Knowledge-Based Systems 100 (2016) 175–187

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

User profiling approaches for demographic recommender systems

Mohammad Yahya H. Al-Shamri [a,b,*]

[a] Computer Engineering Department, College of Computer Science, King Khalid University, Abha, Saudi Arabia
[b] Electrical Engineering Department, Faculty of Engineering and Architecture, Ibb University, Ibb, Yemen

ARTICLE INFO

Article history:
Received 22 September 2015
Revised 4 March 2016
Accepted 5 March 2016
Available online 11 March 2016

Keywords:
Recommender system
User profile
Demographic data
Similarity computation

ABSTRACT

Many of our daily life decisions rely on demographic data, which is a good indicator for closeness of people. However, the lack of these data for many online systems let them search for explicit or implicit alternatives. Among many, collaborative filtering is the alternative solutions especially for e-commerce applications where many users are reluctant to disclose their demographic data. This paper explores, discusses and examines many user-profiling approaches for demographic recommender systems (DRSs). These approaches span many alternatives for profiling users in terms of the attribute types, attribute representations, and the profiling way. We present layout, description, and appropriate similarity computation methods for each one of them. A detailed comparison between these different approaches is given using many experiments conducted on a real dataset. The pros and cons of each approach are illustrated for more advantage that may open a window for future work.

1. Introduction

The huge amount of data and the emerging new ways of marketing enforce the administrators of online systems to search for automatic tools that may facilitate their systems. These systems offer many services to their users ranging from a joke to read or listen to many expensive things to buy online. Literature calls these automatic systems as recommender systems (RSs) with the aim to personalize the user navigation through the Web and direct their action. Today, these systems cover social networks, e-commerce, e-Business, e-Tourist, and many others [1–5]. Recently, Lu et al. [6] reviewed the applications of recommender systems, clustered them into eight main categories and summarized the related recommender system types used for each category.

Formally, there are five phases for building a RS, namely, data collection, user profiling, similarity computation, neighborhood selection, and finally predictions and recommendations. Based on the profile data, RSs can be content-based RSs (CBRSs), collaborative RSs (CRSs), or demographic RSs (DRSs) [1–5]. If the user profile is a set of features extracted from the descriptions of the items user liked before then we have a content-based RS. However, if the user profile is a set of attributes that describe the demographic class or group of the user then we have a DRS. Finally, if the user profile is a list of ratings for items the user has provided before, then we have a CRS which may follow a user-based approach or an item-based approach. Zhang et al. [7] developed a hybrid fuzzy collaborative recommender system which combined user-based and item-based approaches of CRS for mobile products and service recommendations.

The main goal of recommender systems is to address the online information overload problem and to improve the relationship between the system and its customers (users). Both issues are closely related to how the system represents the users and how much processing time is required for fulfilling the customer desire. Among many, DRS is the only system that has a limited number of features that can be fast for thousands if not millions of users. This makes DRS a suitable candidate for many online systems that faces rapid increasing of items and users.

Actually, DRSs do not gain that much popularity due to security and privacy concerns which stand on the top for the user hesitation and the difficulty to obtain true demographic data from the users. However, DRSs are available with a good percentage in our daily life and many online services will be more personalized if this data is taken into account. Age, gender, occupation, income, nationality, and many other demographic data are essential for many applications. For example, age groups are very important when suggesting movies while income ranges are very important when suggesting tourist places. In marketing, male and female shopping requirements are sometimes totally different and we cannot recommend some items without taking the gender of the targeted user into account. Moreover, some RSs suffer from many inherent problems that cannot be solved without hybridization between them and the DRS.

* Correspondence address: Electrical Engineering Department, Faculty of Engineering and Architecture, Ibb University, Ibb, Yemen. Tel.: +967 777796023 (mobile), +966 558137212; fax: +967 4 408068.
E-mail address: mohamad.alshamri@gmail.com

http://dx.doi.org/10.1016/j.knosys.2016.03.006
0950-7051/

This motivates us to explore in some details the profiling approaches of DRS along with the advantages and the appropriate similarity computation methods for each profiling approach. Intuitively, the most important demographic data are age, gender and occupation. Sometimes Zip Code is considered as a demographic data but it is important for some applications only. This makes the research in this field too difficult as the available options are very limited. However, this is not the case if we consider the ways of representing each attribute and the associated similarity methods for comparing them.

This paper studies the user profiling and the similarity computation phases and assumes that the other phases are the same for all approaches. The contributions of this paper are four-fold.

1. Many approaches for profiling users of DRS are studied.
2. We introduce similarity measures for some profiling approaches.
3. We propose a cascaded profiling approach for the neighborhood set generation.
4. We propose a single-attribute profiling approach by treating each attribute as an isolated profile and then merge their predictions.

The rest of this paper is organized as follows: a literature review is given in Section 2. Section 3 is an introduction to DRSs which gives a brief description of DRSs and discusses their advantages and disadvantages. Section 4 introduces the demographic user profile and the nature of the basic attributes for building it. The user profiling approaches for DRSs and the corresponding similarity measures are introduced in Section 5. Section 6 describes the experiments' dataset and the experimental methodology used for examining the profiling approaches. Section 7 discusses the results of the conducted experiments. Finally, we conclude our work in the last section.

2. Literature review

The roots of DRS dated back to 1979 [8], earlier than the notion of recommender system itself in the 90s of the last century [9,10]. Some pure examples for DRS are Grundy [5] which is the first DRS proposal for suggesting books and Lifestyle Finder [11] which aimed to market a range of products and services. To the best of our knowledge, a few number of research papers studied DRS and most of this work was a hybridization between DRS and the other types of RSs for overcoming the weaknesses of these systems like the cold-start problem of the CRSs [12–26].

Al-Shamri and Bharadwaj [12, 13] proposed a compact user model that exploits the user demographic data beside rating-driven features. Demographic data in this model benefited the user and allowed the system to overcome the cold-start problem. Moreover, the added demographic data allows the users with less number of ratings to enjoy the system. A same approach is used in [14] but with a different similarity method. Vozalis and Margaritis [15] proposed a feature combination hybrid RS that used demographic correlations to enhance the prediction accuracy. This work blended different features from different recommendation data sources into a single recommendation algorithm. They used dot product as a similarity measure between the profile vectors. Another work of Vozalis and Margaritis [16] utilized SVD and demographic data at various points of the filtering process in order to improve the predictions quality.

Safoury and Saleh [17] introduced a solution for the cold-start problem by utilizing the demographic data of the new user instead of their ratings. This allowed the system to serve the user even he had no ratings yet. The hybrid RS of Junior et al. [18] employed demographic data to discover and analyze the contextual constraints in a real world recommendations scenario. Ghazanfar and Prugel-Bennett [19] proposed a cascading hybrid RS that combined CBRS, CRS, and DRS. This approach somehow lets each RS to compensate the weaknesses of the others. The importance of demographic data for a research paper RS is studied by Beel et al. [20].

Sobecki [21] proposed two consensus-based hybrid RSs that mixed CBRS, CRS, and DRS at some way. The first proposal mixed CRS and DRS with some contents of the items while the second proposal mixed demographic, collaborative and content-based approaches at different components of the user model. Pazzani [22] proposed an approach that combines recommendations from multiple sources. Traveler agent [23] combined CBRS, CRS, and DRS to bring to the light the positive aspects of each recommender system. Moreno et al. [24] proposed SigTur/E-Destination for tourism and leisure activities using ontologies for guiding the reasoning process.

Said et al. [25] extended CRS to include some demographic features. They argued that these features hold implicit information about users taste and interest. Lu et al. [26] proposed a hybrid fuzzy semantic recommender system that combined item-based fuzzy semantic recommender system and fuzzy item-based fuzzy collaborative recommender system. This hybridization overcomes the semantic limitations of classical collaborative recommender system.

Yujie et al. [27] used the demographic data of new user within a social network to find similar users for him. Chen and He [28] proposed a system that generates user demographic vector from the user information and then employs number of common terms and term frequency for similarity computation. Weber and Castillo [29] studied the effect of some demographic data on the online searching behavior of US people and described how different segments of the population differ in their searching behavior. They argued that revealing the hidden relation between the demographic data and the query type might improve Web search relevance and provide better query suggestions.

3. Demographic recommender systems

DRS is a stereotypical system as it categorizes users based on their demographic attributes. Later, DRS uses the user opinions for the items of the system as a basis for recommendations. It is worth noting that both DRS and CRS utilize user-to-user correlations but based on different data. Therefore the advantages of DRS are almost similar to that of CRS in terms of their unique capacity in identifying cross-genre niches, enticing the users to jump outside the familiar, and their ability to improve themselves over time [3,5].

Formally, DRS has M users, $U = \{u_1, \ldots, u_M\}$, having N demographic attributes, $D = \{a_1, \ldots, a_N\}$. Usually, DRS collects demographic attributes during the registration process using questionnaire about the user demographic data and the user's characteristics [4,5]. Through interacting with the system, the user is asked explicitly or implicitly to rate K items, $S = \{s_1, \ldots, s_K\}$, such as news, Web pages, books, movies, or CDs. Initially, each user u_i may rate a subset of items S_i. The declared rating if available of user u_c for an item s_k is denoted by $r_{c,k}$ [2,10].

After constructing the user profile, DRS calculates the similarity value between the current active user and the remaining training users using a suitable similarity measure. This value indicates how closely the two users in consideration resemble each other. Accordingly, a set of neighbors is selected for this active user from the ranked list of the training users. After that DRS assigns a predicted rating to all the items seen by the neighborhood set and not by the active user. The predicted rating, $pr_{x,k}$, indicates the expected interestingness of the item s_k to the user $\mathbf{u}_x$ [2,3]. The predicted rating, $pr_{x,k}$, is usually computed as an aggregate of the ratings of

Table 1
Occupation distribution of the 100K MovieLens dataset.

Occupation ID	Occupation name	Frequency
1	Administrator	79
2	Artist	28
3	Doctor	7
4	Educator	95
5	Engineer	67
6	Entertainment	18
7	Executive	32
8	Healthcare	16
9	Homemaker	7
10	Lawyer	12
11	Librarian	51
12	Marketing	26
13	None	9
14	Other	105
15	Programmer	66
16	Retired	14
17	Salesman	12
18	Scientist	31
19	Student	196
20	Technician	27
21	Writer	45

$\mathbf{u}_x$s neighborhood set for the same item S_k [2]:

$$pr_{x,k} = \frac{\sum_{u_y \in N_x} sim(\mathbf{u}_x, \mathbf{u}_y) \times r_{y,k}}{\sum_{u_y \in N_x} |sim(\mathbf{u}_x, \mathbf{u}_y)|} \quad (1)$$

where N_x denotes the set of neighbors for $\mathbf{u}_x$ who have rated item S_k. DRS does not require a list of ratings for user profiling that are required by other RSs like CRS and CBRS. This makes DRS strong against "new user" problem. More interestingly, DRS follows the same way we get recommendations in our real life. A school student for example gets most of his recommendations from his classmates. Moreover, DRS is easy, quick, and straight forward as the profiling fields are always very few compared to ratings. This is very important when the number of users is very large. For other RSs, the system accuracy relies largely on the number of ratings because the larger the number of ratings the system get from the user, the higher the quality of its recommendations. This is not the case for DRS, because the profile is fixed for long time once the profiling attributes are obtained from the user.

On the other hand, the basic disadvantage of DRS lies in its sensitivity to security and privacy issues especially for e-commerce applications. Usually, online users are reluctant to share a big amount of personal information with a system due to their security. Due to their privacy, some users assume that disclosing demographic data breaks the anonymity of these systems. In terms of recommendations, the generated recommendations from the demographic groups may be too general [3–5]. However, this is not true if there are many tastes and modes within the demographic groups.

4. Demographic user profile

The demographic user profile (model) representation varies from system to system. Grundy system [8], the first DRS, offered interactive dialogue and used hand-crafted attributes with numeric confidence values for profiling users. Krulwich [11] used a short survey for gathering the demographic data for categorizing the user. Pazzani [22] extracted demographic features from the users' home pages for building a classifier.

Mostly the demographic user profile consists of three attributes, age, gender and occupation. Age is numeric or quantitative attribute while gender and occupation are categorical attributes that take on discrete unordered values. Sometimes, gender and occupation values are called binary attributes because they have two cases, belonging to the category or not. Gender takes on two values, male or female while occupation is a moving attribute, keeps going due to the emerging new occupation titles, and thus may change from time to time. However, the nature of occupation attribute is not changing and still a categorical attribute.

Table 2
General profile form for two users.

$\mathbf{U_x}$	$\mathbf{x_1}$	$\mathbf{x_2}$	$\mathbf{x_3}$
	$x_{11}, x_{12}, x_{13}, ..., x_{1n}$	$x_{21}, x_{22}, x_{23}, ..., x_{2m}$	$x_{31}, x_{32}, x_{33}, ..., x_{3q}$
$\mathbf{U_y}$	$\mathbf{y_1}$	$\mathbf{y_2}$	$\mathbf{y_3}$
	$y_{11}, y_{12}, y_{13}, ..., y_{1n}$	$y_{21}, y_{22}, y_{23}, ..., y_{2m}$	$y_{31}, y_{32}, y_{33}, ..., y_{3q}$

For the 100K MovieLens dataset[1], occupation as a categorical attribute has 21 different categories. Each category corresponds to one possible value of the occupation attribute. Therefore, exactly one of the 21 categories takes on the value 1, and the remaining take on the value 0. Table 1 illustrates the categories of the occupation attribute and their frequency in the 100K MovieLens dataset.

The demographic profile attributes may have only one real value or many categorical values. Table 2 shows a general profile format for two users, user-x, $\mathbf{u}_x$, and user-y, $\mathbf{u}_y$. The first attribute consists of n categories (values) while the second and third attributes consists of m and q categories, respectively. Once formed, user profiles have to be matched to each other to infer some similarity between them. Usually, comparing two users can be done using a distance or a similarity measure which is defined as a function from the users' space to a unipolar interval [0,1] or a bipolar interval [−1,1].

$$\begin{aligned} &dis : \mathbf{u}_x \times \mathbf{u}_y \rightarrow [0, 1] \\ &sim : \mathbf{u}_x \times \mathbf{u}_y \rightarrow [0, 1] \\ &sim : \mathbf{u}_x \times \mathbf{u}_y \rightarrow [-1, 1] \end{aligned} \quad (2)$$

In reality, there are many measures for finding the distance or similarity between two profiles. If we consider the user profile as a vector in the demographic data space, then the most popular measure is the cosine similarity [1,2]. This measure finds the cosine of the angle between the two vectors as a measure for the similarity between them.

$$\cos(\mathbf{u}_x, \mathbf{u}_y) = \frac{\sum_{k=1}^{L} x_{i,k} \times y_{i,k}}{\sqrt{\sum_{k=1}^{L} x_{i,k}^2 \sum_{k=1}^{L} y_{i,k}^2}} \quad (3)$$

where L is the total number of categories (values) of the profile attributes which equals $L = n + m + q$ for the user profile of Table 2. Actually, the dot product of two vectors equals the cosine similarity function if the norm of each vector is unity.

$$\mathbf{u}_x \cdot \mathbf{u}_y = |\mathbf{u}_x||\mathbf{u}_y| \cos(\theta) = \cos(\theta) \quad (4)$$

Otherwise, the dot product has to be divided by the vector norms to get the cosine similarity measure of Formula (3).

$$\cos(\theta) = \frac{\mathbf{u}_x \cdot \mathbf{u}_y}{|\mathbf{u}_x||\mathbf{u}_y|} \quad (5)$$

The range of this measure is actually bipolar −1 to 1 but if all values are positive then we have a positive space and the range is only 0–1.

5. Demographic user profiling approaches

This section introduces five user's profiling approaches for DRS. These approaches take many forms according to the nature

[1] http://www.movielens.umn.edu

of the attributes, the way of representing the profile, and the way of calculating the similarity between profiles. Literature has discussed simple formats of the first three approaches as parts of the user profile for hybrid recommender systems. However, this paper will discuss them in details for the pure DRS along with their representation ways, and variants. We associate a suitable similarity measure with each profiling approach and propose similarity measures for some variants whenever required.

For approach-A, we will discuss all issues related to this approach regarding mapping distance value to a similarity value, age scaling and its effect on the similarity computation, and finally unusual attributes and the way of dealing with them. For approach-B, we will discuss unipolar similarity and will introduce a bipolar similarity at two-levels of abstraction, attribute-level and category-level. The bipolar similarity is important when we take into account the dissimilar cases. Hence, the zero similarity value will represent the neutral case which will occur when the positive cases equal exactly negative cases. Calculating similarity at two levels facilitates the similarity computation process but the similarity value must be the same at these two levels. Approaches D and E are novel and are new ways for profiling users based on different attributes.

5.1. Approach-A (mixed profiling approach)

This approach takes age as a numeric attribute while gender and occupation as categorical attributes. Thus, the user profile represents a mixed-attribute data. In this case, we cannot use a unified similarity measure for this heterogeneous profile. Instead, we have to measure the similarity of each attribute separately and then aggregate them for an overall similarity value. Overlap measure (simple matching measure) (OM) is suitable for the similarity computation between gender and occupation categorical attributes [30].

$$sim(x_i, y_i) = \begin{cases} 1 & x_i = y_i \\ 0 & x_i \neq y_i \end{cases} \tag{6}$$

On the other hand, a simple distance measure, Manhattan distance (MD) [30], can be used for the age numeric attribute.

$$dis(x_i, y_i) = |x_i - y_i| \tag{7}$$

From the data mining point of view, age, gender, and occupation represent different scales of reference and therefore are not comparable to one another using a unified measure with their raw values. If so, gender and occupation attributes will be dominated by the age attribute which has larger magnitude. This deemphasizes the contribution of the gender and occupation attributes in the similarity computation process. For example, Manhattan distance between two age values may exceed 1 by a large value while it is either 0 or 1 only for gender and occupation.

Another point has to be arisen here about the distance value of the age attribute which has to be mapped to a similarity value using some mapping. Usually, the similarity measure is assumed a monotonically deceasing function of the distance measure. This paper maps the distance value to a similarity value using a negative exponential decay function [31].

$$sim(x_i, y_i) = \exp\left(-\eta\ dis^{\alpha}(x_i, y_i)\right) \tag{8}$$

where $0 < \eta < \infty$ is a scale factor, and α is a positive parameter. Al-Shamri [32] found that $\eta = 3.8$, and $\alpha = 2$ give a good mapping of similarity values that nearly cover the whole range between 0 and 1. Formula (8) maps distance values that range from 0 to 1 to similarity values between 0 and 1. However, this is not the case when Manhattan distance is used for the age attribute which exceeds one. To solve this problem, we use min–max scaling to normalize age attribute values to the range [0, 1]. The scaled value of the age of user-x is given by the following formula [30]:

$$xn_1 = \frac{x_1 - \min_a}{\max_a - \min_a} \tag{9}$$

where $\max_a$ and $\min_a$ represent the minimum and maximum age values in the dataset, respectively. At the profile level, the overall similarity value between any two profiles having N attributes is the sum of the individual attribute similarities, $sim(x_i, y_i)$ [30].

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \sum_{i=1}^{N} sim(x_i, y_i) \tag{10}$$

However, the range of this formula is [0, N] not [0,1]. Therefore we have to divide it by N to get a unipolar range, [0,1]. Dividing this formula by N means that all individual similarities are equiprobable at the aggregation process with a probability $1/N$. Alternatively, we can use the following weighted aggregation function [30]:

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \sum_{i=1}^{N} \lambda_i \cdot sim(x_i, y_i)$$
$$\sum_{i=1}^{N} \lambda_i = 1 \tag{11}$$

The weight λ_i gives the relative importance of the ith attribute. This weight can represent the fraction of that attribute in the dataset. Another point may arise about the large number of occupation categories associated with 100K MovieLens dataset. However, this number can be reduced if we merge close occupations together. This will reduce the number of occupation values to eight instead of 21 as given in Table 3.

Table 3
Occupation catogories after merging.

Occupation ID	Occupation name	Frequency
1	Administrator, executive, marketing, salesman	149
2	Doctor, healthcare	23
3	Homemaker, retired	21
4	None, other	114
5	Scientist, programmer, engineer, technician	191
6	Educator, librarian, lawyer	158
7	Entertainment artist, writer	91
8	Student	196

At this point one may ask about the unusual similarities or differences which are statistically more significant than those are common. These values are more informative than the other values. There are two approaches to generalize OM to deal with unusual cases, namely the inverse occurrence frequency and Goodall measure [30]. These measures assign a higher similarity weight to a match if the attribute value is infrequent. The inverse occurrence frequency (IOF) weights the similarity between attribute values by an inverse function of the frequency of the attribute value. If $p_k(x_i)$ is the fraction of users in which the kth attribute takes on the value of x_i in the dataset, then the similarity value will be:

$$sim(x_i, y_i) = \begin{cases} p_k(x_i)^{-2} & x_i = y_i \\ 0 & x_i \neq y_i \end{cases} \tag{12}$$

However, this measure is not appropriate for our application as it is because its value will be very large for small fractions. For example, $p_k(\text{'Female'}) = 0.29$ as illustrated in Table 4 and thus IOF for this ratio is $0.29^{-2} = 11.89$. This factor will surpass that of age by 12 times if we assume the maximum age similarity of 1. The case becomes worse for the occupation attribute because some of its categories have very less number of occurrences. For example, 'doctor' occupation category has a ratio of 0.0074 and thus the IOF

Table 4
Gender distribution of the 100K MovieLens dataset.

Gender	Frequency	Percentage
Male	670	71
Female	273	29

Table 5
Set of users for Example 1.

Name	Age	Gender	Occupation
Alex	23	Male	Engineer
John	55	Male	Scientist
Jeniffer	30	Female	Healthcare
Sarah	42	Female	Doctor

for this ratio is $0.0074^{-2} = 18,148$. We have two ways to deal with this case, the first one is to use another measure like Goodall measure and the second one is to normalize this measure as below.

$$sim(x_i, y_i) = \begin{cases} \frac{p_k(x_i)^{-2}}{\sum_{j=1}^{K} p_k(x_j)^{-2}} & x_i = y_i \\ 0 & x_i \neq y_i \end{cases} \quad (13)$$

On the other hand, a simple variant of Goodall measure (GM) is given below [30].

$$sim(x_i, y_i) = \begin{cases} 1 - p_k(x_i)^2 & x_i = y_i \\ 0 & x_i \neq y_i \end{cases} \quad (14)$$

The range of this measure lies always between 0 and 1 thus it does not require any normalization as we do with the first measure. This approach reflects the real case of each attribute of the profile. However, it needs calculating individual similarities separately and then combines them to get the overall similarity value.

Example 1:

Assume we have four users whose demographic data are illustrated in Table 5. The scaled age values for them and the age distance values between them are illustrated in Table 6. The raw age distance is high and this will dominate the distance of the other attributes in the profile. However, the scaled age distance is always between 0 and 1. The maximum and minimum age values of the 100K MovieLens dataset are 7 and 73, respectively. Table 7 lists the computed similarity values between the four users for mixed-attribute profile approach. Each cell gives individual attribute similarity values and the overall similarity value. For example (0.409, 0.143, 0) represent individual similarity values for age, gender, and occupation attributes respectively and 0.184 represents the overall similarity value. The similarity values of the Goodall measure variant are higher than that of IOF variant and hence this will increase the neighbor contribution in the prediction process.

5.2. Approach-B (categorical profiling approach)

This approach converts age attribute to a categorical attribute as the MovieLens team do with the 1M MovieLens dataset. Table 8 lists the age categories and their frequencies for the 100K MovieLens dataset. This way of profiling unifies the attribute types and thus the similarity value is 1 if the categories of the two attributes are similar. Otherwise, the similarity value is 0 for unipolar (one-sided) similarity range, [0,1], and −1 for bipolar (two-sided) similarity range, [−1,1].

The profile vector can be represented at the attribute-level and thus it is an N-dimensional vector or at the category-level and thus it is an L-dimensional vector, where L is the total number of categories of all attributes of the profile. For example, Alex profile who is 23 years old, male and an engineer is {23, male, engineer} at the attribute-level while it is ⟨0,1,0,0,0;1,0;0,0,1⟩ at the category-level. The corresponding vector of a given attribute is formed by inserting a value 1 for the true category of the attribute and 0 for the remaining categories. For example, age 15 belongs to the first category of Table 8 and thus the corresponding vector of this attribute is ⟨1, 0, 0, 0, 0, 0, 0⟩.

Accordingly, two similarity measures can be defined for the profile vectors, attribute-level similarity measure and category-level similarity measure. The output value of the two measures has to be the same and the difference is only at their granularity level.

5.2.1. Attribute-level similarity computation

Simply, we can define the similarity measure at the attribute-level by the number of attributes with the same category values, N^+, divided by the total number of attributes, N.

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{N^+}{N} \quad (15)$$

Definitely, the range of this measure is [0,1]. This measure counts only the number of similar attributes and ignores the number of dissimilar attributes, N^-. The similarity value for individual attribute is obtained using Formula (6) and thus N^+ can be expressed as below:

$$N^+ = \sum_{i=1}^{N} sim(x_i, y_i) \quad (16)$$

In order to get a bipolar similarity range, [−1,1], we have to subtract the number of dissimilar attributes from that of similar attributes given by Formula (15).

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{N^+}{N} - \frac{N^-}{N} = \frac{N^+ - N^-}{N} = \frac{2N^+ - N}{N} \quad (17)$$

That means the following bipolar similarity measure is used for individual values instead of Formula (6):

$$sim(x_i, y_i) = \begin{cases} +1 & x_i = y_i \\ -1 & x_i \neq y_i \end{cases} \quad (18)$$

Consequently, the numerator, $N^+ - N^-$, can be expressed by

$$N^+ - N^- = \sum_{i=1}^{N} sim(x_i, y_i) \quad (19)$$

Subtracting the dissimilarity value from the similarity value filters neighbors by deleting those are more than 50% dissimilar to

Table 6
Age distance values for users of Example 1 with and without scaling.

			Alex	John	Jeniffer	Sarah	
			23	55	30	42	Age
			0.242	0.727	0.348	0.530	Scaled age
Age distance	Alex	23		0.485	0.106	0.288	Scaled age distance
	John	55	22		0.379	0.197	
	Jeniffer	30	7	25		0.182	
	Sarah	42	19	13	8		

Table 7
Mixed-attribute profile similarity values for users of Example 1 with Goodall measure and IOF.

		Alex	John	Jeniffer	Sarah	
Similarity (Goodall)	Alex		(0.409, 0.143, 0) 0.184	(0.958, 0, 0) 0.319	(0.73, 0, 0) 0.243	Similarity (IOF)
	John	(0.409, 0.496, 0) 0.301		(0.579, 0, 0) 0.193	(0.863, 0, 0) 0.287	
	Jeniffer	(0.958, 0, 0) 0.319	(0.579, 0, 0) 0.193		(0.882, 0.857, 0) 0.579	
	Sarah	(0.73, 0, 0) 0.096	(0.863, 0, 0) 0.287	(0.882, 0.916, 0) 0.599		

Table 8
Age categories and their frequency.

Age group ID	Group specifications	Frequency
1	Under 18	36
18	18–24	198
25	25–34	310
35	35–44	194
45	45–49	80
50	50–55	73
56	56+	52

a given active user. For unipolar similarity, $sim(\mathbf{u}_x, \mathbf{u}_y) = 0$ means either that the users are dissimilar or they are not related to each other at all. There is no measure for neutral cases. However, for bipolar similarity, $sim(\mathbf{u}_x, \mathbf{u}_y) = 0$ means either that the two users have the same value for similarity and dissimilarity (neutral case) or they are not related to each other at all. Popular correlation coefficients are bipolar in their values and hence they can identify both extremes easily.

5.2.2. Category-level similarity computation

The similarity measure at the category-level is the dot product of the two vectors in consideration. Keep in mind that 1 means belongingness to the category and 0 not.

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{\mathbf{u}_x \cdot \mathbf{u}_y}{N} \tag{20}$$

The dot product is divided by the total number of attributes of the profile to get a unipolar similarity value that ranges from 0 to 1. This measure is 1 if the value of the dot product is N which means that all corresponding attributes have the same value. Similar approach is used by [15] with a 27 profile size and binary values for each attributes. On the other hand, we can find a bipolar similarity value at the category-level using the following formula:

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{2(\mathbf{u}_x \cdot \mathbf{u}_y) - N}{N} \tag{21}$$

This bipolar similarity measure has a value of 1 when the value of the dot product is N. That means all attributes agree with each other. On the opposite side, it is -1 if the dot product is zero. That means all attributes disagree each other. The similarity value is 0 if there are as much agreements as disagreements at the attribute level. However, this measure is correct only if the values of each attribute is 1 for belonging to the category and zero otherwise. This approach simplifies the similarity computation using the dot product of two vectors but at the cost of losing some exact information about some attributes.

Example 2: Assume a user profile consists of three attributes having three categories each. The first user profile is $X = \langle 1, 0, 0; 0, 1, 0; 0, 0, 1\rangle$ and the second user profile is $Y = \langle 0, 1, 0; 0, 1, 0; 0, 0, 1\rangle$. Clearly, the category of the first attribute of the first user is category one while it is category two for the second user. The categories of the other two attributes are similar. The unipolar similarity value at the attribute-level is obtained as below:

$$N^+ = \sum_{i=1}^{3} sim(x_i, y_i) = 2$$

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{N^+}{N} = \frac{2}{3}$$

The bipolar similarity value at the attribute-level is obtained as below:

$$N^+ - N^- = \sum_{i=1}^{3} sim(x_i, y_i) = -1 + 1 + 1 = 1$$

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{N^+ - N^-}{N} = \frac{1}{3}$$

The unipolar similarity value at the category-level is:

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{\mathbf{u}_x \cdot \mathbf{u}_y}{N} = \frac{1 \times 0 + 1 \times 1 + 1 \times 1}{3} = \frac{2}{3}$$

The bipolar similarity value at the category-level is

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{2(\mathbf{u}_x \cdot \mathbf{u}_y) - N}{N}$$

$$sim(\mathbf{u}_x, \mathbf{u}_y) = \frac{2(1 \times 0 + 1 \times 1 + 1 \times 1) - 3}{3} = \frac{1}{3}$$

Clearly, the bipolar similarity value is less because we consider the dissimilarity value between the two profiles in the process.

5.3. Approach C (fuzzy profiling approach)

This approach tries to exploit the vague nature of the age attribute. Subtracting the crisp age values may mislead the system because any two users having close ages are similar 100% from the fuzzy point of view. The crisp age value can be represented as a fuzzy value that has membership values to many fuzzy sets. This paper treats age as a fuzzy variable with three fuzzy sets, young, middle-aged and old with the following membership functions [12]:

$$A_{Young}(x) = \begin{cases} 1 & x \le 20 \\ (35 - x)/15 & 20 < x \le 35 \\ 0 & x > 35 \end{cases} \tag{22a}$$

$$A_{Middle}(x) = \begin{cases} 0 & x \le 20, x > 60 \\ (x - 20)/15 & 20 < x \le 35 \\ 1 & 35 < x \le 45 \\ (60 - x)/15 & 45 < x \le 60 \end{cases} \tag{22b}$$

$$A_{Old}(x) = \begin{cases} 0 & x \le 45 \\ (x - 45)/15 & 45 < x \le 60 \\ 1 & x > 60 \end{cases} \tag{22c}$$

The graph of the age fuzzy sets is given in Fig. 1. Accordingly, to compare two age values we need a distance measure that should reflect the fuzzy nature of these values. One choice is the fuzzy

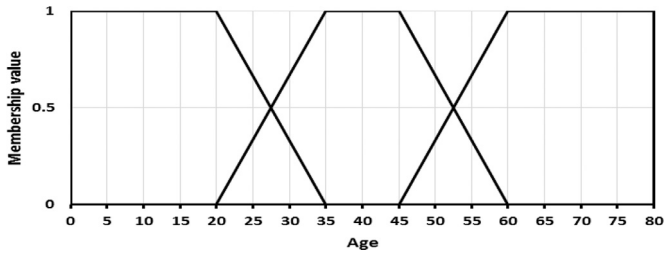

Fig. 1. Age fuzzy sets.

distance function defined by Al-Shamri and Bhardwaj [10].

$$fd(x_1, y_1) = d(\mathbf{x}_1, \mathbf{y}_1) \times d(x_1, y_1)$$

$$d(\mathbf{x}_1, \mathbf{y}_1) = \sqrt{\sum_{j=1}^{l} (x_{1,j} - y_{1,j})^2} \quad (23)$$

where l is the total number of fuzzy sets for the age attribute, and $x_{1,j}$ is the membership value of the age value in the jth fuzzy set. The membership vectors, $\mathbf{x}_1$ and $\mathbf{y}_1$, correspond to two age crisp values, x_1, and y_1. This measure finds the distance at two different levels, crisp value level and fuzzy sets level and then multiplies the two values. The remaining two attributes, gender and occupation can be considered as fuzzy points with a membership value of one to the corresponding category.

The maximum and minimum age values of the 100K MovieLens dataset are 7 and 73, respectively. Hence, to get a reasonable difference range that easily mapped to a similarity value, we have to divide the fuzzy distance value by the maximum possible value for this dataset. The maximum fuzzy distance value [10] for age values 7 and 73 is

$$fd(x_1, y_1) = d(\mathbf{x}_1, \mathbf{y}_1) \times d(x_1, y_1) = \sqrt{2} \times 66 = 93.332$$

Another variant of this approach can be obtained by considering the fuzzy sets of the crisp age value as categories. The membership value of each fuzzy subset will be the category value. Hence, Formulas (20) and (21) can be used for comparing two vectors (profiles). This approach reveals the importance of the age attribute as a fuzzy variable. However, the processing complexity increases somehow because of calculating the fuzzy distance.

Example 3: This example computes the similarity values for the fuzzy profiling approach with fuzzy distance and with cosine similarity as illustrated in Table 9. For the fuzzy distance, each cell gives individual attribute similarity values and the overall similarity value. The first value represents fuzzy similarity value for age while the remaining two are for gender, and occupation attributes respectively. For cosine similarity variant, the user profile for Alex who is 23 years old, male and an engineer is ⟨0.8, 0.2, 0; 1, 0; 0, 1, 0⟩. Each fuzzy subset is considered as a category of the age attribute and its value is the membership value to that fuzzy subset. The similarity values of the cosine similarity variant are higher than that of fuzzy distance variant and hence this will increase the neighbor contribution in the prediction process.

5.4. Approach-D (cascaded profiling approach)

This approach takes age as a numeric attribute and the only factor for obtaining the user similarity in the beginning. Then, the system elects a big set of neighbors based on the age similarity values. This big set will be the input to another system to refine it based on both age and gender attributes. The result of this is a new smaller set of neighbors that finally are refined based on age, gender, and occupation attributes. The resulting neighborhood set is the smallest among all sets. The process of this approach is illustrated in Fig. 2.

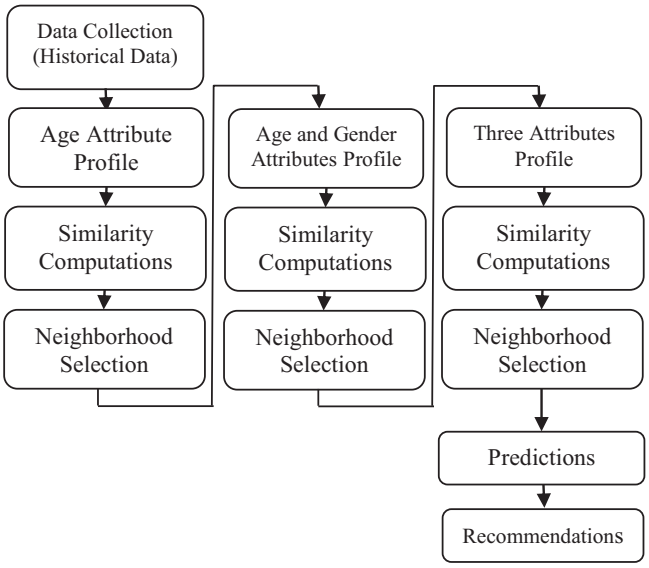

Fig. 2. Approach-D cascaded profiling approach.

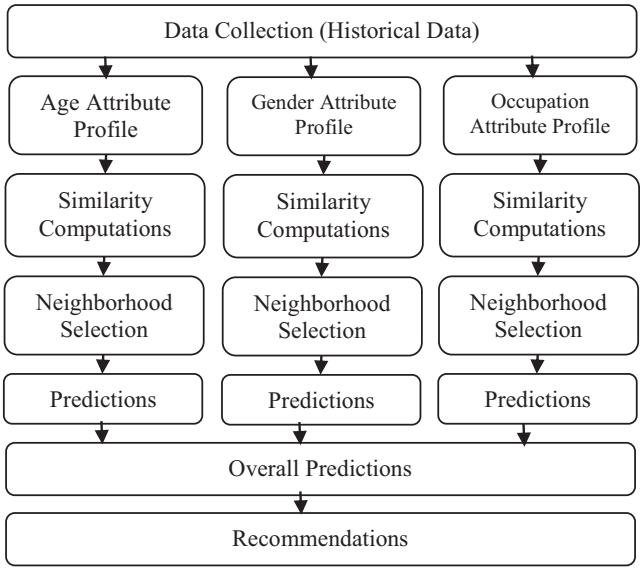

Fig. 3. Approach-E single attribute profiling approach.

Equivalently, we can see the user profile as a hierarchy of three coincide circles (set of neighbors). The first and the larger one is the age circle, followed by a smaller circle of both age and gender attributes and then the smallest circle of age, gender and occupation attributes. Hence, three incomplete DRSs are built to produce set of neighbors for the next system thus the circle getting smaller each time. Incomplete DRS means that the system does not go through the all five phases of the usual DRS. The first stage in the hierarchy carries out only four phases while the second stage implements three phases out of five to produce the set of neighbors.

5.6. Approach-E (single-attribute profiling approach)

This approach sees each attribute as a profile and thus we have three isolated DRSs, age DRS (ADRS), gender DRS (GDRS), and occupation DRS (ODRS). Each one of them completes four phases of the recommender system separately. Only the fifth phase is aggregated for all of them as illustrated by Fig. 3. Hence, each system gives a prediction separately for the unpredicted items and then

Table 9
Fuzzy profiling approach similarity values for users of Example 1 with fuzzy distance and cosine similarity.

		Alex	John	Jeniffer	Sarah	
Cosine similarity	Alex		(0.611, 1, 0) 0.537	(0.991, 0, 0) 0.330	(0.818, 0, 0) 0.272	Fuzzy distance
	John	0.689		(0.834, 0, 0) 0.278	(0.937, 0, 0) 0.312	
	Jeniffer	0.467	0.407		(0.986,1, 0) 0.662	
	Sarah	0.400	0.444	0.889		

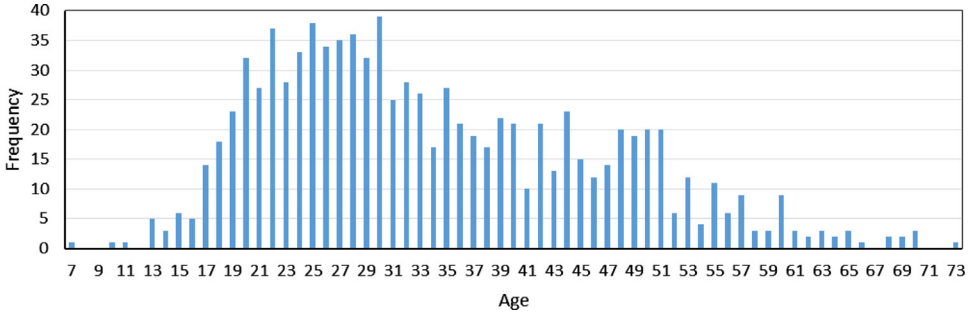

Fig. 4. 100K MovieLens dataset age distribution.

the overall prediction is aggregated using a simple averaging or a maximum averaging.

$$pr_{i,k}(x) = \frac{1}{P}\sum_{i=1}^{P} apr_{i,k}(i) \tag{24a}$$

$$pr_{i,k}(x) = \max_{i=1}^{P} apr_{i,k}(i) \tag{24b}$$

where P is the number of systems contributing a prediction for the item in consideration and $apr_{i,k}(i)$ is the prediction value of the ith system . This new approach brings to light the contribution of each attribute and therefore its advantages in the process as it is the only profile for the system. This hinders the compensation effect that other approaches have where the effect of some attributes may be compensated by others.

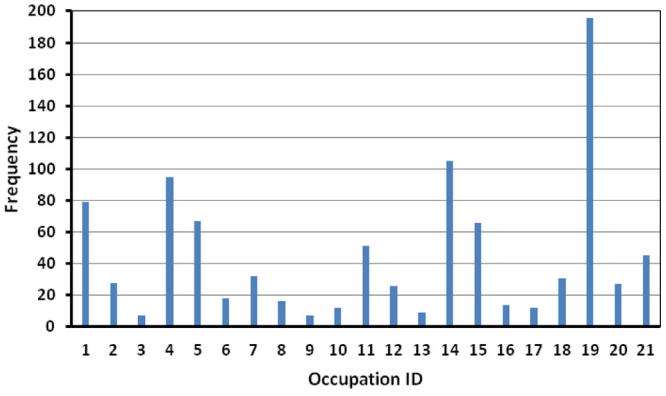

Fig. 5. 100K MovieLens dataset occupation distribution.

6. Dataset and the experiments

This section discusses the dataset, the methodology used for preparing it for our experiments, the conducted experiments, and the metrics used for evaluating the system performance.

6.1. Experiment dataset

We conduct our experiments using the 100K MovieLens dataset which consists of 100,000 ratings assigned by 943 users on 1820 movies. The actual age value for each user is given in this dataset which is unreachable with 1M MovieLens dataset. Other demographic data is given also like gender, and occupation. The age, gender, and occupation distributions of this dataset are given in Fig. 4, Table 5, and Fig. 5, respectively. Male population is more than twice that of female population.

Direct division of the dataset into five splits gives 188.6 users for each split. However, we round it to 185 users and discarded 18 male users having 'none' or 'other' occupation categories with ratings less than 35. The remaining 925 users are divided into 5 mutually exclusive folds, fold(1), …, fold(5), each of which having the same size, 185 users and each fold mimics the whole dataset gender distribution.

We follow leave-one-out cross validation approach used by Al-Shamri [33] for training and testing the system. That is in Split-i dataset, fold(i) is reserved as the test set and the remaining folds are collectively used to train the system. Hence, each fold is used the same number of times for training and once for testing. The number of total users, training users, and active users are $M = 925$, $M_T = 740$, and $M_A = 185$, respectively. During the testing phase, the system treats the user ratings as unseen ratings that the system would attempt to predict [2].

6.2. Evaluation metrics

This paper employs MAE (Mean Absolute Error) and RMSE (Root Mean Squared Error) as prediction accuracy metrics. The MAE measures the deviation of predictions generated by the DRS from the true ratings specified by the active user [2,33–35]. The split MAE over all the active users (M_A) is given by:

$$\text{MAE} = \frac{1}{M_A}\sum_{i=1}^{M_A}\left(\frac{1}{N_i^p}\sum_{k=1}^{N_i^p}|pr_{i,k} - r_{i,k}|\right) \tag{25}$$

Lower MAE corresponds to more accurate predictions of a given DRS. The split RMSE over all the active users (M_A) is given by [34]:

Table 10
Approach-A (mixed profiling approach).

Age (numeric)	Gender (categorical)	Occupation (categorical)	P_x	Similarity measure	Code
Normalized Real Value	F M	1 … 20	23	MD for age and OM for gender and occupation	AMO23
		1 … 8	11		AMO11
		1 … 20	23	MD for age and IOF for gender and occupation	AMI23
		1 … 8	11		AMI11
		1 … 20	23	MD for age and GM for gender and occupation	AMG23
		1 … 8	11		AMG11

Table 11
Approach-B (categorical profiling approach).

Age (categorical)	Gender (categorical)	Occupation (categorical)	P_x	Similarity measure	Code
1 … 7	F M	1 … 20	29	Unipolar Dot product	UDP29
		1 … 8	16		UDP17
		1 … 20	29	Bipolar dot product	BDP29
		1 … 8	16		BDP17

$$\text{RMSE} = \frac{1}{M_A}\sum_{i=1}^{M_A}\left(\sqrt{\frac{1}{N_i^p}\sum_{k=1}^{N_i^p}\left(pr_{i,k} - r_{i,k}\right)^2}\right) \tag{26}$$

RMSE penalizes large errors more heavily because it squares errors before summing them. The MAE (RMSE) over all splits is the average of all splits' MAEs (RMSE) [33].

Coverage and percentage of the correct predictions (PCP) are used also for evaluating the system performance [2, 34–35]. Coverage measures the percentage of items for which a DRS can provide predictions. We compute the active user coverage as the number of items for which the DRS can generate predictions for that user over the total number of unseen items. The split coverage over all the active users is:

$$\text{Coverage} = \frac{\sum_{i=1}^{M_A} N_i^p}{\sum_{i=1}^{M_A}\left|S_i^{TE}\right|} \tag{27}$$

Here, N_i^p is the total number of predicted items for user u_i, and M_A is the total number of the active users. The PCP is the percent of the correctly predicted items by the system to the total number of items in the test ratings set of the active user. The set of correctly predicted items for a given active user and the split PCP over all the active users are defined by the following formulae [12,33]:

$$CorrectSet(u_a) = \left\{s_k | s_k \in S_a^{TE},\ pr_{a,k} = r_{a,k}\right\} \tag{28}$$

$$PCP = \frac{\sum_{i=1}^{M_A}|CorrectSet(u_i)|}{\sum_{i=1}^{M_A}\left|S_i^{TE}\right|} \times 100\% \tag{29}$$

Low Coverage value indicates that the CRS will not be able to assist the user with many of the items he has not rated. Over all splits, we compute PCP (Coverage) by summing all correct predictions (predictions) over all active users over all splits and divided it by the sum of all testing set sizes of all active users over all splits.

6.3. Conducted experiments

This subsection discusses the conducted experiments for each approach and the constraints applied on each one of them. The neighborhood set size N_x is varied from 10 to 100 by a step size of 10 each time, $N_x = \{10, 20, ..., 100\}$ for all experiments so that the system performance under different neighborhood set sizes is tested. For each split, the experiments are performed ten times with each neighborhood set size and the final split results are the average over all neighborhood set sizes. Similarly, at the split-level, all experiments are performed five times with each split and the final evaluation results are the average over all splits.

6.3.1. Approach-A experiments

We conduct six experiments based on six variants of this approach as detailed in Table 10 where $P_{\mathbf{x}}$ is the vector size. The code of each variant is provided for later reference. These variants vary in their similarity computation methods, age values, and the number of values for each categorical attribute. Accordingly, the profile size may be either 10 when we use 8 occupation groups or 23 when we use 20 occupation groups. The employed similarity measures are MD for age attribute and OM, scaled IOF, or GM for gender and occupation attributes. The age value is scaled using the min–max scaling so that the decaying exponential mapping function can be used and hence no one attribute dominates the others.

6.3.2. Approach-B experiments

Four experiments based on four variants are conducted for this approach. Table 11 gives the details and the code of these variants. The profile size can be either 17 or 29. All attributes are categorical attributes therefore the employed similarity measure maybe the unipolar or bipolar similarity measure that are obtained through dot product of the two profile vectors. The first two variants use the unipolar similarity while the second two variants use the bipolar similarity.

6.3.3. Approach-C experiments

For this approach, we conduct four experiments based on four variants as illustrated in Table 12. The profile size is either 10 or 23 based on the number of the categories of each attribute. The first two variants of this approach treat age as a fuzzy variable with three fuzzy sets and treat gender and occupation as fuzzy points.

Table 12
Approach-C (fuzzy profiling approach).

Age (3 fuzzy sets)	Gender (fuzzy point)	Occupation (fuzzy point)	P_x	Similarity measure	Code
Young, Middle-aged, old	F, M	1 … 20	25	Fuzzy distance	CFD25
		1 … 8	12		CFD12
		1 … 20	25	Cosine similarity	CDP25
		1 … 8	12		CDP12

Table 13
Approach-D (cascaded profiling approach).

Age (numeric)	Gender (categorical)	Occupation (categorical)	P_x	Similarity	Code
Normalized Real Value	F, M	1 … 20	1,2,20	Simple average	DFS23
		1 … 8	1,2,8		DFS11
		1 … 20	1,2,20	Weighted average	DFW23
		1 … 8	1,2,8		DFW11

Table 14
Approach-E (single profiling approach).

Age (numeric)	Gender (categorical)	Occupation (categorical)	P_x	Predictions	Code
Normalized Real Value	F, M	1 … 20	1,2,20	Simple average	EAF23
		1 … 8	1,2,8		EAF11
		1 … 20	1,2,20	Maximum	EMF23
		1 … 8	1,2,8		EMF11

We used the normalization process discussed in Section 5.3 for the first two variants of this experiment. The second two variants fill directly the profile fields by the membership value of each fuzzy set and simply employ cosine similarity between these two vectors.

6.3.4. Approach-D experiments

The conducted experiments of this approach are four experiments based on four variants as depicted in Table 13. This approach follows the previous approach by treating gender and occupation attributes as fuzzy points and age attribute as a fuzzy variable with three fuzzy subsets. The difference between this approach and the previous approach lies in the way of profiling users. In this approach, the user is gradually profiled based on one attribute, then two attributes and finally on three attributes. The first two variants of this approach accumulate the similarity values of different attributes and then simply average them.

The second two variants use a weighted averaging where ⟨0.6, 0.4⟩ is used for weighting age and gender attributes that represents the second stage for profiling users while ⟨0.5, 0.3, 0.2⟩ is used for weighting age, gender and occupation attributes respectively for the last stage of profiling. These weights somehow reflect the estimated importance of each attribute in the two-attribute and three-attribute user profiles.

6.3.5. Approach-E experiments

This approach differs with the previous two approaches only in the way of profiling users. Here each attribute represents a profile and hence a DRS system. That means we have three DRSs predicting items to a given active user. These predictions are aggregated for a unified prediction value. The conducted four experiments of this approach are illustrated in Table 14 with two different groups for occupation attribute. The first two variants use a simple averaging for the prediction values while the second two variants use a maximum averaging.

Table 15
Results of the variants of approach-A.

System	PCP	Coverage	RMSE	MAE
AMO23	33.0209752	**92.781837**	1.3399984	0.9901001
AMO11	33.1176542	92.7406412	1.3377216	0.9876198
AMI23	32.8241405	92.474828	1.3426572	0.9940717
AMI11	32.9189147	92.5920458	1.3389854	0.9910757
AMG23	33.0284243	92.7679427	1.3413182	0.9900091
AMG11	**33.1441941**	92.7789409	1.3379914	**0.9869632**
IPV	0.99564	0.33088	0.369285	0.720394

7. Analysis of the experiments results

The impact of the way of profiling users is crucial on the performance of the recommender system as the profile is the platform for acquiring a list of neighbors for the active user. The results show that the performance of all approaches is good even for small neighborhood set size. That means close set of neighbors is usually found from the beginning. The averaged results of all approaches for all metrics are listed in Tables 15–19. The last row in each table is the improvement percentage value between the results of best and worst variant within each approach. The following formulas are used for measuring the increase improvement percentages (for PCP and Coverage) and the decrease improvement percentages (for RMSE and MAE).

$$P_x^+ = \frac{x_{new} - x_{old}}{x_{old}} \times 100\% \tag{30}$$

$$P_x^- = \frac{x_{old} - x_{new}}{x_{old}} \times 100\% \tag{31}$$

Table 16
Results of the variants of approach-B.

System	PCP	Coverage	RMSE	MAE
UDP29	**33.0624246**	**92.4051267**	1.3513237	0.9988486
UDP17	33.0289918	92.3414292	**1.3492657**	**0.9972309**
BDP29	32.7672638	92.1772977	1.3659925	1.0097277
BDP17	32.8426006	92.213867	1.3576811	1.0037884
IPV	0.90	0.25	1.22452	1.23764

The following subsection discusses in details the results of each individual approach. For comparison, we identify the best variant for each approach and then compare it with the best of the other approaches in the last subsection.

7.1. Approach-A results

The results are very close for all variants in terms of all evaluation metrics as shown in Table 15. The increase improvement or decrease improvement percentages between the results of the best and the worst variant are very small and do not exceed 1% for all evaluation metrics. The worst variant is AMI23 which uses 20 occupation categories and IOF for weighting gender and occupation attributes. Actually, IOF gives high weight for rare (odd) values and low weight for popular values. For example, 'doctor' occupation category occurred only 7 times and thus its weight is high, 0.2346 while 'student' occupation category occurred 196 times and thus its weight is low, 0.0003. Consequently, the IOF weighting factor deemphasizes occupation attribute, some occupation categories are nearly ignored, and thus the performance of AMI23 and AMI13 is poor.

The best variant in terms of PCP and MAE is the variant AMG11 which uses Goodall measure and eight occupation groups. Moreover, the Coverage and RMSE of this variant are very close to the best values. That means grouping occupation does enhance the system performance a little bit. Goodall measure performance is better than that of IOF because the weights are always greater than 0.9 even if the frequency of the attribute category is low. That means all occupation categories are contributing with little bit difference in their weights. The highest weight is 0.9999 and the lowest weight is 0.9568.

7.2. Approach-B results

Table 16 depicts the results of this approach which are very close for all variants in terms of all metrics. The increase improvement or decrease improvement percentages between the results of the best and the worst variant are very small and do not exceed 1.5% for all evaluation metrics. The results of this approach show that letting similarity cases cancel dissimilarity cases do not improve the system performance and hence the performance of the variants that use bipolar similarity is poor. Algebraically, summing negative similarity and positive similarity will increase the closeness of the active user to its neighbors. However, this is not always true as the dataset may be skewed as it is with our dataset where ratings 4 and 5 are the popular ratings among users. The best variant in terms of PCP and Coverage is UDP29 with less than 1% improvement percentages while UDP17 is the best in terms of RMSE and MAE with more than 1% improvement percentages.

7.3. Approach-C results

The results listed in Table 17 show that the variants of this approach are almost similar. The improvement percentage values lie below 1%. The best variant in term of Coverage, RMSE, and MAE is CFD25 and the worst one is CDP12. This indicates that fuzzy distance does enhance the system performance and occupation attribute does not matter whether it is 20 or 8 for this approach. Utilizing fuzzy membership values directly into the profile does not improve the system anymore.

Table 17
Results of the variants of approach-C.

System	PCP	Coverage	RMSE	MAE
CFD25	33.0231292	**92.7487387**	**1.3401675**	**0.9902982**
CFD12	33.0751364	92.5774995	1.3426339	0.9911262
CDP25	**33.0921143**	92.6185246	1.3456758	0.9930192
CDP12	33.0205695	92.2988719	1.3532036	0.99857
IPV	0.22	0.49	0.96	0.83

Table 18
Results of the variants of approach-D.

System	PCP	Coverage	RMSE	MAE
DFS23	32.6713795	91.8697137	1.3572379	1.0066882
DFS11	32.7023467	91.8697137	1.3570085	1.0061749
DFW23	32.7129535	91.8697137	1.3565376	1.0056567
DFW11	**32.7208687**	**91.8697137**	**1.3563625**	**1.0054787**
IPV	0.15	0	0.06	0.12

Table 19
Results of the variants of approach-E.

System	PCP	Coverage	RMSE	MAE
EAF23	34.467222	**96.4392071**	**1.2245409**	**0.9109598**
EAF11	**34.4698393**	96.4392071	1.2253386	0.911394
EMF23	34.2306014	96.4392071	1.2824707	0.9435531
EMF11	34.139996	96.4392071	1.2898977	0.9485946
IPV	0.97	0	5.07	3.97

7.4. Approach-D results

All variants of this approach have the same Coverage value, 91.8697137% as shown in Table 18. The improvement percentage values are very small and do not exceed 0.15%. Thus varying occupation groups or the way of aggregating similarity values have a minor effect on the system performance. The best variant is DFW11 and the worst one is DFS23. That indicates weighting the similarities of each stage is important in the process of finding close neighbors for a given active user.

7.5. Approach-E results

The results of this approach are depicted in Table 19. They show a considerable improvement percentage values especially for RMSE and MAE where they reach 5% for RMSE and 4% for MAE. The worst variant is EMF11 while the best variant is EAF23 in terms of Coverage, RMSE, and MAE. The PCP value is very close to that of EAF11. The main differences between EAF23 and EMF11 are prediction aggregation process and the occupation categories. EAF23 uses weighted average while EMI11 uses maximum predictions. That means weighted averaging is better than that of maximum averaging for aggregating predictions.

7.6. All approaches results

The results of the best variant of each approach are depicted in Table 20 and Figs. 6–9 in terms of PCP, Coverage, RMSE and MAE respectively. The best variant of approach-E, EAF23, is the best variant in all aspects and the best variant of approach-D,

Table 20
Results of the best variant of each approach.

System	PCP	Coverage	RMSE	MAE
AMG11	33.1441941	92.7789409	1.3379914	0.9869632
UDP17	33.0289918	92.3414292	1.3492657	0.9972309
CFD25	33.0231292	92.7487387	1.3401675	0.9902982
DFW11	32.7208687	91.8697137	1.3563625	1.0054787
EAF23	**34.4672220**	**96.4392071**	**1.2245409**	**0.9109598**
IPV	5.34	4.97	9.72	9.40

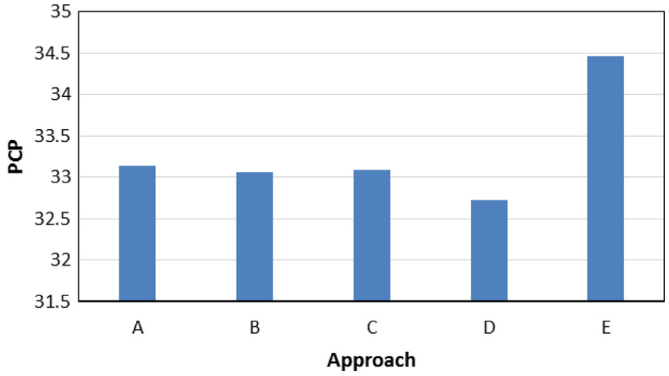

Fig. 6. PCP of the best variant of each approach.

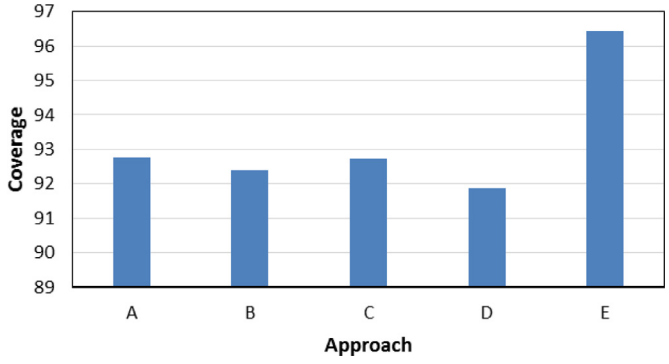

Fig. 7. Coverage of the best variant of each approach.

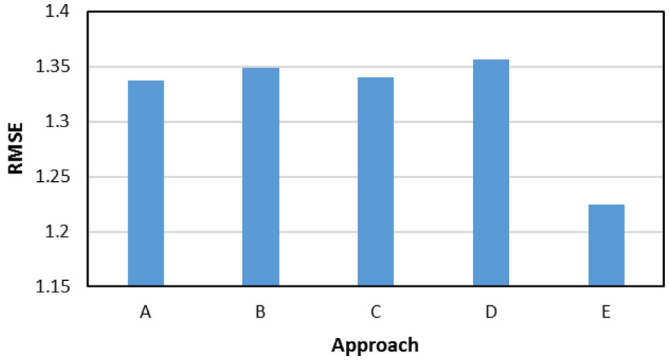

Fig. 8. RMSE of the best variant of each approach.

DFW11, is the worst variant in all aspects. The best variants of the first three approaches are less or more similar in all evaluation metrics and all of them are lying in the middle between the best and worst approaches. The improvement percentage values are high ranging from 4.97% for Coverage to 9.72% for RMSE. These results show that the profiling way of approach-E is the best and gives very good results as compared to the other approaches. This profiling way gains the benefit of each attribute in identifying

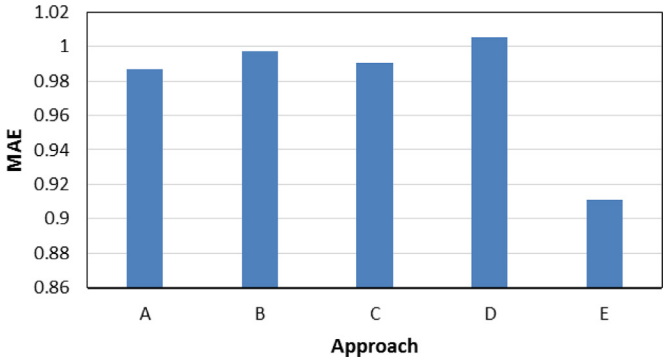

Fig. 9. MAE of the best variant of each approach.

true neighbors and then aggregating their prediction into a unified prediction value. On the other hand, the profiling way of approach-D is the worst as it keeps aside some good possible neighbors from the list of neighbors due to its hierarchical nature.

Obviously, the superiority of EAF23 occurs when we compare its results with that of the second best variant. The improvement percentages of EAF23 over the second best variant, AMG11, are 3.99%, 3.95%, 8.48%, and 7.7% in terms of PCP, Coverage, RMSE, and MAE, respectively. The results of approach-D reveal that all attributes are important if they are considered separately. However, their contribution compensates each other if they considered in a unified profile. The time complexity of each profiling approach does not exceed $O(P_x)$, where P_x is the profile size which cannot exceed 29 for all the examined approaches.

8. Conclusions

This work is an attempt to uncover some aspects of demographic recommender systems which have been underestimated and do not gain much attention till now although they mimic many real life decisions. Broadly, we can summarize profiling approaches of demographic recommender systems into two main approaches, a unified profiling approach (approaches A, B, and C) and an isolated profiling approach (approaches D and E). All approaches belonging to the unified profiling approach are very close in their results and preforming closely similar with some minor differences. The variants within each profiling approach are almost similar in their results and the improvement of the best variant over the worst one is minor for many approaches. This improvement is not a good superiority indicator of the best variant over the worst one. However, the minor changes show that there are some priorities for some attributes, for example age is more important than the others for some approaches.

The results of the unified profiling approach show that the system performance does not improve considerably with different ways of representing the profile attributes. However, once we change the profiling way, new results occur positively as in approach-E or negatively as in approach-D. The reasonable improvements mean that close neighbors and consequently predictor are found and hence the way of profiling users play an important role for enhancing the system performance.

This paper conducts experiments using three demographic attributes, as they are the only available for MovieLens dataset. However, many other can be used for other applications. In the future work, we will examine the studied approaches for different datasets and try to explore more approaches for them. Moreover, we will try hybridization between DRS with different approaches and the collaborative recommender system.

Acknowledgment

The author would like to thank very much the anonymous reviewers for their valuable comments.

References

[1] J. Bobadilla, F. Ortega, A. Hernando, A. Gutirrez, Recommender systems survey, Knowledge-Based Syst. 26 (2013) 109–132.
[2] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions, IEEE Trans. Knowl. Data Eng. 17 (6) (2005) 734–749.
[3] R. Burke, Hybrid recommender systems: survey and experiments, User Model. User-Adapted Interact. 12 (2002) 331–370.
[4] M. Montaner, B. Lopez, J.L. De La, A taxonomy of recommender agents on the Internet, Artif. Intell. Rev. 19 (2003) 285–330.
[5] M. de Gemmis, L. Iaquinta, P. Lops, C. Musto, F. Narducci, G. Semeraro, Preference learning in recommender systems, in: Proceedings of ECML/PKDD-09 Workshop on Preference Learning (PL 09), 2009.
[6] J. Lu, D. Wu, M. Mao, W. Wang, G. Zhang, Recommender system application developments: a survey, Decis. Support Syst. 74 (2015) 12–32.
[7] Z. Zhang, H. Lin, K. Liu, D. Wu, G. Zhang, J. Lu, A hybrid fuzzy-based personalized recommender system for telecom products/services, Inf. Sci. 235 (2013) 117–129.
[8] E. Rich, User modeling via stereotypes, Cognit. Sci. 3 (1979) 329–354.
[9] D. Goldberg, D. Nichols, B.M. Oki, D. Terry, Using collaborative filtering to weave an information Tapestry, Commun. ACM 35 (12) (1992) 61–70.
[10] U. Shardanand, P. Maes, Social information filtering: algorithms for automating 'Word of Mouth, in: Proceedings of the Conference on Human Factors in Computing Systems (CHI'95), Denver, 1995, pp. 210–217.
[11] B. Krulwich, Lifestyle finder: intelligent user profiling using large-scale demographic data, Artif. Intell. Mag. 18 (2) (1997) 37–45.
[12] M.Y.H. Al-Shamri, K.K. Bharadwaj, Fuzzy-genetic approach to recommender systems based on a novel hybrid user model, Expert Syst. Appl. 35 (3) (2008) 1386–1399.
[13] M.Y.H. Al-Shamri, K.K. Bharadwaj, A compact user model for hybrid movie recommender system, in: Proceedings of the Seventh IEEE International Conference on Computational Intelligence and Multimedia Applications (ICCIMA'07), Tamil Nadu, India, 2007, pp. 519–524.
[14] M.Y.H. Al-Shamri, K.K. Bharadwaj, A hybrid preference-based recommender system based on fuzzy concordance/discordance principle, in: Proceedings of the Third Indian International Conference on Artificial Intelligence (IICAI'07), Pune, India, 2007, pp. 301–314.
[15] M. Vozalis, K.G. Margaritis, Collaborative filtering enhanced by demographic correlation, in: Proceedings of the 18th World Computer Congress AIAI Symposium on Professional Practice in AI, 2004.
[16] M. Vozalis, K. Margaritis, Using SVD and demographic data for the enhancement of generalized collaborative filtering, Inf. Sci. 177 (15) (2007) 3017–3037.
[17] L. Safoury, A. Salah, Exploiting user demographic attributes for solving cold-start problem in recommender system, Lecture Notes Softw. Eng. 1 (3) (2013).
[18] E.B. Junior, M.G. Manzato, R. Goularte, Evaluating the impact of demographic data on a hybrid recommender model, IADIS Int. J. WWW/Internet 12 (2) (2014) 149–167.
[19] M. Ghazanfar, and A. Prugel-Bennett, A scalable, accurate hybrid recommender system, in: Proceedings of the International Conference on Knowledge Discovery and Data Mining (WKDD 2010), pp. 94–98.
[20] J. Beel, S. Langer, A. Nürnberger, M. Genzmehr, The impact of demographics (age and gender) and other user characteristics on evaluating recommender systems, in: Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries (TPDL 2013), Valletta, Malta, Springer, 2013, pp. 400–404.
[21] J. Sobecki, Implementations of web-based recommender systems using hybrid methods, Int. J. Comput. Sci. Appl. 3 (3) (2006) 52–64.
[22] M.J. Pazzani, A framework for collaborative, content-based and demographic filtering, Artif. Intell. Rev. 13 (5–6) (1999) 393–408.
[23] SilviaN. Schiaffino, Analía Amandi, Building an expert travel agent as a software agent, Expert Syst. Appl. 36 (2) (2009) 1291–1299.
[24] Antonio Moreno, Aida Valls, David Isern, Lucas Marin, Joan Borràs, SigTur/E-Destination: ontology-based personalized recommendation of tourism and leisure activities, Eng. Appl. Artif. Intell. 26 (2013) 633–651.
[25] A. Said, T. Plumbaum, W.E. De Luca, S. Albayrak, A comparison of how demographic data affects recommendation, in: Proceedings of the 19th International Conference on User Modeling, Adaption, and Personalization, 2011.
[26] J. Lu, Q. Shambour, Y. Xu, Q. Lin, G. Zhang, A web-based personalized business partner recommendation system using fuzzy semantic techniques, Comput. Intell. 29 (1) (2013) 37–69.
[27] Y. Yujie, Z. Huizhi, W. Xianfang, On alleviation of new user problem in collaborative filtering using SNA theory, Int. J. u- e-Serv. Sci. Technol. 6 (2013) 133–144.
[28] T. Chen, L He, Collaborative filtering based on demographic attribute vector, in: Proceedings of the International Conference on Future Computer and Communication, 2009, pp. 225–229.
[29] I. Weber, C. Castillo, The demographics of web search, in: Proceedings of the Conference on Research and development in information retrieval (SIGIR), ACM, 2010.
[30] C.C. Aggarwal, Data Mining: The Textbook, Springer Cham, Heidelberg, New York, 2015.
[31] U. Luxburg, Statistical learning with similarity and dissimilarity functions (Ph.D. thesis), Technical University of Berlin, Berlin, Germany, 2004.
[32] M.Y.H. Al-Shamri, Fuzzy and genetic algorithm approaches to recommender systems (Ph.D. thesis), Jawaharlal Nehru University, New Delhi, India, 2008.
[33] M.Y.H. Al-Shamri, Power coefficent as a similarity measure for collaborative recommender system, Expert Syst. Appl. 41 (2014) 5680–5688.
[34] J.L. Herlocker, L.A. Konstan, L.G. Terveen, J.T. Riedl, Evaluating collaborative filtering recommender systems, ACM Trans. Inf. Syst. 22 (1) (2004) 5–53.
[35] E. Vozalis, K. Margaritis, "Analysis of recommender systems' algorithms, in: Proceedings of the Sixth Hellenic-European Conference on Computer Mathematics and its Applications (HERCMA), Athens, Greece, 2003.