of the most popular predictors of buying or information seeking behavior in consumer behavior theory.

This research demonstrates that lifestyle can serve as the basis of effective recommendation methods and proposes novel algorithmic approaches that exploit this factor. The proposed approaches address the sparsity problem, which is considered as one of the most important limitations in several recommendation techniques and occurs when few interaction data are available (Konstan et al., 1997; Balabanovic and Shoham, 1997).

The application domain of our research is the domain of digital interactive television advertisements, considered as an audiovisual information item that promotes (recommends) products, services, or information to individual viewers (users). The developments in digital interactive television technology (Milenkovic, 1998) provide the ability to replicate (in this environment) methods and techniques applied on other interactive environments and at the same time draw conclusions that can be applicable in other domains as well.

The remaining of the paper is organized as follows. In the next section, background work is presented, followed by the presentation of the proposed personalization approach. The empirical results of the application of the proposed approach are then discussed. Next, the proposed approach is extended to demonstrate that its performance can drastically increase within certain hybridization techniques. Conclusions and further research issues are discussed in the last section.

2. Background Work

The original idea underlying recommender systems was based on the observation that people very often rely upon opinions and recommendations from friends, family, or associates to make selections or purchase decisions. This "social" approach to information filtering (Malone et al., 1987) motivated the development of recommender systems defined as systems that "*produce* *individualized recommendations as output or have the effect of guiding the user in personalized way to interesting or useful objects in a large space of possible options"* (Burke, 2002).

The primary concern of recommendation approaches is to identify which of the information items (or objects) available are interesting or likeable to the users. The recommendation process usually takes user ratings on observed items and/or item features as input and produces the same type of output for unobserved items. Ratings can be collected either implicitly, by monitoring the user's interactive behavior (Breese et al, 1998), or explicitly, by asking users to rate the observed items. Two major approaches are used for processing input data and formulating the prediction: collaborative filtering (CF) and content-based filtering (CBF).

Collaborative filtering operates upon the assumption that users who have exhibited similar behavior in the past (or present some form of similarity) can serve as recommenders for each other on unobserved data items. So, given the target user's ratings on observed items, the idea is to trace relationships or similarities between him/her and the remaining of the users in the database, aggregate the "similar" users' ratings, and use them as a prediction for the target user.

On the other hand, content-based filtering makes predictions upon the assumption that a user's previous preferences or interests are reliable indicators for his/her future behavior. CBF requires that items are described by features, and is typically applied upon text-based documents, or in domains with structured data (Balabanovic and Shoham, 1997; Pazzani, 1999). For example, content-based filtering has been utilized in book recommendation tasks (Mooney and Roy, 2000), using features such as title, author, or theme. In such cases, the user's previous preferences on the respective features are used to filter the available books and recommend the most relevant to the user.

In contrast to content-based filtering, collaborative filtering is applicable to any type of content (Balabanovic and Shoham, 1997), while it can also capture concepts that are hard to represent, such as quality and taste (Herlocker et al., 2002). Additionally, collaborative filtering does not restrict the spectrum of recommendations to items similar to the ones that the user has previously evaluated. Collaborative filtering has been acknowledged as the most successful and most widely implemented recommendation technique to date (Burke, 2002; Sarwar et al., 2000). For the above reasons our research is focused on the collaborative filtering strategy.

The collaborative filtering task concerns the prediction of the target user's rating for a specific item (the target item), based upon his/her ratings on observed items. Each user is represented by item-rating pairs, and can be summarized in a user × item table which contains the ratings Rij that have been provided by the i^{th} user for the j^{th} item.

Collaborative filtering approaches can be distinguished into two major classes: model-based and memory-based (Breese et al, 1998). Model-based methods develop a model, which is applied upon the target user's ratings to make predictions for unobserved items (e.g. Breese et al., 1998; Basu et al., 1998; Billsus and Pazzani, 1998). On the other hand, memory-based methods operate upon the entire database of users to find the closest neighbors of the target user and weight their recommendation according to their similarities. The fundamental algorithm of the memory-based class is the nearest neighbor (denoted as NN, hereafter), which is considered as one of the most effective collaborative filtering approaches (Good et al., 1999; Herlocker et al., 2002; Schafer et al., 2001; Hofmann, 2004) and can serve as a suitable and reliable algorithm for our recommendation task. It can be described as a process divided in three steps (Resnick et al., 1994): (a) computation of similarities between the target user (the user we aim to make prediction) and the remaining users, based on the commonly rated items; (b) selection of the most similar users to form the target user's neighborhood, and (c) computation of prediction based on the weighted average of the neighbors; ratings on the target item.

Nearest neighbor algorithms present high-accuracy levels in terms of prediction (Herlocker et al., 2002), confirmed by empirical studies (Breese et al., 1998; Sarwar et al., 2000; Schafer et al., 2001; Basu et al., 1998; Good et al., 1999). The disadvantages of nearest neighbor algorithms are associated with their limited scalability, as the size of the database increases. In addition, nearest neighbor algorithms inherit the limitations of collaborative filtering algorithms in cases of limited availability of user ratings on observed items, which seriously affects the predictive performance, as discussed next.

3. The role of lifestyle

The sparsity problem has been acknowledged as the most important drawback in collaborative filtering algorithms (Billsus and Pazzani, 1998; Claypool et al., 1999; Pennock et al., 2000), and refers to the low ratio of rated items to the total of available items. In general, recommender systems users rate only a small fraction of the available items, since they are not willing to invest time and effort in rating (Aggarwal et al., 1999). Even in systems where ratings are collected implicitly through monitoring of users' interactive behavior, the vast amount of available items and the requirement that users have actually observed and reviewed an item makes the collection of a sufficient number of ratings problematic. In nearest neighbor algorithms, sparsity affects the measurement of similarities, thus leading to unreliable measures and consequently to reduced prediction accuracy. This problem is also significant when a new item is introduced in the system (the *"new item"* problem): since no rating exists for the new item, it cannot be recommended to any of the users. Similarly, the *"new user"* problem occurs when a new user is introduced in the system and therefore no ratings are available for the measurement of his/her similarities with the remaining users.

A major goal in collaborative filtering applications is to improve the accuracy of algorithms by effectively addressing the sparsity problem. We argue that the concept of *lifestyle* can be utilized to overcome this problem and produce more accurate predictions.

Lifestyle is defined as the patterns in which people live and spend their time and money (Gunter and Furnham, 1992). It represents the central notion in the Consumer Behavior Model (Hawkins et al., 1998) which suggests that consumers' actual and desired lifestyle (i.e. the way they would like to think and feel about themselves) are translated into daily behaviors including purchase and consumption behavior. Lifestyle is affected by a number of external (culture, subculture, demographics, social status, reference groups, family, and marketing activities) and internal factors (perception, learning, memory, motives, personality, emotions, and attitudes). It can be quantified through psychographic research that measures constructs revealing attitudes, values and beliefs, interests and activities, demographics, media consumption, and product usage rates. The measurement of these constructs and the application of clustering techniques upon these data lead to the lifestyle segmentation, one of the most effective and popular market segmentation methods (Vyncke, 2002). The clustering process also provides a set of classification rules, which can be applied to consumers' demographic and media consumption data to classify them into the lifestyle segments. Subsequently, the product usage rates attached to the description of the segments are used to infer the preferences of consumers and target products accordingly.

Given the widespread and successful use of lifestyle segments in market targeting activities, we can hypothesize that the concept of lifestyle may also be utilized for the recommendation of products and services, in particular in sparsity conditions. More specifically, the underlying idea is that, instead of developing the neighborhood of the target user based on unreliable similarities computed upon a few ratings, we can develop a "lifestyle" neighborhood. Since people can be discriminated upon their lifestyles (Chaney, 1996), and consumers found in the same lifestyle segment (or neighborhood) present common behavior, then the members of a "lifestyle" neighborhood can be considered as reliable recommenders to each other.

In personalization research, personality traits (such as lifestyle) have been acknowledged as potential personalization factors (Brusilovsky, 2001), but lifestyle has not been adequately studied to date, except in a few cases, such as SeAN (Ardissono et al., 2001) and Lifestyle Finder (Krulwich, 1997). However, both systems limit the exploitation of lifestyle to the classification of users in lifestyle segments. In research indirectly involving personality or behavioral factors, Pennock et al. (2000) propose a Personality Diagnosis (PD) algorithm, based on the assumption that there is an association between how people rate items and their personality type (modeled as a latent variable). Other approaches aim at clustering users assuming that behavioral relationships exist among them (Breese et al, 1998; Ungar and Foster, 1998). Besides the fact that such approaches are not concerned with the notion of lifestyle per se, another major difference is that they rely upon available data and are therefore affected by the sparsity problem, while the personality factor upon which clusters are developed is not specified.

In contrast, lifestyle is a meaningful factor that can group users of personalized systems and eventually overcome the sparsity effect. The development of a personalization approach incorporating the notion of lifestyle is presented in the next section.

4. Developing the Lifestyle approach

The main idea underlying the use of lifestyle to overcome the sparsity effect is to avoid considering irrelevant neighbors as relevant (or ignoring actually relevant neighbors) due to the absence of a sufficient number of ratings. Indeed, sparsity affects the most important step in the prediction process, which is the identification of users that will be considered as recommenders

We contend that the incorporation of an additional step in the prediction process, involving the identification of similar users in terms of their lifestyle, will restrict the search space among users that present this form of similarity (lifestyle), thus avoiding the effects of misleading similarity computations. A direct implementation of this additional step is to exploit existing lifestyle segments, relying on their successful utilization in market targeting activities. More specifically, if users are classified into one of these segments then we achieve our objective of restricting the search space within each segment, which includes similar users. Prediction can then be made by assigning the cluster's centroid (the mean rating value of the ratings of the segment members for a specific item) or assigning an expert's predefined prediction. However, these predictions are assigned uniformly to all users and therefore the level of similarity between the target user and his/her neighbors is not taken into account. Thus, an alternative approach would be to weigh each neighbor's contribution in the final prediction by his/her similarity to the target user, following a nearest neighbor reasoning.

Both options require that users have been pre-classified in lifestyle segments and anchor the whole process to the availability and validity of existing proprietary segments (Mowen and Minor, 1998; Gunter and Furnham, 1992; Beatty et al., 1998). In addition, the classification of users in lifestyle segments is performed under the assumption that each user belongs to one segment only, and therefore behaviorally similar users with different lifestyles cannot be traced. The development of an approach that may overcome these problems is discussed next.

To address these limitations we investigate a segment-independent way of identifying lifestyle neighbors directly at the individual level, instead of classifying users into lifestyle segments. This could be achieved by measuring "lifestyle" similarities directly between the target and the remaining users to develop a "personal" neighborhood for the target user. Thus, each user must be described by a set of features that meet the following requirements:

(a) They should be lifestyle indicators, i.e. they should be significantly associated with the membership of a user into a lifestyle.

(b) They should be independent from the availability of ratings to reduce the sparsity effect.

(c) They should be easily collectable to avoid increasing the cost of collecting additional information and engaging the users into time-consuming questionnaire filling processes.

Evidence suggests that user demographics can play this role. Indeed, demographics have been successfully used to classify users in *Lifestyle Finder* (Krulwich, 1997) and *SeAN* (Ardissono et al., 2001). However, demographic data are usually too generic to achieve accurate classification results when used in isolation. For example, in *SeAN*, Ardissono et al. (2001) had to combine demographics with user hobbies, while in cross-selling of banking services (Peltier et al., 2002), demographics were used in conjunction with customer credit data.

We performed statistical analysis on the data (including demographic and television program preferences) provided by a sample of 502 users. The results demonstrated that demographics in combination with TV program preferences are significant indicators of membership in lifestyle segments. In particular, significant discriminating power with respect to lifestyle has demonstrated the combination of the demographic variables "age", "marital status", and "education" with the eight program genres "Documentaries", "Cartoons", "Football/ basketball/ volleyball games", "Video clips", "Domestic comedy series", "Discussions/ interviews", and "News".

These indicators also satisfy the second and third requirements set above. For example, in the domain of digital television, demographics can be collected at the subscription to the service and updated online if necessary, while television program preferences can be easily traced online through the set-top box.

The television program preferences as well as the demographic data are encoded in a uniform binary format and each user is represented by a vector of feature-value pairs. Hence, similarities among users can be computed upon these features and the target user's neighbor consisting of users with "similar" lifestyle patterns can be traced. While several similarity measures can be applied on binary variables (for example, simple matching, Jaccard coefficient, and phi 4-point correlation) the Pearson correlation coefficient is selected since it has been used in the measurement of similarities upon demographic data (Pazzani, 1999) as well as for consistency reasons with the nearest neighbor (NN) algorithm that will be used for the evaluation of the proposed approach. Let us call this approach the "*lifestyle*" approach, described by the following steps:

1. Measure similarities between the target and the remaining users based upon data associated with their lifestyle by applying equation (3).

$$w(i,j) = \frac{\sum_k (I_{i,k}-\overline{I_i})(I_{j,k}-\overline{I_j})}{\sqrt{\sum_k \left(I_{i,k}-\overline{I_i}\right)^2 \sum_k \left(I_{j,k}-\overline{I_j}\right)^2}}, \text{ (Eq. 1)}$$

   where $I_{i,k}$ and $I_{j,k}$, refer to k^{th} lifestyle indicator available in common for the i^{th} (target user) and j^{th} users, and $\overline{Ii}$ and $\overline{Ij}$ to the corresponding means.

2. Formulate the target user's neighborhood, based on the similarity measures described in step 1, by selecting users who score above a certain threshold.

3. Predict the target user's rating on the target item by aggregating lifestyle neighbors' ratings

weighted by the lifestyle similarities developed at step 1. Aggregate the target user's preferences into a prediction for the target item by applying equation (2) (Resnick et al., 1994; Shardanand and Maes, 1995):

$$R_{\mathbf{i,p}} = \overline{R_i} + \frac{\sum_{j=1}^{m} w(i,j)\,(R_{j,p} - \overline{R_j})}{\sum_{j=1}^{m} |w(i,j)|}\ , \quad \text{(Eq. 2)}$$

where $R_{i,p}$ is the rating to be predicted for user i and for item p, $\overline{R_i}$ is the mean of the ratings of user i for all items that user has provided his/her ratings, the weight w(i,j) is the similarity measure between user i and j and $R_{j,p}$ is the rating of user j for item p and $\overline{R_j}$ is the mean of ratings of user j in a neighborhood of size m.

The main difference between the lifestyle and the NN approach is at the first step of the proposed method where lifestyle indicators, instead of ratings, are used for the computation of similarities. This step results into different neighbors and different weights indicating the importance of each neighbor's rating utilized in the prediction formula. Thus, the prediction accuracy of the proposed approach depends of the validity of the hypothesis that lifestyle neighbors are more reliable than ratings-based neighbors in sparsity conditions. The validity of this hypothesis will be examined in the next section.

5. Empirical Evaluation of the Lifestyle Approach

In the experiment, we employed a sample of 37 individuals drawn from our research group. The sample includes academic (19%), research (73%), and technical staff (8%), consisting of 62.2% males and 37.8% females, aged 18-24 (10.8%), 25-34 (67.6%), 35-44 (18.9%) and 45-54 (2.7%). The users were shown 65 advertisements selected from seven product categories (food and drink, fast moving consumer goods, computer and technology, family and home, books and magazines, public services, finance and investment, and autos) and provided their ratings for each advertisement in a one-to-five scale. Similarly to other domains (e.g. movies, books) a rating concerns the degree of likeability (i.e the overall taste) for each item. Furthermore, participants filled-in a questionnaire providing their demographic and TV program preferences data as required by the lifestyle algorithm.

Empirical findings suggest that the best predictive performance of the NN approach is achieved within a neighborhood of 20-30 users (Herlocker, 1999; 2002; Sarwar et al, 2001). Thus, our sample size ensures that the performance of the NN algorithm will not be pessimistic due to a small sample size.

To compare the performance of the two algorithms, a cross-validation technique was used, appropriate for small-sample experiments (Dietterich, 1998). In particular, we employed the leave-one-out (LOO) cross-validation, which is the recommended technique for small-sample model selection problems (Cawley and Talbot, 2003). This method replicates the error estimation process *n* times for a sample of size *n* by considering each user in the original sample as the test set (target user) and the remaining sample of size *n-1* as the training set. A certain number of randomly selected ratings are considered available for each target user to make a prediction for the remaining (removed) ratings. The ratings are removed following the experimental design for the empirical analysis of collaborative filtering algorithms introduced by Breese et al. (1998) who describe a set of experimental protocols, called *Given 2, Given 5, and Given 10*.The *Given n* protocol involves the random selection of 2, 5, or 10 votes (corresponding to "n") from each test user as the observed ratings. These ratings are then used to predict the remaining ratings. The various "given" protocols examine the performance of the algorithms when relatively little is known about the active user. This set of experimental protocols also includes the "all-but-one" protocol where a single rating is removed and predicted given all the other ratings. It measures the performance of the algorithms when given as much data as possible (Breese et al, 1998) and therefore it is beyond the scope of our empirical evaluation, which concerns the performance of the algorithms in sparsity conditions.

The corresponding prediction error for each user is measured using the *Mean Absolute Error (MAE)*, which is the average difference between the predicted and the actual rating value and is commonly used for performance evaluation (Breese et al, 1998; Herlocker et al., 2002; Shardanand and Maes, 1995). The overall error estimator for each of the approaches is the average value of the errors from (65-n)*37 predictions, where 65 is the total number of items in the dataset, *n* is the number of available items in each of the "Given n" experimental protocols, and 37 is the total number of users.

We focus our attention on the Given 2, 5, and 10 protocols since they represent the sparsity conditions according to our objective. In all cases presented below the normality requirement is met and the differences in mean absolute errors are compared using paired t-tests (Mitchell, 1997). The ratings of the training and test items follow the same distribution since they are drawn through stratified random sampling according to the original distribution of ratings.

The averaged results along with the values of the t-statistic are depicted in Table 1.

	Given 2	Given 5	Given 10
Lifestyle	*1.1639*	*0.8265*	*0.7850*
NN	*1.1857*	*0.8416*	*0.7942*
p	0.0001	0.000	0.05

Table 1: Overall performance of the algorithms

The results confirm that the *lifestyle* approach achieves lower error levels for all protocols An interesting finding is that lifestyle approach, although significantly better, is also affected by the number of available items, despite the fact that similarities are computed on the basis of rating-independent data. Considering the prediction formula (equation 2), each user in the target user's neighborhood does not contribute directly by his/her rating but does so by the "amount" of likeness (positive value of the quantity $R_{j,p} - \overline{R_j}$) or "dislikeness" (negative value of the above quantity). The weighted sum of the above quantities is added to the mean value of the available ratings of the target user. Thus, if the target user's mean value is misleading with regard to what the user perceives as neutral value due to few ratings, then the prediction may be negatively affected (but still less significantly than in the NN-based prediction).

One of the additional benefits of the lifestyle approach is that it can be further analyzed based on the elements of lifestyle, in contrast to the limited explanatory power of the collaborative filtering approach. In the following section we analyze the performance of the lifestyle approach in relation to product-related attributes.

6. The Integrated Approach

So far we have considered, measured, and compared the performance of individual algorithms and concluded that the lifestyle approach outperforms the NN one in the absence of a sufficient amount of ratings for each user. Exploiting the behavior of the two algorithms, in this section we apply an *integrated* recommendation mechanism that utilizes the lifestyle approach to make predictions upon the (eventually few) available ratings for the items unobserved by the target user, populating his/her ratings vector. A new vector is introduced for each user, consisting of the original ratings provided by the user and the ratings predicted by the application of the lifestyle approach. Thus, for each user a new user representation is developed, called "pseudo-user", who participates as a significant neighbor in a subsequent application of the NN prediction process, which then produces the final prediction for the unobserved items

More specifically, let us assume a number *u* of users where each $user_i$, i=1,…,*u* has provided his/her ratings for k_i items {$R_{i,1}$, $R_{i,2}$,…, $R_{i,ki}$} in a total of *n* items, and prediction concerns a rating $R_{t,p}$ for a $user_t$ (target user) and an item p (target item). Then the steps of the approach can be described as follows:

Lifestyle Prediction

1. For each pair of users ($user_i$, $user_j$), i,j=1,…,*u*, $i \neq j$ measure their similarity w(i,j) upon their lifestyle indicators, applying equation 3 (section 3.2).
2. For each $user_i$, i=1,…,*u*, develop his/her neighborhood (NBi) as follows: NBi={$user_j$ | w(i,j) > threshold}.
3. For each $user_i$ predict the $n\text{-}k_i$ missing ratings {$L_{i,1}$, $L_{i,2}$, … $L_{i,n-ki}$} utilizing (in equation 4) the available ratings from all $users_j \in$ NBi for each of the above missing ratings.

Pseudo-user table formulation

4. For each $user_i$, introduce in the Pseudo-user table a pseudo-user p_$user_i$, whose ratings Pi,q are defined as follows:

$$P_{i,q} = \begin{cases} R_{i,q}, \text{if rating for item q has been provided by user i} \\ L_{i,q}, \text{if item q has not been rated by user i} \end{cases} ,$$

i=1,…,*u*, q=1,…,*n* and $L_{i,q}$ refers to the predicted rating by the Lifestyle approach.

NN-based prediction

5. Measure similarities between the target $user_t$ and each of the remaining $users_j$, j=1,…,u, $j \neq t$, using the Pearson correlation formula (equation 1). Assign an increased weight to the p_$user_t$ (whose k_t ratings are identical with the target user's ratings and therefore we may be more confident for his/her recommendations on the missing ratings).
6. Develop the target user's neighborhood by selecting users above a threshold value.
7. Produce the prediction by the NN prediction formula (equation 2) for the neighbors selected in step (6) and weights computed in step (5).

The utilization of the *lifestyle* prediction is very promising, since sparsity is completely removed but the *lifestyle* predictions also transfer the errors produced by the lifestyle approach and the effect of massive substitution of all users by their respective pseudo-users remains to be investigated. The main hypothesis to be tested is that the *integrated* approach significantly outperforms the NN approach. The evaluation results will also demonstrate the differences in terms of performance between the *integrated* approach and the lifestyle and NN approaches discussed so far.

7. Empirical evaluation

The design utilizes once again the "Given n" protocols, upon the sparse user × item table. For each user the same number of randomly selected ratings is removed and both algorithms are applied upon the "original table". At this phase of the experimentation series, we extend the number of given ratings beyond the Given 2, 5, and 10 ratings. Specifically, the performance of all algorithms discussed so far is measured upon 2, 5, 10, 15, 20, 25, 30, 35, and 50 ratings, to acquire an overall picture of the performance of the various approaches in different levels of available ratings. In the approaches discussed so far, the focus was on proving their superiority to the NN approach upon few ratings. In addition, the experimental results demonstrated that the NN approach improves its performance as more ratings become available. However, our expectations for improved performance of the integrated approach at any sparsity level (compared to the NN), suggest the extension of the experiment in various amounts of available ratings.

The hypothesis of significant difference of the performance between the integrated and the NN approach is tested by paired t-tests between the performances of the two algorithms for each individual user. In addition to the main hypothesis, the integrated approach is also compared to the lifestyle approach to assess the value and usefulness of the integrated approach (Table 2).

	NN	**Lifestyle**	**Integrated**
G2	1.1071	1.1053	1.1048
p	0.010	0.142	
G5	1.0056	0.9881	0.9783
p	< 0.0001	< 0.0001	
G10	0.9333	0.9286	0.9057
p	< 0.0001	< 0.0001	
G15	0.906	0.9059	0.8776
p	< 0.0001	< 0.0001	
G20	0.8859	0.8884	0.8556
p	< 0.0001	< 0.0001	
G25	0.8679	0.8653	0.8309
p	< 0.0001	< 0.0001	
G35	0.8352	0.8381	0.8081
p	< 0.0001	< 0.0001	
G50	0.7784	0.7867	0.7687
p	0.045	0.006	

Table 2: Predictive performance of the approaches on different sparsity levels

The results clearly suggest that the *integrated* approach is significantly better than all the approaches examined so far, apart from the case of Given 2, where the availability of only two ratings has significant and rather unmanaged effects on all approaches. In the case of the "Given 50" protocol, the significance levels of the *integrated* approach superiority are downscaled, compared to the rest of the experiments – as expected – but still the averaged performance is significantly better.

In contrast to the lifestyle approach, the integrated approach firmly outperforms the NN in the range of 5 to 35 available ratings, while the improvement is decreased for 50 ratings. At the Given 2 protocol, all approaches are affected by the limited number of available ratings and present similar performances (although statistical significance is confirmed for the difference between the integrated and the NN approach). Measuring the improvement in performance by the difference in MAE values between the NN and *integrated* approaches in relation to the various sparsity levels (which correspond to the number of available ratings), it is found that the improvement rate escalates from a very small 0.2% at 97% sparsity (Given 2) to a 2.7% improvement rate (Given 5). Then it follows an incremental route up to a 4.3% at 62% sparsity level (Given 25), which is the maximum rate. The improvement still maintains a high rate of 3.2% for the Given 35 protocol (46% sparsity) and finally drops at 1.2% when sparsity is only at 23% and the NN approach presents its best performance. This trend indicates that as we approximate a completely dense user x item table the NN approach would perform better than the integrated approach.

The performance of the integrated approach compared to the NN one suggests that the management of the sparsity effect can significantly improve the accuracy of the prediction, within the limits set out by the errors inherited by the two approaches upon which the integrated approach has been built. However, the value of the integrated approach lies in its flexibility to accommodate any improved variation of the lifestyle and/or NN approaches. Furthermore, a second important feature of the approach is that it is open to any type of prediction mechanism that is affected by sparse datasets. Indeed, the utilization of lifestyle data and the prediction of missing ratings remove the sparsity effect. This type of outcome (predicted ratings) can be subsequently loaded as input to virtually any learning algorithm (for example, Bayesian Networks) and improve its performance.

8. Conclusions and Future Research

User personality factors, such as lifestyle, are rarely considered in the personalization process despite their theoretical significance in the prediction of a user's future behavior. We have proposed personalization approaches based on the notion of lifestyle, which has been shown to increase the accuracy of prediction in sparsity conditions, compared to established recommendation approaches, such the nearest neighbor algorithm. The main idea underlying the development of the lifestyle approach is to avoid erroneous selection of neighbors due to misleading computation of similarities among users in sparse databases. The lifestyle approach can overcome this problem by performing the prediction task within the range of lifestyle neighbors.

Moreover, we have followed a hybridization technique based on the notion of pseudo-user, to strengthen the performance of the lifestyle approach and we have demonstrated that the integrated approach yields superior predictive performance at varying sparsity levels. This is achieved by combining the ability of the lifestyle approach to make predictions when few ratings are available with the ability of the nearest neighbor approach to make successful predictions upon sufficient amounts of available ratings.

However, a pre-selection of a sub-sample or the application of some other data reduction algorithm is required to apply the integrated approach in large databases due to the increased computational cost of the method. Furthermore, our empirical findings may be limited by the sample size and the population from which the sample was drawn, which consists mainly of well-educated IT literate individuals. Sample size limitations have been managed to a certain extent by applying cross-validation techniques and exploiting particular features of nearest neighbor algorithms concerning the optimal neighborhood size. However, our future research plans include the evaluation of the proposed approaches upon larger datasets from different domains.

One of the challenges emanating from the present research is to apply the proposed personalization approaches in other domains, such as personalized systems over mobile platforms, which are also characterized by limitations in collecting extended user-driven interaction data. The identification of lifestyle indicators, which are easily collectable in other environments, or even the complete disengagement of the process from such indicators would also increase the generalization ability of the lifestyle approach. In addition, the combination of the lifestyle approach as a collaborative filtering method with content-based filtering could further increase the accuracy and transparency of predictions. Although the proposed approaches address the sparsity problem they cannot operate and make predictions for new items (that have not been rated) and therefore the exploitation of existing relationships between items would eventually lead to a personalization approach that operates in any situation related to the absence of ratings.

References

Aggarwal, C., Wolf, J., Kun-Lung, W., & Yu, P. S. (1999). *Horting Hatches an Egg: A New Graph-Theoretic Approach to Collaborative Filtering.* Paper presented at the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining., San Diego, CA.

Ardissono, L., Console, L., & Torre, I. (2001). An adaptive system for the personalized access to news. *AI Communications, 14*(3), 129-147.

Balabanovic, M., & Shoham, Y. (1997). Fab: Content-based collaborative recommendation. *Communications of the ACM, 40*(3), 66-72.

Basu, C., Hirsh, H., & Cohen, W. (1998). *Recommendation as Classification: Using social and Content-based Information in Recommendation.* Paper presented at the 15th National Conference on Artificial Intelligence, Madison, WI.

Beatty, S. F., Homer, P. M., & Kahle, L. R. (1998). Problems with VALS in International Marketing Research: An Example from an Application of the Empirical Mirror Technique. *Advances in Consumer Research, 15*, 375-380.

Billsus, D., & Pazzani, M. (1998). *Learning Collaborative Information Filters.* Paper presented at the Fifteenth National Conference on Machine Learning.

Breese, J. S., Heckerman, D., & Kadie, C. (1998). *Empirical Analysis of predictive algorithms for collaborative filtering.* Paper presented at the Fourteenth Annual Conference on Uncertainty in Artifical Intelligence.

Brusilovsky, P. (2001). Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction, 6*(2-3), 87-129.

Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User Adapted Interaction, 12*, 331-370.

Cawley, G. C., & Talbot, N. L. C. (2003). Efficient Leave-one-out cross-validation of kernel Fisher discriminant classifiers. *Pattern Recognition, 36*(1), 2585-2592.

Chaney, D. (1996). *Lifestyles*. London: Routledge.

Claypool, M., Gokhale, A., Miranda, T., Murnikov, P., Netes, D., & Sartin, M. (1999). *Combining Content-Based and Collaborative Filters in an Online Newspaper*. Paper presented at the ACM SIGIR Workshop on Recommender Systems, Berkeley, CA.

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation, 10*, 1895-1923.

Good, N., Schafer, J. B., Konstan, J., Borchers, A., Sarwar, B., Herlocker, J., et al. (1999). *Combining Collaborative Filtering with personal Agents for Better Recommendations*. Paper presented at the 16th National Conference on Artificial Intelligence (AAAI-99), Menlo Park, CA.

Gunter, B., & Furnham, A. (1992). *Consumer Profiles: an Introduction to Psychographics*. London: Routledge.

Hawkins, I., Best, R. J., & Coney, K. A. (1998). *Consumer Behavior: Building Marketing Strategy*. New York: Irwin/McGraw-Hill.

Herlocker, J., Konstan, J., & Riedl, J. (2002). An Empirical Analysis of Design Choices in Neighborhood-Base Collaborative Filtering Algorithms. *Information Retrieval, 5*, 287-310.

Hofmann, T. (2004). Latent Semantic Models for Collaborative Filtering. *ACM Transactions on Information Systems, 22*(1), 89-115.

Hudlicka, E. (2001). *Integrated model of trait and state: influences on user behavior*. Paper presented at the 2nd Workshop on Attitude, Personality and Emotions in User-Adapted Interaction, in conjunction with User Modeling 2001, Sonthofen, Germany.

Konstan, J. (2001). *Heavyweight Applications of Lightweight User Models: A Look at Collaborative Filtering, Recommender Systems and Real-Time Personalization*. Paper presented at the 8th International Conference in User Modeling, UM 2001, Sonthofen, Germany.

Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L. R., & Riedl, J. (1997). Applying Collaborative Filtering to Usenet News. *Communications of the ACM, 40*(3), 77-87.

Krulwich, B. (1997). Lifestyle Finder: Intelligent User Profiling Using Large-Scale Demographic Data. *AI Magazine*, 37-45.

Malone, T. W., Grant, K. R., Turbak, F. A., & Brobst, S. A. (1987). Intelligent Information Systems. *Communications of the ACM, 30*(5), 390-402.

Mooney, R. J., & Roy, L. (2000). *Content-based book Recommending Using Learning for Text Categorization*. Paper presented at the Fifth ACM Conference in Digital Libraries, San Antonio, TX.

Mowen, J. C., & Minor, M. (1998). *Consumer Behavior* (5th Edition ed.). Upper Saddle River, New Jersey: Prentice-Hall.

Pazzani, M. (1999). A Framework for Collaborative, Content-Based and Demographic Filtering. *Artificial Intelligence Review, 13*(5/6), 393-408.

Peltier, J. W., Schibrowsky, J. A., Schultz, D. E., & Davis, J. (2002). Interactive Psychographics: Cross-Selling in the Banking Industry. *Journal of Advertising Research, 42*(2), 7-22.

Pennock, D. M., Horvitz, E., Lawrence, S., & Giles, C. L. (2000). *Collaborative Filtering by Personality Diagnosis: A Hybrid Memory and Model-Based Approach*. Paper presented at the Sixteenth Conference on Uncertainty in Artificial Intelligence, San Francisco.

Picard, R. W. (1997). *Affective Computing*: MIT Press.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). *GroupLens: An open architecture for collaborative filtering of NetNews*. Paper presented at the ACM Conference on Computer Supported Cooperative Work.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). *Analysis of recommendation algorithms for e-commerce*. Paper presented at the ACM e-commerce conference.

Schafer, J. B., Konstan, J. A., & Riedl, J. (2001). Electronic-commerce recommender systems. *Journal of Data Mining and Knowledge Discovery, 5*(1), 115-152.

Shardanand, U., & Maes, P. (1995). *Social Information Filtering: Algorithms for Automating "Word of Mouth"*. Paper presented at the ACM CHI'95 Conference on Human Factors in Computing Systems, Denver, Colorado.

Strachan, L., Anderson, J., Murray, S., & Evans, M. (2000). Minimalist User Modelling in a Complex Commercial Software System. *User Modeling and User Adapted Interaction, 10*, 109-145.

Ungar, L., & Foster, D. (1998). *Clustering methods for collaborative filtering.* Paper presented at the Workshop on Recommendation Systems at the Fifteenth National Conference on Artifical Intelligence, Madison, Wisconsin.

Vyncke, P. (2002). Lifestyle Segmentation: From Attitudes, Interests and Opinions, to Values, Aesthetic Styles, Life Visions and Media Preferences. *European Journal of Communication, 17*(4), 445-464.