Expert Systems with Applications 39 (2012) 7451–7457

Contents lists available at SciVerse ScienceDirect

Expert Systems with Applications

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

Property-based collaborative filtering for health-aware recommender systems ☆

Martín López-Nores *, Yolanda Blanco-Fernández, José J. Pazos-Arias, Alberto Gil-Solla

Department of Telematics Engineering, University of Vigo, EE Telecomunicación, Campus Universitario s/n, 36310 Vigo, Spain

ARTICLE INFO

Keywords:
Recommender systems
Collaborative filtering
Semantics
Electronic health records

ABSTRACT

Recommender systems aim at solving the problem of information overload by selecting items (commercial products, educational assets, TV programs, etc.) that match the consumers' interests and preferences. Recently, there have been approaches to drive the recommendations by the information stored in electronic health records, for which the traditional strategies applied in online shopping, e-learning, entertainment and other areas have several pitfalls. This paper addresses those problems by introducing a new filtering strategy, centered on the properties that characterize the items and the users. Preliminary experiments with real users have proved that this approach outperforms previous ones in terms of consumers' satisfaction with the recommended items. The benefits are especially apparent among people with specific health concerns.

1. Introduction

Knowing the growing amount of data available through the digital media, the scientific community has been working on *recommender systems* to tailor the information delivered to the consumers' interests, preferences and needs. Typically, these systems were driven by the information captured in such *user profiles* as web navigation or TV watching records (Blanco-Fernández, Pazos-Arias, López-Nores, Gil-Solla, & Ramos-Cabrer, 2006; Tsunoda & Hoshino, 2008), learning or consumption histories (Cheung, Hui, Zhang, & Yiu, 2003; Lim et al., 2008), etc. More recently, bearing in mind that health is a major driver of decisions for many people, there have been approaches to make recommendations considering also the data stored in *electronic health records* (HER), providing additional mechanisms to ensure privacy (Fernández-Luque, Karlsen, & Vognild, 2009; Khan et al., 2007; Lacal, 2007; Pattaraintakorn, Zaverucha, & Cercone, 2007). This idea, for example, makes it possible to advertise over-the-counter (i.e. non-prescription) drugs to receptive users who may benefit from them, while avoiding offering groceries or drinks (e.g. coffee) that may interact with any prescriptions they are following. Likewise, having access to EHR repositories serves to achieve greater targeting in advertising for herbal, first-aid or dietetic products, for rehabilitation or assistance services, etc. following factual information about the users' health problems or factual/inferred information about their hobbies. Among many other uses, it is possible to avoid offering certain products to people touched by diabetes, coeliac disease or allergies in general, and even think of preparing personalized diet plans following the culinary preferences and the metabolic possibilities of each individual.

Our proposal in this paper has to do with the strategy adopted by the current recommender systems to decide which items to offer to a given user (the *target user*). We can differentiate three major approaches in literature and also in commercial use, which have significant drawbacks to be applied in general-purpose health-aware recommender systems:

- Historically, the first strategy considered for recommender systems was *content-based filtering* or *case-based filtering* (which we shall denote by CBF). This strategy consists in suggesting items similar to others that gained the target user's interest in the past (Bridge, Göker, McGinty, & Smyth, 2006), which is quite simple to implement. However, the recommendations tend to be repetitive for considering that a user will always appreciate the same kind of stuff. This *overspecialization* may not pose a problem with users who want to remain informed on specific topics (e.g. people with chronic diseases), but it does so in general.
- In response to the problem of overspecialization, researchers came up with *user-based collaborative filtering* (UBCF), to consider the success of the recommendations previously made to users with similar interests (the *neighbors* of the target user) (Pazzani, 1999). This approach solves the lack of diversity, but works poorly with users (the *gray sheep*) whose preferences or needs are dissimilar to those of the majority. This is a very important issue for health-aware recommenders, inasmuch as health conditions are always a source of uniqueness.

☆ This work was supported by the Ministerio de Educación y Ciencia (Gobierno de España) project TIN2010-20797 (partly financed with FEDER funds), and by the Consellería de Educación e Ordenación Universitaria (Xunta de Galicia) incentives file CN 2011/023 (partly financed with FEDER funds).

* Corresponding author. Tel.: +34 986813967; fax: +34 986812116.
E-mail address: mlnores@det.uvigo.es (M. López-Nores).

0957-4174/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2012.01.112

- The most recent strategy is *item-based collaborative filtering* (IBCF), which consists in recommending items related to others that the target user liked in the past, considering two items related when users who like the one tend to like the other as well (Sarwar, Karypis, & Konstan, 2001). This approach still faces several problems that were also apparent with UBCF. One of those problems is *sparsity*, implying that when the number of items available to recommend is high (as it happens in many domains of recommender systems application nowadays), it is difficult to find users with similar valuations for common subsets. Another important drawback is that of *latency*, related to the inability to recommend recently added items, as long as there are no user ratings available for them.

In this paper, we present a new strategy, called *property-based collaborative filtering* (PBCF), as a means to fight the aforementioned problems in general settings, but especially in the realm of health-aware recommender systems. PBCF depends on having a semantic characterization of the items that may be recommended, which is not necessarily true for CBF, UBCF or IBCF.

The paper is organized as follows. In Section 2, we briefly survey the state-of-the-art in semantics-based recommender systems. Next, we describe the core aspects of PBCF in Sections 3 and 4. Section 5 provides details about how we have implemented PBCF in a health-aware recommender for online shopping. This serves to discuss gains achieved in terms of users' satisfaction in Section 6. Finally, we give a summary of conclusions in Section 7.

2. Background in semantics-based recommender systems

Regardless of the filtering strategy, it is noticeable that most of the recommender systems have relied on heuristics or *syntactic matching* techniques, which relate items by looking for common words in their attached metadata. These techniques —employed, for example, in the personalized advertising systems of Kastidou and Cohen (2006), Lekakos and Giaglis (2004), Thawani et al. (2004)— overlook a lot of knowledge during the personalization processes, because they are unable to reason about the meaning of the metadata (for example, it is not possible to link items including tags like "*golden retriever*" with items including tags like "*boxer*" in any way, because the two words are dissimilar). A syntactic approach is also a source of overspecialization in itself, because the recommendations so computed can only include items very similar to those the users already know.

To go one step beyond in personalization quality and diversity, research is now focused on applying techniques from the *semantic web*, which enable automatic reasoning processes that gain insight into the meaning of words —for example, it is possible to automatically recognize "*golden retriever*" and "*boxer*" as two different breeds of dogs, the latter having nothing to do with a combat sport. The key lies within the use of *ontologies*, which provide vocabularies of terms and relations to represent knowledge in terms of hierarchies of classes, instances of those classes and interrelationships among the instances (Staab & Studer, 2003). Semantic technologies were first thought of as a means to improve the users' navigation through the growing amount of information available on the Internet, but it has been applied in many other areas over the last few years, from publishing, entertainment and government to financial services and life sciences. In line with the purposes of this paper, as noted in Martínez-Costa, Menárguez-Tortosa, Maldonado, and Fernández-Breis (2010), one of the fastest-growing areas has to do with the automatic management of health-related information.

As regards research on recommender systems, many authors have enhanced the traditional filtering strategies with semantic reasoning mechanisms, in order to discover the items that best match the preferences of each user by reasoning about their semantic descriptions. In Hung (2005), proposed a recommender system for one-to-one marketing based on a taxonomy of products, providing evidence of the advantages of such semantics when it came to providing instant online recommendations and identifying potential customers upon release of a new product. In Middleton, Shadbolt, and de Roure (2004), explored a novel ontological approach to user profiling within semantics-based recommender systems, coping with the problem of recommending academic research papers; the experiments showed that profile visualization and feedback outperformed previous user modelling approaches, which led the authors to conclude that the semantics captured by their ontological approach made the profiles easier to understand. Yuan and Cheng (2004) investigated analogy structures between heterogeneous products (i.e., products with different properties) to recommend items that are disparate from others the users had purchased before, using what they called an *ontology-driven coupled clustering* algorithm.

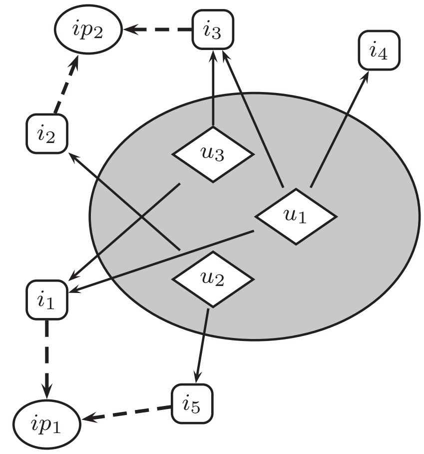

Fig. 1. Conceptual representation of UBCF. (u: user; *i*: item; *ip*: item property).

We have explored ourselves the benefits of semantics-based recommender systems in different domains of application. In Blanco-Fernández et al. (2006), we proposed an ontology-driven recommendation system to select the most appealing TV programs for the users. In Pazos-Arias et al. (2008), we incorporated a similar semantics-enhanced approach into a t-learning platform to recommend personalized educational courses according to the users' preferences and previous knowledge. Later on, we developed MiSPOT (López-Nores, Pazos-Arias, García-Duque, & Blanco-Fernández, 2010) to deliver personalized advertisements for online shopping during TV watching. This system was refined to yield a prototype of HARE (*health-aware recommender*), which could access information stored in electronic health records as a means to improve the quality of the recommendations (López-Nores, Blanco-Fernández, & Pazos-Arias, 2010). Finally, in Blanco-Fernández, López-Nores, Pazos-Arias, Gil-Solla, and Ramos-Cabrer (2010), we exploited the benefits of semantics-driven reasoning in a tourism recommender system.

3. Motivations of PBCF

As noted in Section 1, UBCF is driven by the definition of clusters of users (*neighborhoods*) as per the items they have rated positively or negatively. This approach is depicted in Fig. 1, where solid arrows denote a positive rating given by a user to an item, and dashed arrows link items to the properties or features that characterize them.[1] On the other hand, as depicted in Fig. 2, IBCF

[1] We shall be using the terms *feature* and *property* interchangeably.

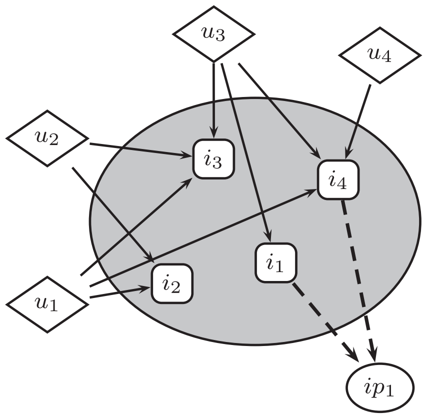

Fig. 2. Conceptual representation of IBCF. (*u*: user; *i*: item; *ip*: item property).

can be seen as clustering together items that have been successfully or unsuccessfully recommended to certain users.

Most of the times, the two collaborative filtering strategies rely on direct links between specific users and specific items, meaning, for example, that "*Jane has liked ProductA*" or "*John has disliked ProductB*". Thus, users u_1 and u_3 are put in the same neighborhood in Fig. 1 (UBCF) because they have given similar ratings to items i_1 and i_3. Likewise, items i_2 and i_3 in Fig. 2 (IBCF) appear together because they have been appealing to two users, namely u_1 and u_2. In some cases, however, the relationships that define the clusters may be identified also through the item properties. In Fig. 1, for example, users u_2 and u_3 are found to have similar likings because there are common properties among items they have rated positively (i_2 and i_3 share the property ip_2, whereas i_1 and i_5 share the property ip_5). This way, having the ability to manage item properties, an UBCF-based system like that of Blanco-Fernández et al. (2010) can treat two users as neighbors if they are fond viewers of nature documentaries, even if they have watched different ones on different TV channels. Similarly, in IBCF-based systems like those surveyed in Sarwar et al. (2001), it is possible to detect that people who purchase tennis rackets tend to purchase sportswear too, regardless of specific brands. Anyway, it is common to all forms of UBCF and IBCF that the links exploited to match items and users are characterized *statically* by the ratings that the latter have given to the former.

Our proposal of PBCF introduces a new level in the reasoning of the recommender systems, by maintaining links between the semantic properties that characterize both users and items. Specifically, the aim of PBCF is to capture knowledge that may be put down in rules like "*diabetic users tend not to like candy, unless they are sugar-free*" or "*pregnant women usually purchase baby clothes*". To this aim, as shown in Fig. 3, we end up defining clusters of items as per the likelihood that they will be appealing to users who match certain features.

The ability of PBCF to reason about item properties and user properties depends on having a formal, explicit specification and representation of domain knowledge. It is here that an ontology and other concepts from the Semantic Web come into play, as it will be seen next.

4. Internal computations of PBCF

Internally, PBCF maintains one matrix with one row per property that may characterize a user, and one column per property that may characterize an item. For the items, in principle, we can merely handle the semantic properties and attributes defined in

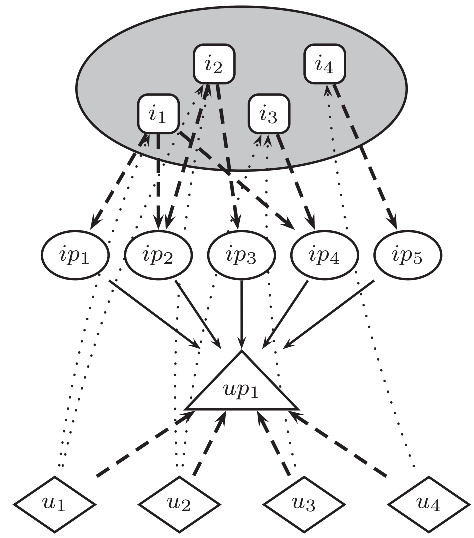

Fig. 3. Conceptual representation of PBCF. (*u*: user; *i*: item; *ip*: item property; *up*: user property).

the ontology. As regards the users, on the contrary, we can manage various types of data:

- Demographic features like age, gender or marital status.
- Conditions bound to consumption patterns, such as what can be inferred from the classical sources of information about the users—for example, web navigation histories may suggest specific topics of interest, whereas TV watching records may provide information about preferred genres of TV programs and movies.
- Conditions not bound (at least, directly) to consumption patterns, such as those retrieved from electronic health records, related to current or past diseases or to medical prescriptions in force.
- The representativeness of certain stereotypes, which can be easily related to the properties and attributes modeled in the ontology. For example, by applying classical metrics of CBF-based systems (see Bridge et al., 2006) it is easy to classify one user within the stereotypes of "*people who like traveling*" or "*do-it-yourself fans*".

The value of cell (*i*,*j*) of the PBCF matrix is an indication of how good an item characterized by the item property ip_j should be for a user characterized by the user property up_i. Fig. 4 shows that we keep the values of the matrix in the range [−1,1] (with −1 representing the most negative effect and 1 representing the opposite), but this is just for convenience in our own implementation of PBCF (details will follow in Section 5).

In contrast with the static links handled by UBCF and IBCF (remember Section 3), the numbers in the PBCF matrix evolve over time, driven by the relevance feedback gathered after recommending items. Intuitively, the facts that a user characterized by

	ip_1	ip_2	ip_3	$\dots$	ip_n
up_1	+0.8	−0.3	−0.1	$\dots$	−0.3
up_2	−0.1	−0.9	+0.4	$\dots$	−0.8
up_3	−0.2	0	−0.1	$\dots$	+0.9
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
up_n	+0.4	+0.2	−0.3	$\dots$	+0.7

Fig. 4. A sample PBCF matrix: one row per user property, one column per item property and no trace of data that may serve to identify any specific user.

properties up_2 and up_3 has given a very positive rating to an item characterized by properties ip_1 and ip_n contributes to increasing the values of cells $(2,1)$, $(2,n)$, $(3,1)$ and $(3,n)$—obviously, negative ratings work in the opposite way. Interestingly, given the right service-level agreements, this dynamic updating can be done in a totally anonymous manner, so a PBCF-based recommender system can be seen as computing an aggregate that does not expose individual users' data.

Knowing the list of properties that characterize a given item, we can traverse the rows of the PBCF matrix to compute the item's level of appeal for the different user properties. This way, we obtain what we call the appeal vector of the item. In this process, it is mandatory to apply some ponderation in order to grant greater weight to the features with the highest and the lowest values in the row, because they are most informative about what is good and bad for every single user property. Weighings derived from classical metrics of information theory (like *entropy* or *amount of information* as defined in Shannon & Weaver (1998)) can work, whereas simple averages (e.g. arithmetic, geometric or harmonic mean) would not do. For example, with an arithmetic mean, item features that conflict with the users' preferences and needs (e.g. sugar-richness in the case of a diabetic person) would dilute as negative contributions to summations of many addends, so unsuitable items could end up appearing in the recommendations anyway.

The left part of Fig. 5, for example, shows the appeal vector computed from the matrix of Fig. 4 for an item *IT* characterized by properties ip_1, ip_2 and ip_3, with relative weights given by exponentiations of the form $|matrix(i,j)|^3$. In Fig. 4, it can be seen that property ip_1 is quite good (+0.8) for users characterized by up_1, whereas ip_2 is quite bad (-0.9) for users characterized by up_2 (the other properties of *IT*, with values between -0.3 and +0.4, have effects closer to neutral). The relative importance of ip_1 and ip_2 is evident in *IT*'s appeal vector due to the values +0.743 and -0.794 corresponding to up_1 and up_2, respectively On the contrary, a vector computed using an arithmetic mean (shown on the right hand side of Fig. 5) would not make it clear that *IT* should be much more suitable for users characterized by up_1 than for users characterized by, say, up_n.

At this point, it is worth noting that, by relying only on properties, PBCF can compute appeal vectors for newly-added items even if there are no user ratings available for them: it suffices to have values in the PBCF matrix for items that share properties with the new one. This way, we avoid the sparsity and latency problems of the other collaborative strategies.

Having obtained the appeal vectors, we can discretize their values to cluster the items as per the likelihood that they will be appealing to users who match certain features. Typically, as shown in Fig. 6, the clusters obtained for any given user property up_i yield a diamond shape, with a comparatively large number of items placed in intermediate levels, meaning that up_i is not a decisive feature to take into account when it comes to deciding about recommending one of those items to a given user.

The discretization and clustering processes are undoubtedly useful for market studies (which are beyond the scope of this paper), but they are not needed to make recommendations. Intuitively, the filtering strategy with PBCF consists in identifying the items that best fit with the features listed in the profile of the target user, considering the 4 types of user properties mentioned at the beginning of this section. To this aim, we can simply compute the level of appeal of each item for the target user by traversing their appeal vectors, again with some ponderation to give more importance to the features with the highest and lowest values.

up_1	$+0.743$	$+0.133$
up_2	-0.794	-0.2
up_3	-0.177	-0.033
$\vdots$	$\vdots$	$\vdots$
up_n	$+0.193$	$+0.1$

Fig. 5. On the left, the appeal vector computed from the matrix of Fig. 4 for an item *IT* characterized by ip_1, ip_2 and ip_3. On the right, the vector that would result from using an arithmetic mean.

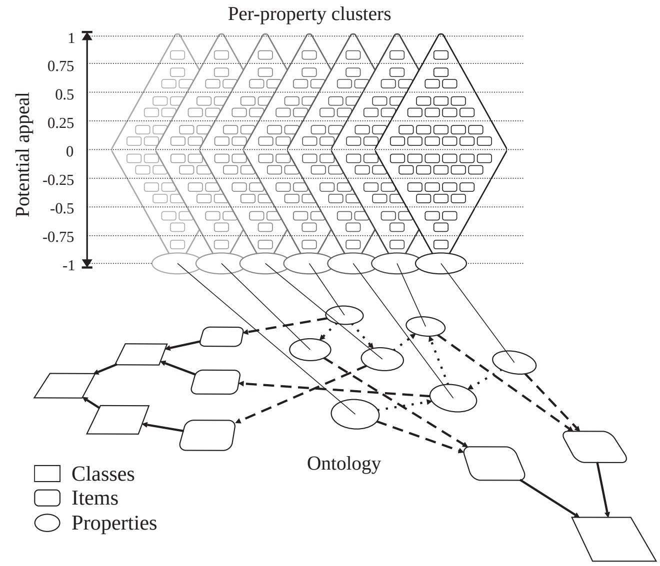

Fig. 6. Clustering items as per user properties.

Going on with the example of Fig. 5, using ponderations of the form $|vector(i)|^3$, we would compute the value -0.102 as the level of appeal of the item *IT* for a user characterized by properties up_1, up_2 and up_3. Thus, the item would not be recommended to one such user, because the very positive effect of the item property ip_1 is neutralized by the very negative effect of ip_2. On the contrary, *IT* would be offered to a user characterized by up_1 and up_3 only, thanks to the very positive effect of ip_1 and the neutral effects of the other *IT* properties for this target. The item would not be recommended if we applied simple averages; for instance, using the vector on the right hand side of Fig. 5, we would have got an appeal value of +0.1 for *IT* (resulting from $0.133 - 0.033$), which is too low.

It is important to note that, with PBCF, there is no unfair treatment to gray sheep (i.e. to users whose preferences and needs are dissimilar to those of the majority) because we can proceed with each one of the users' features *separately*, no matter how uncommon their combination might be. Besides, the ability to identify the properties with greater weights in the computations makes it possible to explain the reasons why the items are recommended. In other words, PBCF can reduce dubiousness/distrust among users with very specific needs by indicating the most decisive features when presenting the recommendations (e.g. emphasizing the words "*popular among diabetics*" or "*gluten free*").

Finally, as explained in Takács, Pilászy, Németh, and Tikk (2009), implementations of UBCF and IBCF do not scale well with the number of users and the number of items available to a recommender system. On the contrary, the computational cost of PBCF is related to the numbers of user properties and item properties considered, which are to a great extent independent of the numbers of users and items, respectively. It is also worth noting that the computation of appeal vectors can be done off-line at any time, separately from the elaboration of recommendations. From our implementation of PBCF (described in Section 5) and our previous experiences with other recommendation strategies, we can confirm that PBCF is indeed more scalable than the other forms of collaborative filtering.

Fig. 7. A snapshot of a recommendation made by the HARE system.

5. Specifics of the HARE system

While the idea of PBCF should be valid for any domain of recommender systems application, we have tried it in the realm of health-aware recommenders. Specifically, we have come up with a revision of HARE, a system introduced in López-Nores et al. (2010) to deliver personalized advertisements during TV watching, through a combination of CBF and UBCF. The previous version of HARE was in turn an extension of the health-unaware recommender MiSPOT (López-Nores et al., 2010), also based on a combination of filtering strategies. The receiver-side parts of these systems were written using the Java APIs provided by the *Multimedia Home Platform* (MHP) standard,[2] an open middleware system that enables the distribution and execution of interactive applications on a TV set.

Fig. 7 shows one snapshot of the HARE system, displaying one advertisement of herbal remedies in the context of a nature documentary, with a note on the upper right corner of the screen due to the fact that many people take garlic derivatives to control their cholesterol levels.

HARE works with metadata from several specifications, including the following:

- TV-Anytime[3] to capture demographic features of individual consumers or groups.
- eCl@ss[4] to categorize classes of commercial products and services as per features like purpose or brand.
- A subset of the *Disease Ontology*[5] to enable reasoning about health conditions, diseases, symptoms and signs, to be matched against information retrieved from electronic health records compliant with the EN13606 standard (CEN, 2008).

We built an ontology following the methodology of Hepp (2006), in such a way that the advertisements that may be recommended inherit and extend the characterization of the corresponding items. As regards the consumer profiles, we supplement the aforementioned demographic features with properties retrieved from EN13606-compliant HER repositories, and also with properties (stereotypes) bound to the ontology as per the consumers' consumption of genres of TV programs or classes of commercial products. The belonging of one consumer to one stereotype is decided by the classical criterion of Pearson-*r* correlation as explained in López-Nores et al. (2010).

The HARE system manages consumer ratings in the range $[-1, 1]$, with -1 representing the greatest disliking and +1 representing the greatest liking. As noted in Section 4, the values in the PBCF matrix are kept between -1 and +1, too. This makes it possible to differentiate important from unimportant features by means of simple exponentiations of the form $|matrix(i,j)|^n$ or $|vector(i)|^n$. After a first round of experiments, we found that taking $n = 3$ served to differentiate relevant from irrelevant features as desired/expected, even though we leave it for future work to try other formulae and find which one works best.

In HARE, just like in the original MiSPOT system (see López-Nores et al., 2010), the ratings may be entered explicitly by the consumers, or inferred from whether they browse the advertisements inserted in the TV programs, whether they launch interactive commercial applications thereafter, how long they take to learn about the items, whether they decide to purchase online, etc. The novelty regarding PBCF is that we use this feedback to update the values in the matrix of user and item properties, with weights determined by (i) the ratings previously given by each consumer to items with the item properties in question, and (ii) the number of ratings received before for any given cell. The updates propagate upwards in the hierarchies defined in the ontology, so that, for example, positive ratings given to "*nappies*" help increase the number corresponding to the class "*baby stuff*" in general. In relation to this, the number of updates made to any cell in the matrix can be seen as an indication of how consolidated its value is, which we believe is worth taking into account during the computation of appeal vectors. If we find that the number corresponding to an item property is too preliminary, we look for more consolidated values among its superclasses. In order to consider one number consolidated, we require a number of updates that (i) grows with the number of ratings received overall, and (ii) decreases with the depth of the property in the hierarchies.

6. Evaluation

We have conducted preliminary experiments to assess the personalization quality achieved by the PBCF-based version of the HARE recommender system, in comparison with the original version introduced in López-Nores et al. (2010) (based on a combination of CBF and UBCF) and the MiSPOT system (López-Nores et al., 2010) (also based on CBF and UBCF, but with no refinements derived from accessing health-related data). To this aim, we recruited 112 users among our graduate/undergraduate students and their relatives or friends. They made up a diverse audience, with disparate demographic data and educational backgrounds; there were nearly as many men as women (57% vs 43%), with ages ranging from 16 to 67 yr old.

The experiments consisted in delivering personalized advertisements during TV shows over a period of 4 weeks, in successive 1-h sessions until every user had received at least 30 suggestions from each one of the three recommendation engines. None of the systems started from scratch, since we could rely on an extensive collection of ratings from more than 150 consumer profiles that had built up during previous experiments with MiSPOT (see López-Nores et al., 2010) and other previous systems (see Blanco-Fernández et al., 2006, 2010; Pazos-Arias et al., 2008).

Prior to making any recommendations, we defined a set of 15 stereotypes by clustering the profiles from the previous experiments and averaging the numbers they contained. Specifically, 14 clusters contained the profiles that had comparatively high (close to +1) or comparatively low (close to -1) ratings for items

[2] http://www.mhp.org/.
[3] http://www.tv-anytime.org/.
[4] http://www.eclass-online.com/.
[5] http://diseaseontology.sourceforge.net/.

Fig. 8. A snapshot of the interfaces provided for the users to enter ratings during our experiments.

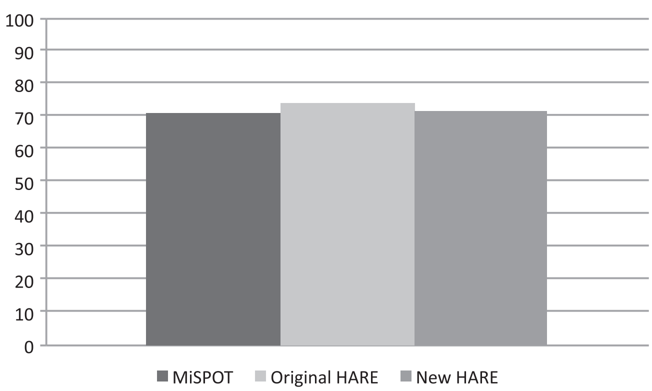

Fig. 9. Overall precision of the three recommendation engines.

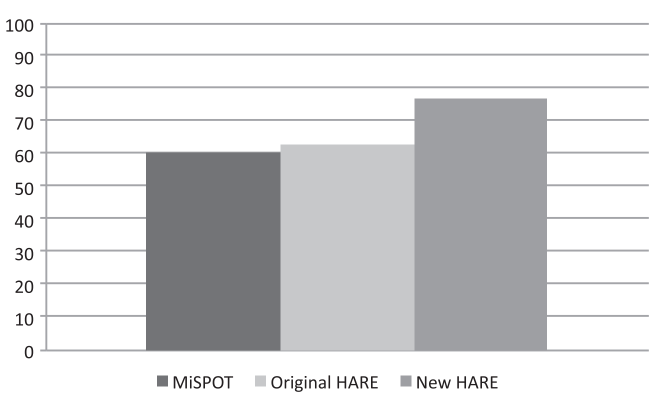

Fig. 10. Precision among users with specific health-related concerns.

classified under *Sports*, *Nature*, *Technology*, *Science*, *Health*, *Culture* or *Traveling*, whereas the 15th cluster gathered the profiles that did not meet any of those conditions. Having done this, we asked each user to rate his/her interest in the aforementioned topics between 0 and 10, and their individual profiles were then initialized by weighing the numbers of the corresponding stereotypes. In the PBCF-based version of HARE, we treated the belonging to the stereotypes as user properties.

The users had to rate the advertised items at the end of each session, again with numbers between 0 and 10 that we would translate to the range $[-1, 1]$. The interfaces provided to this aim looked as shown in Fig. 8.

Having gathered a sufficient number of user ratings, as usual in the literature of recommender systems, we could measure the precision of the recommendations as the percentage of items that had been rated greater than +0.5. As a result, we obtained the chart of Fig. 9.

In principle, we found that the overall quality of the recommendations was practically the same with the three engines. Briefly, 71% of the recommendations provided by MiSPOT received ratings between +0.5 and +1, whereas the first version of HARE achieved 74% and the new one got 72%. Nonetheless, these small differences —which should nevertheless be ratified with measures from many more users over a longer period of time— turn into big ones if we look only at the recommendations made to users with specific health concerns. There were 42 such users in our experiments, with low-medium degrees of seriousness. Considering only their cases, as shown in Fig. 10, we got 60% precision for MiSPOT, 63% for the first version of HARE, and 77% for the new HARE. Undoubtedly, these results evidence the benefits of PBCF for users who would be treated as gray sheep with other approaches.

We have checked the significance of the aforementioned differences through ANOVA tests (Sauer, 2010), which are one popular validation tool in the literature of recommender systems. To this aim, we compared the F statistic value in Fig. 11 against a critical value, which must be queried in predefined tables considering a specific significance value ($\alpha = 0.01$ in our tests) and the degrees of freedom of ANOVA tests: 2 (from the fact that we are comparing 3 methods) and 39 (from the 42 users we are considering).

The resulting F value is 46.86, which is much greater than the tabulated value $F(0.01, 2, 39) = 5.194$. Therefore, we can reject the hypothesis of equal population means and conclude that the precision of the recommendations varies with the filtering strategy. Besides, the p value corresponding to the data in Table 11 is $5.87 \cdot 10^{-5}$ (<0.01), so the test statistic is significant at that level.

In order to analyze the trend of recommendation precision values, we made contrasts and post hoc tests (specifically, Tukey's and Bonferroni's tests using SPSS) to compare mean values, and the results confirmed that PBCF outperforms both the first version of HARE and MiSPOT. Finally, running a poll among the users with specific health concerns, we found that displaying the health-related properties that motivate the recommendations (for example, with a note on screen saying "*popular among people with skin problems*") made them consider items they would disregard otherwise in nearly 20% of the cases. Also, we estimated that users who received PBCF recommendations felt 25% more comfortable with the publicity, reducing its perception as "*a nuisance*" or "*a necessary*

Source of variation	Sum of squares	Degrees of freedom	Mean squares	F	p
Treatments	4787.24	2	2393.62	46.86	$5.87 \cdot 10^{-5}$
Error	1989.31	39	51.078		
Total	6776.55	41			

Fig. 11. ANOVA table for recommendation precision.

evil". This is something that could hardly be done with CBF, UBCF or IBCF.

7. Conclusion

We have presented property-based collaborative filtering as a new filtering strategy for recommender systems, based on measuring the likelihood that the available items will be appealing to users who match certain features. Whereas other strategies rely on direct links between specific users and items, PBCF fully decouples users and their properties on the one hand, and items and their properties on the other. This way, it is possible to build a matrix of values representing how much one item feature influences (positively or negatively) the suitability of an item for someone with a certain user property, which helps solve persistent problems of other collaborative approaches like sparsity, latency and the unfair treatment given to people whose interests and needs are different from those of the majority. Our preliminary experiments have evidenced that PBCF can be particularly advantageous to reckon the distinguishing and decision-driving nature of health-related data.

Nowadays, we are working to use the PBCF paradigm as the basis to improve the consideration of time in recommender systems. Current recommenders can adapt the selection of items as the preferences of the users evolve over time, but the adaptation process always takes it for granted that a consumer's interest for a given type of product (or any of its features) decreases with time from the moment of the last purchase, even though certain products may indeed become more interesting or necessary. For example, if a consumer has just bought a washing machine, it is foreseeable that he/she will not need another one until the average lifetime of such appliances has passed, so any recommender system should prioritize other products for some time. Likewise, the interest for seasonal clothes may vary along the year, while the interest in books and music may remain constant and school equipment may have a peak at the beginning of the academic year. Some authors have addressed this issue by attaching temporal information to the items' metadata (see Lee, Park, & Park, 2009), but missing the point that the influence of time can be very different for different consumers. Due to its treatment of user properties, we are working on PBCF to develop a better management of time, using the relevance feedback to compute functions with which to revise parameterizable time functions linked to classes and properties in the ontology of items. In parallel, we aim to check whether the same PBCF algorithms could work in a social tagging environment, managing the tags in a folksonomy instead of the properties defined in an ontology.

References

Blanco-Fernández, Y., López-Nores, M., Pazos-Arias, J. J., Gil-Solla, A., & Ramos-Cabrer, M. (2010). Exploiting digital TV users' preferences in a tourism recommender system based on semantic reasoning. *IEEE Transactions on Consumer Electronics, 56*(2), 904–912.

Blanco-Fernández, Y., Pazos-Arias, J. J., López-Nores, M., Gil-Solla, A., & Ramos-Cabrer, M. (2006). AVATAR: An improved solution for personalized TV based on semantic inference. *IEEE Transactions on Consumer Electronics, 52*(1), 223–231.

Bridge, D., Göker, M. H., McGinty, L., & Smyth, B. (2006). Case-based recommender systems. *The Knowledge Engineering Review, 20*(3), 315–320.

CEN. (2008). Health informatics – Electronic health record communication – Part 1: Reference Model, International Standard ISO 13606-1.

Cheung, B., Hui, L., Zhang, J., & Yiu, S. M. (2003). SmartTutor: An intelligent tutoring system in web-based adult education. *Journal of Systems and Software, 68*(1), 11–25.

Fernández-Luque, L., Karlsen, R., & Vognild, L. K. (2009). Challenges and opportunities of using recommender systems for personalized health education. *Studies in Health Technology and Informatics, 150*, 903–907.

Hepp, M. (2006). Products and services ontologies: A methodology for deriving OWL ontologies from industrial categorization standards. *International Journal on Semantic Web and Information Systems, 2*(1), 72–99.

Hung, L. (2005). A personalized recommendation system based on product taxonomy for one-to-one marketing online. *Expert Systems with Applications, 29*, 383–392.

Kastidou G., Cohen R. (2006). An approach for delivering personalized ads in interactive TV customized to both users and advertisers, In*Proceedings of the 4th european conference on interactive television*, Athens, Greece.

Khan, S. A., McFarlane, D. J., Li, J., Ancker, J. S., Hutchinson, C., Cohall, A., Kukafka R. (2007). Healthy harlem: Empowering health consumers through social networking, tailoring and web 2.0 technologies, Chicago, USA, October.

Lacal J. C., (Eds). In*Proceedings of the 1st workshop of the IEEEP2407 personalized health informatics working group*, January 2007. <http://www.IEEE2407.org>.

Lee, T. Q., Park, Y., & Park, Y. (2009). An empirical study on effectiveness of temporal information as implicit ratings. *Expert Systems with Applications, 32*(2), 1315–1321.

Lekakos, G., & Giaglis, G. M. (2004). A lifestyle-based approach for delivering personalised advertisements in digital interactive television. *Journal of Computer-Mediated Communications, 9*(2).

Lim, J., Kim, M., Lee, B., Kim, M., Lee, H., & Lee, H.-K. (2008). A target advertisement system based on TV viewer's profile reasoning. *Multimedia Tools and Applications, 36*(1-2), 11–35.

López-Nores, M., Blanco-Fernández, Y., & Pazos-Arias, J. J. (2010). Electronic health records as a new dimension of user profiles in recommender systems. *Global Telemedicine and eHealth Updates, 3*, 137–140.

López-Nores, M., Pazos-Arias, J. J., García-Duque, J., & Blanco-Fernández, Y. (2010). MiSPOT: Dynamic product placement for digital TV through MPEG-4 processing and semantic reasoning. *Knowledge and Information Systems, 22*(1), 101–128.

Martínez-Costa, C., Menárguez-Tortosa, M., Maldonado, J. A., Fernández-Breis, J. T. (2010). Semantic Web, chapter Semantic Web technologies for managing EHR-related clinical knowledge, InTech.

Middleton, S. E., Shadbolt, N. R., & de Roure, D. C. (2004). Ontological user profiling in recommender systems. *ACM Transactions on Information Systems, 22*(1), 54–88.

Pattaraintakorn, P., Zaverucha, G. M., & Cercone, N. (2007). Web-based health recommender system using rough sets, survival analysis and rule-based expert systems. *Lecture Notes in Artificial Intelligence, 4482*, 491–499.

Pazos-Arias, J. J., López-Nores, M., García-Duque, J., Díaz-Redondo, R. P., Blanco-Fernández, Y., Ramos-Cabrer, M., et al. (2008). Provision of distance learning services over interactive digital TV with MHP. *Computers and Education, 50*(3), 927–949.

Pazzani, M. (1999). A framework for collaborative content-based, and demographic filtering. *Artificial Intelligence Review, 13*(5), 393–408.

Sarwar, B., Karypis, G., Konstan, J., Riedl J. (2001). Item-based collaborative filtering recommendation algorithms, In *Proceedings of the 10th international conference on the World Wide Web* (pp. 285–295). Hong Kong, China.

Sauer, M. (2010). The basics of analysis of variance (ANOVA). *Amazon Digital Services*.

Shannon, C. E., & Weaver, W. (1998). *The Mathematical Theory of Communication*. University of Illinois Press.

Staab, S., & Studer, R. (Eds.). (2003). *Handbook on ontologies*. Springer.

Takács, G., Pilászy, I., Németh, B., & Tikk, D. (2009). Scalable collaborative filtering approaches for large recommender systems. *Journal of Machine Learning Research, 10*, 623–656.

Thawani, A., Gopalan, S., Sridhar V. (2004). Context-aware personalized ad insertion in an Interactive TV environment, In *Proceedings of the 4th workshop on personalization in future TV*, Eindhoven, The Netherlands, August.

Tsunoda, T., & Hoshino, M. (2008). Automatic metadata expansion and indirect collaborative filtering for TV program recommendation system. *Multimedia Tools and Applications, 36*(1–2), 37–54.

Yuan, S. T., & Cheng, C. (2004). Ontology-based personalized couple clustering for heterogeneous product recommendation in mobile marketing. *Expert Systems with Applications, 26*(1), 461–476.