Group Recommendations Based on Hesitant Fuzzy Sets

Jorge Castro,[1,2,†] Manuel J. Barranco,[3,‡] Rosa M. Rodríguez,[1,§]
Luis Martínez[3,*]
[1]*Department of Computer Science and Artificial Intelligence, University of Granada, E-18071 Granada, Spain*
[2]*Centre for Quantum Computation & Intelligent Systems (QCIS), University of Technology Sydney, Ultimo NSW 2007, Australia*
[3]*Department of Computer Science, University of Jaén, s/n 23071 Jaén, Spain*

Group recommender systems (GRSs) recommend items that are used by groups of people because certain activities, such as listening to music, watching a movie, dining in a restaurant, etc., are social events performed by groups of people sharing their tastes, and their choices affect all of them. GRSs help groups of people making choices in overloaded search spaces according to all group members preferences. A common GRS scheme aggregates users preferences to generate a group preference profile. However, the aggregation process may imply a loss of information, negatively affecting different properties of the GRS such as *diversity* of group recommendations, which is an important quality factor because of such recommendations are targeted to groups formed by users with individual and possibly conflicting preferences. To avoid and manage the loss of information caused by aggregation, this paper proposes to keep all group members preferences by using hesitant fuzzy sets (HFSs) and interpreting such information like the group hesitation about their preferences that will be used in the group recommendation process. To evaluate the performance and rank quality of the HFS GRS proposal, a case study is carried out. © 2017 Wiley Periodicals, Inc.

1. INTRODUCTION

Frequently, people face situations where they have to choose among a large number of options: books, movies, restaurants, vacations, TV shows, etc. Recommender systems (RSs)[1–3] assist people in these situations, in which the large amount of information makes it hard to find out the relevant items according to their tastes or necessities, by filtering useless items. In the literature, there are different types of

*Author to whom all correspondence should be addressed; e-mail: martin@ujaen.es.
†email: jcastro@decsai.ugr.es.
‡email: barranco@ujaen.es.
§email: rosam.rodriguez@decsai.ugr.es.

View this article online at wileyonlinelibrary.com. • DOI 10.1002/int.21922

RSs, being the most popular ones: (i) collaborative filtering (CF) RS,[4–7] used in this research, (ii) content-based RS,[8–11] and (iii) hybrid RS.[12,13] The main differences among these approaches are the sources of information that they rely on to provide the recommendations.

The performance of RSs has been classically evaluated by *accuracy* metrics,[2] but recently other features are getting more importance in the evaluation of RS, such as *diversity*,[14] which measures how dissimilar the items recommended by the system are to the target user. The improvement of the diversity can lead to lower accuracy,[15] hence a trade-off should be considered between them. Therefore, a RS should balance both features to provide quality recommendations.

RSs focus on recommending items to individuals. However, there are items with social features that are consumed by groups, e.g., some friends who want to eat at a restaurant, to choose a movie to watch, to select the holidays destination, etc. In these cases, the recommendations should satisfy not only to a unique individual but also to the whole group. To deal with issues introduced by group recommendations, group recommender systems (GRSs) propose solutions that have been widely explored in the literature.[16–18]

A traditional GRS is based on classic RSs oriented to individuals, which are extended by means of aggregation mechanisms to address a group of users instead of a unique user. Unfortunately, aggregation processes may imply loss of useful information such as the distribution, diversity, or shape of the initial data. This loss of information can either bias or lead to wrong results. Therefore, an important challenge for GRSs would be to keep the maximum information about the group during the fusion of information for providing better recommendations. It is then necessary to find out and develop the right preference modeling and information fusion tools to achieve this goal in GRSs.

Hesitant fuzzy sets (HFSs), were introduced by Torra[19] as an extension of fuzzy sets[20] in which, given a reference set, the membership function does not provide only one value but a set of them, which provides a way of modeling hesitation. Hence, we can interpret the different users' preferences of a group, for a given item, as the group hesitation about the preference of such an item.

Our proposal will consist of a hesitant group recommender model (HGRM), based on CF and HFS, which is able to recommend to a group of users, while it keeps all information avoiding the aggregation process. This proposal will be evaluated and compared with traditional GRSs by a case study on a widely used data set for RSs and discussing then the results obtained from different points of view, such as accuracy and diversity. The results show several improvements of our proposal compared to the baseline.

This paper is structured as follows. Section 2 provides a background of GRSs and HFSs. Section 3 introduces in further detail our proposal of a HGRM. Section 4 shows a case study performed to evaluate the proposal and discuss the findings. Finally, Section 5 concludes the paper.

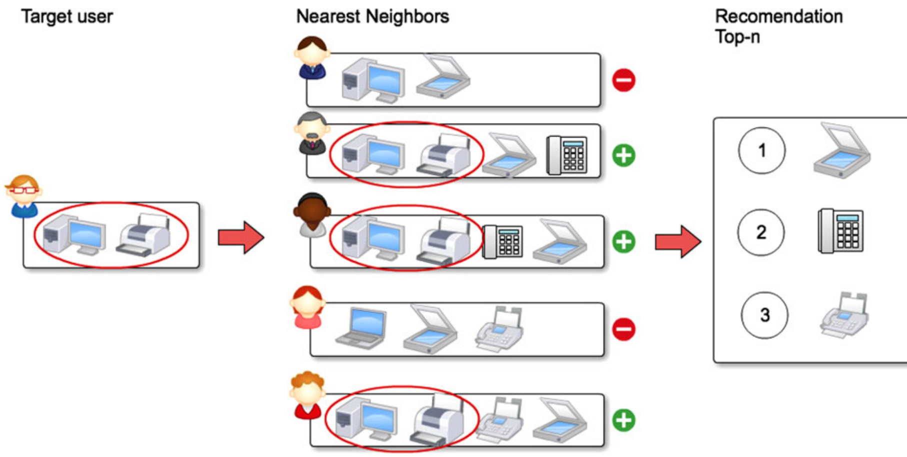

Figure 1. Collaborative filtering.

2. PRELIMINARIES

This section reviews the main concepts regarding HFSs and GRSs, which are necessary to understand our proposal.

2.1. Group Recommender Systems

A GRS works with a set of users, $U = \{u_1, \ldots, u_q\}$, a set of items, $I = \{i_1, \ldots, i_m\}$, and a set of ratings, $R = \{r_{11}, \ldots, r_{ij}, \ldots, r_{qm}\}$, where r_{ij} is the rating of the user u_i for the item i_j, expressed in the domain interval $D = [d_{min}, \ldots, d_{max}]$. Besides, there is a group of users, $G = \{u_1, \ldots, u_n\} \subset U$, $n << q$, which is the target group; without loss of generality, we will consider that G are the first n users of U.

Typically, a GRS is based on traditional RS approaches, but GRSs can cope with additional issues that do not appear on recommendation targeted to individuals. Some authors present new challenges that a GRS should cope with.[16,17] As aforementioned, our proposal is focused on the traditional CF RS,[6] which recommends to the target user the items that others with similar tastes liked in the past. User profiles, obtained from rating history, are compared to measure the similarity among users. Given a target user, the strategy is to find his/her *nearest neighbors* and combine their profiles to discover and rank the most suitable items (see Figure 1).

The most widely used approach for GRSs is a CF GRS based on a CF RS in which an aggregation process is added.[21,22] Regarding when the aggregation process is applied, there are different approaches for CF GRS[16] (see Figure 2):

(a) *Aggregating individual preferences*: The users' preferences, stated by their ratings, $\{r_{ij}\}$, are aggregated to obtain the group's preferences. The rest of steps are the same as for a CF RS (see Figure 2a)).

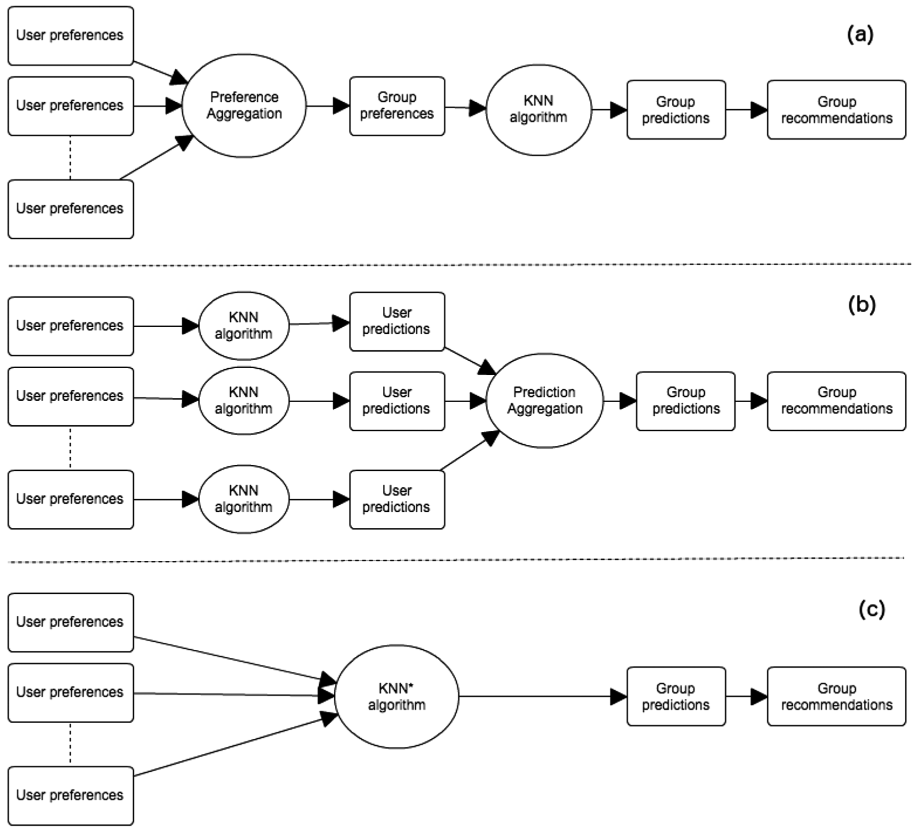

Figure 2. Three approaches for CF GRS.

(b) *Aggregating individual predictions*: The similarity measure is applied individually to each user, u_i, using an algorithm to find the K nearest neighbors (KNN), obtaining individual user predictions. These ones are aggregated to obtain a group recommendation. The process finalizes like a CF RS (see Figure 2b).

Our aim is to develop a GRS approach without an aggregation process following the scheme shown in Figure 2c.

One of the most popular approaches for GRSs aggregates the individual preferences as shown in Figure 2a, known as a *pseudo-user* approach. Once a group profile represents a summary of users profiles, any traditional recommendation approach can be applied. There exist many aggregation strategies to obtain a group profile,[17] highlighting the following ones:

- *Mean*: The average of the ratings is one of the most intuitive aggregation strategy regarding fairness. The group rating r_{Gj}, for the item i_j, is calculated as

$$r_{Gj} = \frac{1}{n} \sum_{i=1}^{n} (r_{ij})$$

- *Root mean squared* (RMS): Every rating is taken into account but high ratings have more weight than low ratings.

$$r_{Gj} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(r_{ij})^2}$$

- *Least misery*: When the priority of the GRS is to reduce the possibility of a user unsatisfied with the recommendations, the minimum aggregation is used.
- *Average without misery*: Averaging individual ratings, after excluding items with individual ratings that are under a certain threshold.

Aggregation always implies a summarization of original information that can imply the loss of information from different points of view such as distribution, diversity, or shape of data. This loss of information can either bias or lead to wrong results regardless of the aggregation operator as can be seen in the following examples:

1. If the *mean* aggregation is used, the following cases are indistinguishable: (a) user u_1 rates 6 to item i_1 and user u_2 rates 5, mean: 5.5, (b) user u_1 rates 10 to item i_1 and user u_2 rates 1, mean: 5.5. Taking into account this group rating for the item i_1 (5.5 in both situations), the GRS can recommend an item i_2 similar to i_1. However, the two situations are pretty different: in situation (a) both users should be satisfied with the recommended item (they like item i_1, similar to i_2), but in situation (b) the user u_2 would probably be unsatisfied because he/she dislikes i_1, although user u_1 would be quite satisfied.
2. The *minimum* operator, used in *least misery* and *average without misery* aggregation operators, is also commonly used in GRS, because it tries to minimize the misery of members to satisfy most users of the group. However, it has an important drawback, mainly for large groups. In this case, only one low rating is enough to penalize an item, because the group rating is the minimum of the whole set of ratings. For large groups, it is likely that every item has at least one low rating and, therefore, the group profile associated with a large group would be mainly composed by negative ratings, leading to low-quality recommendations.[23]

Our aim is to propose a new group recommender model that keeps all group ratings (n ratings for each item, being n the size of the user group) instead of only one aggregated rating for each item, removing the aggregation stage to avoid the loss of information. To do so, we will use the concept of HFSs, which is introduced in the coming section.

2.2. Hesitant Fuzzy Sets for Group Recommendations

A HFS is an extension of fuzzy sets,[24,25] which is defined by a function that returns a set of membership degrees for each element in the domain.

Definition 1.[19] *Let X be a reference set, a HFS on X is a function h that returns a nonempty subset of values in [0,1]:*

$$h : X \rightarrow \wp([0, 1]) \tag{1}$$

Moreover, a HFS can be defined as a set of fuzzy sets.

DEFINITION 2.[19] *Let $M = \{\mu_1, \ldots, \mu_n\}$ be a set of membership functions. The HFS associated with M, h_M, is defined as*

$$h_M : X \to \wp([0,1])$$
$$h_M(x) = \bigcup_{\mu \in M} \{\mu(x)\} \tag{2}$$

Xia and Xu[26] completed the original definition of HFS including the concept of hesitant fuzzy element (HFE), which is a particular subset of values in [0,1] for a particular $x \in X$.

DEFINITION 3.[26] *Let X be a reference set, an HFS on X can be represented as*

$$E = \{\langle x, h_E(x)\rangle : x \in X\} \tag{3}$$

and the set of values $h_E(x)$, for a particular $x \in X$, is called a HFE, which denotes the possible membership degrees of the particular element x.

In this way, for each item in X we have a HFE, that is, a set of membership values in [0,1]. In our case, the hesitation comes from the cardinality of the group: For each item, we have not a unique rating but a set of n ratings, one for each user.

For applying HFSs in group recommendations, it would be necessary to extend some functions defined for crisp values or fuzzy sets.Torra and Narukawa[27] proposed an extension principle that allows to export operations from fuzzy sets to HFS.

DEFINITION 4.[27] *Let $E = \{H_1, \ldots, H_n\}$ be a set of n HFS and Θ a function, $\Theta : [0,1]^n \to [0,1]$, we then export Θ on fuzzy sets to HFSs defining*

$$\Theta_E = \bigcup_{\gamma \in H_1(x) \times \cdots \times H_n(x)} \{\Theta(\gamma)\} \tag{4}$$

Particularly, this principle has been applied to the Pearson correlation coefficient (PCC),[28,29] a function widely used in RSs. Gonzalez-Arteaga et al.[30] extended the PCC, noted as ρ, to the hesitant Pearson correlation coefficient (HPCC), ρ_{HFS}. Figure 3 shows an example of this situation. Given two valuations, X and Y, and three items, i_1, i_2, and i_3, the goal is to measure the correlation between both valuations. Owing to the hesitation that might appear to rate each item, instead of providing only one value, a HFS is used to represent each valuation. The correlation between X and Y is measured by HPCC.

DEFINITION 5. *Let X and Y be two HFSs on S and $h_X(s_i) \times h_Y(s_i)$ be the collection of all pairs of HFEs,*

$$\left((h_X(s_i))^{(j)}, (h_Y(s_i))^{(k)}\right)$$

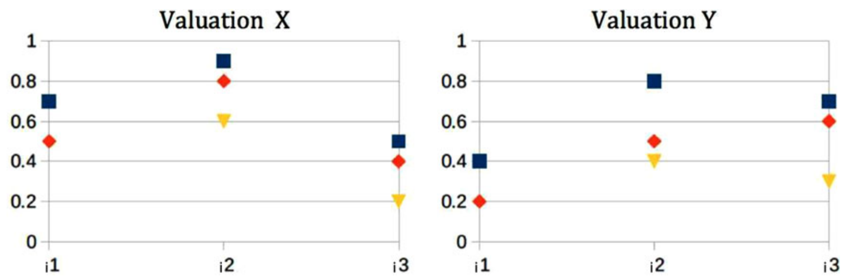

Figure 3. Correlation of hesitant valuations.

where $j \in \{1, \ldots, l_X(s_i)\}$ *and* $k \in \{1, \ldots, l_Y(s_i)\}$, *being* $l_X(s_i)$ *and* $l_Y(s_i)$ *the cardinals of* $h_X(s_i)$ *and* $h_Y(s_i)$, *respectively.*

The set of all pairs HFEs for each $s_i \in S$ *is given by*

$$R_{HFS} = \cup_{s_i \in S} h_X(s_i) \times h_Y(s_i) \tag{5}$$

where $i \in \{1, \ldots, n\}$.

The number of pairs of values in R_{HFS} *is computed as*

$$|R_{HFS}| = \sum_{i=1}^{n} (l_X(s_i) \times l_Y(s_i)) \tag{6}$$

DEFINITION 6.[30] *Let X and Y be two HFSs on S, the HPCC,* ρ_{HFS}, *is defined as follows:*

$$\rho_{HFS}(X, Y) = \frac{SSC(h_X, h_Y)}{\sqrt{SS(h_X)}\sqrt{SS(h_Y)}}, \tag{7}$$

where SSC corresponds to the covariance of both sets and is defined as

$$SSC(h_X, h_Y) = \sum_{i=1}^{n} \sum_{j=1}^{l_X(s_i)} \sum_{k=1}^{l_Y(s_i)} \left((h_X(s_i))^{(j)} - \overline{h_X}\right)\left((h_Y(s_i))^{(k)} - \overline{h_Y}\right), \tag{8}$$

being $\overline{h_X}$ *and* $\overline{h_Y}$ *the arithmetic mean of the corresponding values of the first and second elements of the pairs, respectively.*

$$\overline{h_X} = \frac{1}{|R_{HFS}|} \sum_{i=1}^{n} l_Y(s_i) \left(\sum_{j=1}^{l_X(s_i)} (h_X(s_i))^{(j)} \right) \tag{9}$$

$$\overline{h_Y} = \frac{1}{|R_{HFS}|} \sum_{i=1}^{n} l_X(s_i) \left(\sum_{j=1}^{l_Y(s_i)} (h_Y(s_i))^{(j)} \right) \tag{10}$$

And $SS(h_X)$ and $SS(h_Y)$ correspond to the standard deviation of the respective sets, defined as

$$SS(h_X) = \sum_{i=1}^{n} l_Y(s_i) \left(\sum_{j=1}^{l_X(s_i)} \left((h_X(s_i))^{(j)} - \overline{h_X} \right)^2 \right) \tag{11}$$

$$SS(h_Y) = \sum_{i=1}^{n} l_X(s_i) \left(\sum_{j=1}^{l_Y(s_i)} \left((h_Y(s_i))^{(j)} - \overline{h_Y} \right)^2 \right) \tag{12}$$

This measure is used in the proposed GRS with the aim of improving the recommendations quality.

3. HESITANT GROUP RECOMMENDER MODEL

Here, a new proposal for group recommendation based on CF GRS and HFS is introduced. When the target of a CF RS is not only one user but a group of them, the strategy to find the nearest neighbors is adapted to deal with groups instead of individual users, by finding the nearest neighbors to a group of users. In this way, the proposal changes the traditional KNN algorithm, which uses a classic correlation measure, by another one that uses the correlation coefficient HPCC.

The proposed model fits into the GRS scheme shown in Figure 2c, because it uses a modified KNN algorithm, based on HPCC, which does not imply an aggregation process since it computes the similarity between the target group and the neighbors using the whole set of group ratings. The general scheme of the proposal, HGRM, is depicted in Figure 4, which is divided into three phases:

1. *Similarity measure with HPCC*. The modified KNN algorithm, using HPCC, provides a set of KNN for the group G.
2. *Rating prediction*. Once the neighborhood of G is computed, the system is able to predict ratings for the group according to the ratings of the KNN.
3. *Group recommendation*. The top N items with the highest prediction are recommended to the group.

The remainder of this section is structured as follows. First, Section 3.1 introduces the notation used in the proposal. Afterwards, each phase of the HGRM scheme is explained in further detail. Section 3.2 describes how HPCC is used to find the neighborhood of the group. Section 3.3 explains how the prediction is computed. Finally, Section 3.4 shows how to provide the recommendation.

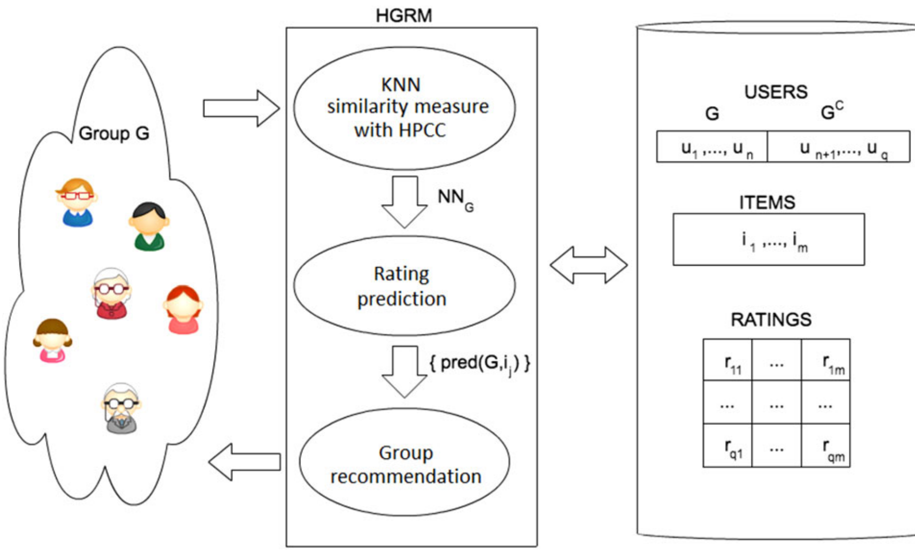

Figure 4. General scheme of HGRM.

3.1. Notation

A CF GRS works with the following information:

- $U = \{u_1, \ldots, u_q\}$ is a set of users.
- $I = \{i_1, \ldots, i_m\}$ is a set of items.
- $R = \{r_{11}, \ldots, r_{ij}, \ldots, r_{qm}\}$ is a set of ratings, where r_{ij} is the rating of the user u_i for the item i_j.
- $D = [d_{min}, \ldots, d_{max}]$ is the domain interval that users utilize for rating items. The most used domain is the 1–5 stars scale.
- $G = \{u_1, \ldots, u_n\} \subset U$, is the target group, with $n \ll q$.
- $G^c = \{u_{n+1}, \ldots, u_q\} = G \setminus U$ is the complementary set of G.

Besides, the proposed HGRM will deal with two types of profiles represented by HFSs:

- The group's profile is defined by a HFS, X_G, to deal with the multiplicity of ratings provided by the members of group G, for each item. X_G is a HFS for the group profile that includes the whole set of ratings of the group G

$$X_G = \{\langle i_j, h_X(i_j)\rangle : i_j \in I\}$$
$$h_X : I \to \wp([0, 1])$$
$$h_X(i_j) = \left\{\frac{r_{ij}}{d_{max}}, \text{for } i = 1, \ldots, n\right\}$$

- Y_k is a HFS for the user profile, $u_k \in G^c$. This is a HFS composed only by one membership

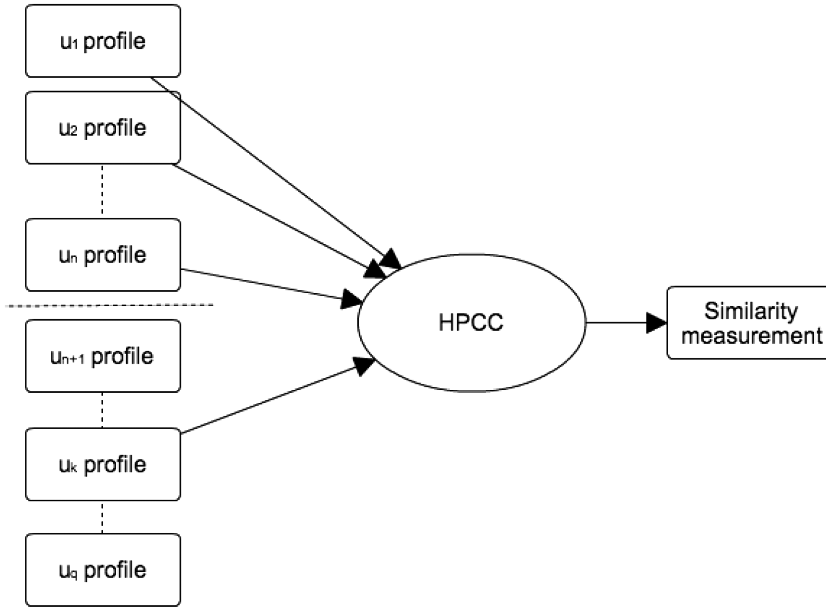

Figure 5. Hesitant Pearson correlation coefficient.

function, that is a special case of a HFS.

$$Y_k = \left\{ \langle i_j, h_Y(i_j) \rangle : i_j \in I \right\}$$
$$h_Y : I \rightarrow \wp([0, 1])$$
$$h_Y(i_j) = \left\{ \frac{r_{kj}}{d_{max}} \right\}$$

Both, $h_X(i_j)$ and $h_Y(i_j)$ are sets; therefore, to refer to an element of the set, superscripts are used. For instance, $(h_X(i_j))^{(z)}$ refers to the zth element of the set.

3.2. Similarity Measure with HPCC for KNN

The KNN algorithm is implemented by using HPCC instead of a traditional similarity measure. It calculates the similarity between the profile of a group G and users in G^c (see Figure 5).

Given a group G and a user $u_k \in G^c$, HPCC (see Equation 7) measures the similarity between their profiles, defined as two HFSs, X_G and Y_k.

$$\rho_{HFS}(X_G, Y_k) = \frac{SSC(h_X, h_Y)}{\sqrt{SS(h_X)}\sqrt{SS(h_Y)}} \tag{13}$$

where

$$SSC(h_X, h_Y) = \sum_{j=1}^{m}\sum_{z=1}^{n}\left((h_X(i_j))^{(z)} - \overline{h_X}\right)\left(\left(h_Y(i_j)\right)^{(1)} - \overline{h_Y}\right)$$

$$\overline{h_X} = \frac{1}{n \cdot m}\sum_{j=1}^{m}\sum_{z=1}^{n}(h_X(i_j))^{(z)}$$

$$\overline{h_Y} = \frac{1}{m}\sum_{j=1}^{m}(h_Y(i_j))^{(1)}$$

and

$$SS(h_X) = \sum_{j=1}^{m}\sum_{z=1}^{n}\left((h_X(i_j))^{(z)} - \overline{h_X}\right)^2$$

$$SS(h_Y) = n \cdot \sum_{j=1}^{m}\left((h_Y(i_j))^{(1)} - \overline{\mu_Y}\right)^2$$

The KNN algorithm will provide a list of the KNN to the group G noted as NN_G, i.e., users u_k who maximize $\rho_{HFS}(X_G, Y_k)$. Given that negative correlations do not lead to good results,[31] only users with positive correlations to the group are considered for the neighborhood.

The HGRM has been developed and tested in two different versions:

- *Standard* version. This is the one described in the previous formulas. The whole set of ratings has been considered, which can include duplicated ratings.
- *No-duplicate* version. It works with unique ratings, deleting repeated ones: if there are two or more users who give the same rating for an item, the system considers only one of these ratings.

3.3. Rating Prediction

Once NN_G has been obtained, the system combines their ratings to compute the predictions. According to the traditional CF approach, recommended items are the best rated items of the KNN. There are two main prediction approaches to compute the predicted rating for a given item, according to[6]

- *Direct prediction*: Predictions are obtained directly from the ratings of the nearest neighbors, weighted by their similarity to the target group. To obtain values correctly scaled, the equation is normalized by dividing by the sum of the neighbors' similarities.

$$pred\left(G, i_j\right) = \frac{\sum_{u_k \in NN_G} \rho_{HFS}\left(X_G, Y_k\right) \cdot r_{kj}}{\sum_{u_k \in NN_G} \rho_{HFS}\left(X_G, Y_k\right)} \tag{14}$$

- *Compensated prediction*: Optimistic users may rate favorite items with four or five stars, whereas pessimistic ones rate with two or three stars. To compensate the ratings scale variations, the deviation from users ratings averages is used. This is the prediction used in our case study.

$$pred(G, i_j) = \overline{r}_G + \frac{\sum_{u_k \in NN_G} \rho_{HFS}(X_G, Y_k) \cdot (r_{kj} - \overline{r}_k)}{\sum_{u_k \in NN_G} \rho_{HFS}(X_G, Y_k)} \tag{15}$$

where $\overline{r}_G$ is the average value of the set of ratings $\{r_{ij}\}$, being $u_i \in G$, and $\overline{r}_k$ of the set $\{r_{kj}\}$, for each item i_j.

3.4. Group Recommendation

Finally, the system recommends an ordered list $\tilde{I}$ consisting of the top N items with the highest predicted ratings.

$$\tilde{I} = (\langle \tilde{i}_1, \tilde{r}_1 \rangle, \cdots \langle \tilde{i}_N, \tilde{r}_N \rangle) \tag{16}$$

where $\tilde{r}_j = pred(G, \tilde{i}_j)$ and $\tilde{r}_j \geq \tilde{r}_k$ for each $j < k$.

4. CASE STUDY

This section presents a case study to evaluate the proposed HGRM by comparing it with a classical GRS model.

4.1. Experiment Description

The HGRM model is compared with two versions of the traditional GRS (see Figure 2a): (a) using the *Mean* function for aggregating user ratings and (b) using the *RMSMean* function for this aggregation process (see Section 2.1). Moreover, the evaluation has been performed for the two versions of the HGRM described previously: *standard* and *no-duplicate*. Thus, four models have been compared:

- *Mean*: Traditional GRS with *Mean* aggregation.
- *RMSMean*: Traditional GRS with *RMSMean* aggregation.
- *HGRM*: It is the standard version of the proposed model.
- *HGRM no-dup.*: It is the no-duplicate version of the proposal.

Outcomes of the system is a list $\tilde{I}$ of N recommended items, testing with $N \in \{3, 4, 5, 6\}$, for the sake of clarity, only the results for $N = 5$ are shown. The results of each evaluation measure have been calculated for 20 executions, and each execution with a 5-cross fold validation.

4.2. Data Set

The data set used for evaluating the proposal is the ml-100k[a], which consists of 1682 items, 943 users, and 100k ratings. Users evaluate items by using a *five stars* domain, being one star the minimum value and five stars the maximum one (ratings are normalized to deal with HFS).

The MovieLens data set does not contain information about the groups; therefore, they could be selected using three different group formations:

- *Random groups*: Random group formation matches the situation of a number of users who group to do an activity.[32]
- *Similar groups*: Users group following the principle of homophily, that is, the groups are formed by users with similar features, such as interests, beliefs, education, or age.
- *Dissimilar groups*: Users group following the principle of heterophily, that is, the groups are formed by users with diversity of features.

Our study is focused on random groups, which is the most challenging type of groups for GRSs.

Different group sizes have been considered in the experiments, ranging from 1 to 500 users. But for the sake of clarity, only the results for groups of size 20, 25, 50, 100, 200, and 500 are shown.

4.3. Evaluation Measures

We aim at evaluating the GRSs from different points of view: accuracy, rank quality, and diversity. Hence, three widely used evaluation measures: *normalized root mean squared error* (NRMSE), *normalized discounted cumulative gain* (NDCG), and *intra list similarity* (ILS), are used respectively.[33,34]

We stated the hypothesis for our proposal that keeping all information from group members by avoiding aggregation processes, the GRS performance will improve taking into account different properties.

Hence, to validate our hypothesis the four approaches for group recommendation are compared by using the following evaluation measures:

- NRMSE[35] is a quantity commonly used to calculate the accuracy of a RS, which measures how close predictions are to the true ratings. As aforementioned, the evaluation has been done using 5-cross fold validation, where, for each recommended item, we have the predicted value $\tilde{r}_i$ and the real value r_i.

$$NRMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\tilde{r}_i - r_i}{d_{max} - d_{min}}\right)^2} \tag{17}$$

- NDCG[36] measures the ranking quality through the utility or gain of a recommendation list. It provides a quantity of the performance of a recommendation system based on the graded relevance of the recommended entities. It ranges from 0.0 to 1.0, with 1.0

[a] http://grouplens.org/datasets/movielens/

Table I. Results for NRMSE.

	Size 20	Size 25	Size 50	Size 100	Size 200	Size 500
Mean	0.25558	0.25572	0.25620	0.25665	0.25747	0.25890
RMSMean	0.25561	0.25576	0.25624	0.25669	0.25751	0.25894
HGRM	0.25591	0.25609	0.25659	0.25702	0.25781	0.25926
HGRM no-dup.	0.25603	0.25620	0.25676	0.25728	0.25815	0.25996

representing the ideal ranking of the entities.

$$\mathrm{DCG} = \sum_{i=1}^{N} \frac{\tilde{r}_i - 1}{\log_2(i+1)}$$

$$\mathrm{NDCG} = \frac{\mathrm{DCG}}{\mathrm{IDCG}} \tag{18}$$

where IDCG is the ideal or maximum possible value of DCG and the relevance of a recommended item is its predicted rating $\tilde{r}_i$. The other measures considered, NRMSE and ILS, are measures to minimize. Therefore, for consistency of the results, we use $1 - \mathrm{NDCG}$ rather than NDCG.

- ILS.[37] Diversity is an important ingredient to user satisfaction, in any RS. This metric intends to capture the similarity of a list, referred to all kinds of features that describe items. The less similarity, the more diversity of the recommendation.

$$\mathrm{ILS}\left(\tilde{I}\right) = \frac{\sum_{i_j \in \tilde{I}} \sum_{i_k \in \tilde{I}, j \neq k} \mathrm{cosine}\left(v_j, v_k\right)}{2} \tag{19}$$

where v_j and v_k are, respectively, characteristic vectors of items i_j and i_k built using the singular value decomposition (SVD)[7] with 20 features.

Therefore, we will analyze initially these three properties in our case study according to the previous metrics. But as it was mentioned in the Introduction, a GRS should balance these properties, because they can be conflicting among them, to achieve better recommendations from a more general and realistic point of view and not only from accuracy viewpoint. Hence, after the analysis of each property we will also present a study about their combination to balance these three properties.

4.4. Results

In this section, the results obtained for the different approaches are shown and analyzed to evaluate the performance of our proposal. This analysis considers the accuracy and diversity for each approach and then analyzes their performance regarding the combination of both properties for the group recommendation.

Tables I and II show results for NRMSE and 1-NDCG, respectively. The results are shown for group sizes ranging from 20 to 500. These measures should be minimized, i.e., the lower value, the better.

Table II. Results for 1-NDCG for a recommendation list of 5 items.

	Size 20	Size 25	Size 50	Size 100	Size 200	Size 500
Mean	0.07250	0.07252	0.07256	0.07247	0.07265	0.07266
RMSMean	0.07250	0.07252	0.07256	0.07247	0.07265	0.07266
HGRM	0.07250	0.07252	0.07256	0.07247	0.07266	0.07266
HGRM no-dup.	0.07250	0.07252	0.07256	0.07247	0.07265	0.07266

Table III. Results for ILS for a recommendation list of five items.

	Size 20	Size 25	Size 50	Size 100	Size 200	Size 500
Mean	0.84599	0.83347	0.78712	0.71462	0.63706	0.44160
RMSMean	0.84833	0.83739	0.79293	0.71737	0.63849	0.44423
HGRM	0.85172	0.84111	0.79121	0.71535	0.62920	0.38945
HGRM no-dup.	0.85216	0.84227	0.79437	0.71180	0.62109	0.40837

These tables show that the four approaches have an identical performance in ranking quality (1-NDCG) and with differences less than one thousandth for accuracy (NRMSE).

Once the proposal has been evaluated regarding the NRMSE and NDCG, the diversity of the group recommendations is taken into account by measuring the ILS of the recommendations.

Table III shows the results of ILS in which HGRM and HGRM no-dup. show a clear tendency, *the bigger the group size the better the diversity*. Consequently, we can conclude that all approaches perform identically regarding *ranking quality*, there are small differences in *accuracy* and significant differences in *diversity* in favor of the HGRM proposals when the group size grows up, which validates our hypothesis for big groups.

Additionally, according to our view of balancing different GRS properties to obtain better recommendations from a global point of view, it can be considered that ranking quality can be ignored in this case study because all approaches perform identically. However, accuracy and diversity should be balanced to show the performance of the four approaches. We then propose to this analysis a convex combination in which different weights, α, can be assigned to each property:

$$\mathrm{NRMSE} \cdot \alpha + \mathrm{ILS} \cdot (1 - \alpha)$$

The quotient $ratio = \alpha/(1 - \alpha)$ provides the importance rate of accuracy over diversity. The combined metric facilitates the comparison of GRSs taking into account both properties at the same time.

The models have been evaluated with different values for α, which determines the relative importance of the accuracy over the diversity. We have set $\alpha \in \{0.75, 0.50, 0.25\}$ to combine the measures, giving more importance to the accuracy ($ratio = 3/1$), equal importance to both ($ratio = 1/1$), and less importance to the accuracy with respect to the diversity ($ratio = 1/3$), respectively. The

Table IV. Results for NRMSE $\cdot \alpha +$ ILS $\cdot (1-\alpha)$, $\alpha = 0.75$.

	Size 20	Size 25	Size 50	Size 100	Size 200	Size 500
Mean	0.40318	0.40015	0.38893	0.37114	0.35236	0.30457
RMSMean	0.40379	0.40116	0.39041	0.37185	0.35275	0.30526
HGRM	0.40486	0.40234	0.39024	0.37160	0.35065	0.29180
HGRM no-dup.	0.40505	0.40271	0.39116	0.37090	0.34888	0.29706

Table V. Results for NRMSE $\cdot \alpha +$ ILS $\cdot (1-\alpha)$, $\alpha = 0.50$.

	Size 20	Size 25	Size 50	Size 100	Size 200	Size 500
Mean	0.55078	0.54459	0.52166	0.48563	0.44726	0.35025
RMSMean	0.55197	0.54657	0.52458	0.48702	0.44799	0.35158
HGRM	0.55381	0.54859	0.52390	0.48618	0.44350	0.32435
HGRM no-dup.	0.55409	0.54923	0.52556	0.48453	0.43961	0.33416

Table VI. Results for NRMSE $\cdot \alpha +$ ILS $\cdot (1-\alpha)$, $\alpha = 0.25$.

	Size 20	Size 25	Size 50	Size 100	Size 200	Size 500
Mean	0.69839	0.68902	0.65439	0.60012	0.54216	0.39592
RMSMean	0.70015	0.69198	0.65875	0.60219	0.54324	0.39790
HGRM	0.70276	0.69485	0.65755	0.60076	0.53635	0.35690
HGRM no-dup.	0.70312	0.69575	0.65996	0.59816	0.53035	0.37126

results for each α are shown in Tables IV, V and VI, respectively. In the same way, Figures 6a–6c show the results for the three different combinations.

As it can be observed for all α values used, HGRM and HGRM no-dup. achieve a remarkable difference as compared to the traditional models as the group size increases, showing clear improvements for group size greater than 100 users. This difference suggests that the improvements of HGRM and HGRM no-dup. in terms of diversity of the recommendations do not have a negative influence in the accuracy of the system.

We can conclude that HGRM and HGRM no-dup. properly balance the accuracy and diversity when recommending to large groups, making them suitable for recommendations in contexts in which the relative importance of accuracy over diversity is less or equal to 3. These results confirm our hypothesis, which is that keeping all information from group members by avoiding aggregation processes, the GRS performance will improve taking into account different properties.

Remark 1. We have just focused on random groups in this section, but we carried out the same experiments for similar and dissimilar groups. Obtaining similar comparative performance among the different algorithms, but on the one hand, for similar groups the performance of each algorithm is better because it is easier to recommend to groups of users with similar interests. On the other hand, for dissimilar groups it

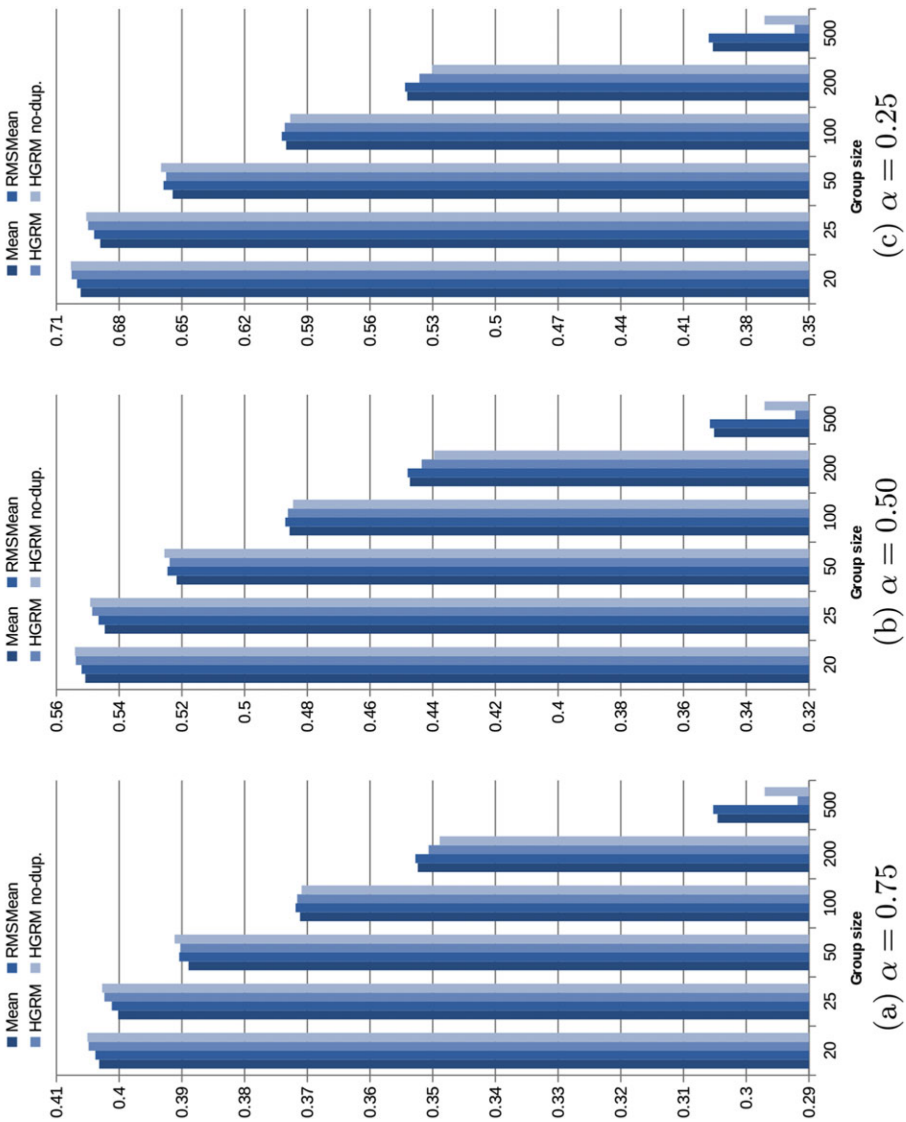

Figure 6. Results for NRMSE $\cdot \alpha +$ ILS $\cdot (1 - \alpha)$.

is more difficult to satisfy the members; hence, the results for dissimilar groups are worse than previous ones. These trends are consistent for all group sizes explored.

5. CONCLUSIONS

This paper has proposed a model of GRSs based on HFSs to model the hesitation of group preferences, which avoids the aggregating process and the bias of information associated with it. To evaluate and validate the better performance of the proposal, a case study has been carried out on a GRS based on CF and it shows improvements using HFS in RSs for large groups randomly formed, mainly regarding *diversity* in recommendations.

The diversity of recommendations is a desired property in GRSs for large groups because members may have either different or conflicting interests; therefore, recommending diverse items increases the possibility of all members being satisfied with at least one item.

The case study reveals that the proposed model based on HFSs enhances the diversity of the recommendations as long as group size increases. Additionally, a mixed metric of *accuracy* and *diversity* presents remarkable improvements for different relative importance of accuracy over diversity. Hence, the new proposal not only provides accurate recommendations but also obtains recommendations composed of diverse items increasing the group members satisfaction regarding the recommended items.

The promising results obtained by the use of HFSs in GRSs encourage us to further research about their application in other types of GRSs, such as content based or hybrid GRSs, as a future work.

Acknowledgments

This paper was partially supported by the Spanish National Research Project TIN2015-66524-P, Spanish Ministry of Economy and Finance Postdoctoral Training (FPDI-2013-18193), Spanish FPU fellowship (FPU13/01151), and ERDF.

References

1. Schafer JB, Konstan J, Riedl J. Recommender systems in e-commerce. In: EC '99 Proceedings of the 1st ACM conference on Electronic commerce, Denver, Colorado, USA, November 3–5, 1999. pp 158–166.
2. Adomavicius G, Tuzhilin A. Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions. IEEE Trans Knowl Data Eng 2005;17:734–749.
3. Ricci F, Rokach L, Shapira B, Kantor P, editors. Recommender systems handbook. Berlin: Springer; 2011.
4. Resnick P, Iacovou N, Suchak M, Bergstrom P, Riedl J. An open architecture for collaborative filtering of netnews. In: Proceeding CSCW '94 Computer-Supported Cooperative Work 94, Chapel Hill, North Carolina, USA, October 22–26, 1994. pp 175–186.
5. Herlocker JL, Konstan JA, Borchers A, Riedl J. An algorithmic framework for performing collaborative filtering. In: Proceedings: 22nd Annual International ACM SIGIR Conference

on Research and Development in Information Retrieval, Berkeley, CA, USA, August 15–19, 1999. pp 230–237.
6. Schafer J, Frankowski D, Herlocker J, Sen S. Collaborative filtering recommender systems. In: The adaptive web. Berlin: Springer; 2007. pp 291–324.
7. Koren YR, Volinsky RBC. Matrix factorization techniques for recommender systems. Computer 2009;42:30–37.
8. Martínez L, Pérez L, Barranco M. A multi-granular linguistic content-based recommendation model. Int J Intell Syst 2007;22:419–434.
9. Pazzani M, Billsus D. Content-based recommendation systems. In: Brusilovsky P, Kobsa A, Nejdl W, editors. The adaptive web: Methods and strategies of web personalization, volume 4321 of LNCS. Berlin: Springer-Verlag; 2007. pp 325–341.
10. Lops P, de Gemmis M, Semeraro G. Content-based recommender systems: state of the art and trends. In: Ricci F, Rokach L, Shapira B, Kantor P, editors. Recommender systems handbook. Berlin: Springer-Verlag; 2011. pp 73–105.
11. Castro J, Rodríguez RM, Barranco MJ. Weighting of features in content-based filtering with entropy and dependence measures. Int J Comput Intell Syst 2014;7:80–89.
12. Burke R. Hybrid recommender systems: survey and experiments. User Model User-Adap 2002;12:331–370.
13. Albadvi A, Shahbazi M. A hybrid recommendation technique based on product category attributes. Expert Syst Appl 2009;36:11480–11488.
14. Vargas S, Castells P. Rank and relevance in novelty and diversity metrics for recommender systems. In: Proceedings: RecSys '11 Fifth ACM Conference on Recommender Systems, Chicago, IL, USA, October 23–27, 2011. pp 109–116.
15. Zhou T, Kuscsik Z, Liu J, Medo M, Wakeling J, Zhang Y. Solving the apparent diversity-accuracy dilemma of recommender systems. Pro Nat Acad Sci U S A 2010;107: 4511–4515.
16. Jameson A, Smyth B. Recommendation to groups. In: Brusilovsky P, Kobsa A, Neidl W, editors. The adaptive web: Methods and strategies of web personalization, volume 4321 of Lecture Notes in Computer Science. Berlin: Springer-Verlag; 2007. pp 596–627.
17. Masthoff J. Group recommender systems: combining individual models. In: Ricci F, Rokach L, Shapira B, Kantor P, editors. Recommender systems handbook. Berlin: Springer; 2011. pp 677–702.
18. Castro J, Quesada F, Palomares I, Martínez L. A consensus-driven group recommender system. Int J Intell Syst 2015;30:887–906.
19. Torra V. Hesitant fuzzy sets. Int J Intell Syst 2010;25:529–539.
20. Zadeh L. Fuzzy sets. Inform Control 1965;8:338–353.
21. McCarthy JF, Anagnost TD. Musicfx: an arbiter of group preferences for computer supported collaborative workouts. In: The 1998 ACM Conference on Computer Supported Cooperative Work. New York: ACM; 1998. pp 363–372.
22. Lieberman H, Van Dyke NW, Vivacqua AS. Let's browse: A collaborative web browsing agent. In: The 4th International Conference on Intelligent User Interfaces, Redondo Beach, CA, Jan 5–8, 1999. New York: ACM; 1999. pp 65–68.
23. Boratto L, Carta S. Art: group recommendation approaches for automatically detected groups. Int J Mach Learn Cybern 2015;6:953–980.
24. Rodríguez R, Martínez L, Torra V, Xu Z, Herrera F. Hesitant fuzzy sets: state of the art and future directions. Int J Intell Syst 2014;29:495–524.
25. Rodríguez R, Bedregal B, Bustince H, Dong Y, Farhadinia B, Kahraman C, Martínez L, Torra V, Xu Y, Xu Z, Herrera F. A position and perspective analysis of hesitant fuzzy sets on information fusion in decision making towards high quality progress. Inform Fusion 2016;29:89–97.
26. Xia N, Xu Z. Hesitant fuzzy information aggregation in decision making. Int J Approx Reason 2011;52:395–407.
27. Torra V, Narukawa Y. On hesitant fuzzy sets and decision. In: Proceedings: 18th International Conference on Fuzzy Systems, Jeju Island, Korea, August 20–24, 2009. pp 1378–1382.

28. Pearson K. Notes on the history of correlation. Biometrika 1920;13:25–45.
29. Rodgers JL, Nicewander WA. Thirteen ways to look at the correlation coefficient. Am Stat 1998;42:59–66.
30. González-Arteaga T, Alcantud J, de Andrés Calle R. New correlation coefficients for hesitant fuzzy sets. In: Proceedings: 16th World Congress of the International Fuzzy Systems Association (IFSA), June 30th to July 3rd, Gijón, Spain, 2015. pp 427–434.
31. Massa P, Avesani P. Trust-aware collaborative filtering for recommender systems. In: Meersman R, Tari Z, editors. The International Conference on Cooperative Information Systems (CoopIS). Berlin: Springer; 2004. pp 492–508.
32. Wang W, Zhang G, Lu J. Member contribution-based group recommender system. Decis Support Syst 2016;87:80–93.
33. Candillier L, Chevalier M, Dudognon D, Mothe J. Diversity in recommender systems. In: Proceedings: The Fourth International Conference on Advances in Human-oriented and Personalized Mechanisms, Technologies, and Services. CENTRIC 2011, October 23–29, Barcelona, Spain, 2011. pp 48–53.
34. Ghanghas Y, Rana C, Dhingra S. Diversity in recommender system. Int J Eng Trends Technol 2013;4:2344–2348.
35. Cremonesi P, Koren Y, Turrin R. Performance of recommender algorithms on top-n recommendation tasks. In: The Fourth ACM Conf on Recommender Systems, 2010. pp 39–46.
36. Baltrunas L, Makcinskas T, Ricci F. Group recommendations with rank aggregation and collaborative filtering. In: The Fourth ACM Conf on Recommender Systems, 2010. pp 119–126.
37. Ziegler C, McNee S, Konstan J, Lausen G. Improving recommendation lists through topic diversification. In: 14th Int Conf on World Wide Web, Chiba, Japan, 2005. pp 22–32.