Expert Systems With Applications 184 (2021) 115444

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Content-based group recommender systems: A general taxonomy and further improvements

Yilena Pérez-Almaguer [a], Raciel Yera [b], Ahmad A. Alzahrani [c], Luis Martínez [d,*]

[a] University of Holguín, Holguín, Cuba
[b] University of Ciego de Ávila, Ciego de Ávila, Cuba
[c] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia
[d] Computer Science Department, University of Jaén, Jaén, Spain

ARTICLE INFO

Keywords:
Content-based recommendation
Group recommender systems
Machine learning
Personalization

ABSTRACT

Group recommender systems have emerged as a solution to recommend interesting, suitable, and useful items that are consumed socially by groups of people, rather than individually. Such systems have pushed for the use of new recommendation methods within such an emerging scenario, in which the use of the collaborative filtering paradigm is the core of the recommender algorithm. However, collaborative filtering presents several drawbacks and limitations in this scenario, such as the need for lots of rating values, as well as their co-occurrence across several items and users (scarcity). In order to overcome these drawbacks, this research explores a taxonomy for content-based group recommendation systems (CB-GRS), and subsequently the paper discusses and analyzes three specific models that can be used to build CB-GRS, which are (1) CB-GRSs supported by recommendation aggregation and individual ranking, (2) CB-GRSs supported by recommendation aggregation and user-item matching, and (3) CB-GRSs supported by the aggregation of user profiles. Furthermore, the paper presents a hybrid CB-GRS that combines the models (2) and (3) and integrates feature weighting and aggregation function switching. An experimental protocol over well-known datasets is then developed in order to evaluate the proposals. The current study aims at providing a basis to develop a research branch concerning content-based group recommender systems.

1. Introduction

Recommender systems (RSs) are systems that produce personalized recommendations as output or that have the effect of guiding the user to choose, in a personalized way, interesting and useful products in an overloaded product space (Burke, 2002). RSs perform several core tasks in their recommendation process (Ricci, Rokach, & Shapira, 2011): (1) to predict unknown ratings, (2) to generate a listing of the items recommended.

RSs have become powerful tools, considering that they are used by most important e-marketplaces obtained from a diversity of domains. Key scenarios are Amazon, thanks to the recommendation of diverse products; YouTube, as users are recommended different videos that might be of interest to them according to the their previous viewing history in the system; Facebook, which has recommendation systems that suggest friends with common interests; or Spotify, for music recommendation (Lu, Wu, Mao, Wang, & Zhang, 2015). Other application areas in which RSs have been widely applied are: e-learning (Yera, Caballero Mota, & Martínez, 2018; Yera & Martínez, 2017), e-tourism (Nilashi, bin Ibrahim, Ithnin, & Sarmin, 2015), e-health (Yera, Alzahrani, & Martínez, 2019), etc.

Most RSs have been developed by using two main paradigms: (1) content-based recommendation (Lops, Gemmis, & Semeraro, 2011) which are centered on suggesting items that contain similar attributes to other items that were preferred in the past by the same user, and (2) collaborative filtering-based recommendation (Ekstrand, Riedl, & Konstan, 2011), centered on suggesting items preferred by users that are similar to the active user.

On the other hand, the majority of RSs research has been focused on recommending items to individual users (Adomavicius & Tuzhilin, 2005; Bobadilla, Ortega, Hernando, & Gutiérrez, 2013; Yera & Martínez, 2017). However, in the last decade there has been a significant increasing interest in group recommendation thanks to certain items, called social items, that tend to be consumed in groups and not by an

* Corresponding author.
E-mail addresses: yilena@uho.edu.cu (Y. Pérez-Almaguer), ryera@unica.cu (R. Yera), aalzahrani8@kau.edu.sa (A.A. Alzahrani), martin@ujaen.es (L. Martínez).

https://doi.org/10.1016/j.eswa.2021.115444
Received 22 October 2020; Received in revised form 14 May 2021; Accepted 12 June 2021
Available online 30 June 2021

individual user (Castro, Yera, & Martínez, 2017; De Pessemier, Dooms, & Martens, 2014). TV programs and touristic packages are examples of social items. In this way, group recommender systems (GRSs) are focused on extending individual RS to recommend an item to a group of users, supported by the aggregation of the information associated to each individual user (Castro et al., 2017).

Even though group recommendation could be identified as a research area of considerable development over the last few years, a common feature of such recent studies is that collaborative filtering and the relationship between users have been considered the main sources of information used to build the proposed models (Castro, Quesada, Palomares, & Martínez, 2015; Castro et al., 2017; Ghazarian & Nematbakhsh, 2015; Mahyar et al., 2017; Ortega, Hernando, Bobadilla, & Kang, 2016; Seo, Kim, Lee, Seol, & Baik, 2018; Wang, Jiang, Sun, Liu, & Liu, 2018; Wang, Zhang, & Lu, 2016). Furthermore, other hybrid developments focused on integrating the item features into collaborative filtering-based models can be identified (De Pessemier, Dhondt, Vanhecke, & Martens, 2015; Kaššák, Kompan, & Bieliková, 2016).

The use of collaborative filtering in GRSs demands the availability of a huge amount of rating values as well as their co-occurrence across several items and users. However, RSs are typically associated to highly sparse scenarios where collaborative filtering may then achieve a low performance. In such a context, content-based recommendation might play a relevant role, as it only depends on the information associated with the active user that requests the recommendation, and does not depend on rating co-occurrences across the users. Meanwhile, whilst typical content-based recommendation methods have been widely used since the 90s (Lops et al., 2011) for individual recommendation, there is a lack of research exploring the use of such methods with GRSs, with a brief mention in the recent GRS book (Felfernig, Boratto, Stettinger, & Tkalčič, 2018).

Consequently, multiple factors have motivated the analysis and study of applying content-based recommendation for group recommendation. Recently, some research has suggested that content-based recommenders alone, or their hybridization with other recommendation approaches, can be more effective than their collaborative filtering counterparts in some real-world application domains (Kirshenbaum, Forman, & Dugan, 2012; Lops, Jannach, Musto, Bogers, & Koolen, 2019). GRS is one of such domains of improvement. Additionally, such studies show that content-based techniques would allow for direct exploitation, in the group scenario, of benefits formerly associated with the content-based paradigm such as the typical handling of the cold-start problem (Adomavicius & Tuzhilin, 2005). Furthermore, in the near future it would open up new research possibilities in the group scenario to explore well-known RS goals that usually depend on content information, such as diversity reaching (Kaminskas & Bridge, 2016), semantic-awareness (de Gemmis, Lops, Musto, Narducci, & Semeraro, 2015), or the use of rich item descriptions (Lops et al., 2019).

As such, this paper is devoted to introduce a taxonomy for content-based recommendation approaches in the group recommendation scenario, starting with a previously proposed taxonomy for group recommendation (De Pessemier et al., 2014), and a few existing developments in this area. Hence, our research goes beyond the current state of art. Specifically, the main contributions of the paper are:

- The introduction of a taxonomy for content-based group recommendation system (CB-GRSs) models.
- The screening of three CB-GRSs models, which are (1) CB-GRSs supported by recommendation aggregation and individual ranking, (2) CB-GRSs supported by recommendation aggregation and user-item matching, and (3) CB-GRSs supported by the aggregation of users' profiles.
- A hybrid CB-GRS that combines the CB-GRS supported by recommendation aggregation and user-item matching and the CB-GRS supported by the aggregation of users' profiles, which also integrates feature weighting and aggregation function switching.
- An experimental study designed to characterize the behavior of such models to find the best option to be used in a practical scenario.

The paper is structured as follows. Section 2 presents the necessary background for the proposal presentation, including content-based recommendation, group recommender systems, as well as previous studies on content-based group recommender systems. Section 3 introduces a taxonomy for content-based group recommendation, which is specifically composed of the three alternative models for building content-based group recommender systems, supported by recommendation aggregation and individual ranking, by recommendation aggregation and user-item matching, and by the aggregation of users' profiles. Furthermore, Section 4 presents a hybrid CB-GRS that combines the CB-GRS supported by the last two models, as well as integrating feature weighting and aggregation function switching. Section 5 introduces a comparison among the models presented in previous sections, in terms of their advantages and disadvantages. Section 6 develops several experiments to study the performances of the proposed frameworks and each design alternative individually, and to compare this framework with a collaborative filtering-based group recommendation framework, in order to measure the value of the current proposal as compared to pre-existing schemes. Section 7 concludes the paper and points out future research proposals.

2. Background

This section is focused on revising some background content which is essential to be able to understand the proposal of this paper. Specifically, it revises some key concepts regarding content-based recommendation, group recommender systems, as well as content-based group recommender systems.

2.1. Content-based recommendation

Content-based recommendation has become a relevant paradigm in the development of recommender systems (Bobadilla et al., 2013; Yera & Martínez, 2017). In this approach, the value of item i for user u is calculated using those values assigned by user u to items s_i that are more similar to item i (Adomavicius & Tuzhilin, 2005; Pazzani & Billsus, 2007). Here, the features associated with the items (e.g. actors, director, genres, in the case of movies), play a relevant role in recommendation.

Specifically, in content-based recommendation (in Fig. 1) the item profile $Content(i)$ represents a set of attributes that characterize item i, and is usually calculated by extracting a set of features from i. Several approaches have been proposed for this feature representation, including the creation of binary profiles to directly represent the presence/absence of a feature (Adomavicius & Tuzhilin, 2005), weighting schemes taken from text-based measures (Aizawa, 2003), the use of latent semantic analysis to represent textual items (Castro et al., 2019), the use of tag-based profiles (Gedikli & Jannach, 2013), and other strategies.

The calculated item profiles are then used in combination with items previously rated by the same user in order to recommend those items that are related. To do so, we must define a user profile $ContentBasedProfile(u)$, which is built based on the previously identified item profiles associated with the items preferred by the associated user. Such a user profile represents the preference of the user u over the set of items, and there are several ways to create it, including aggregation approaches (Adomavicius & Tuzhilin, 2005), machine learning-based approaches (Pazzani, 1999), or semantic approaches (Movahedian & Khayyambashi, 2014).

Considering both profiles, the utility $v(u,i)$ of the item i for the user u is calculated as (Eq. 1):

$$v(u,i) = score(ContentBasedProfile(u), Content(i)) \quad (1)$$

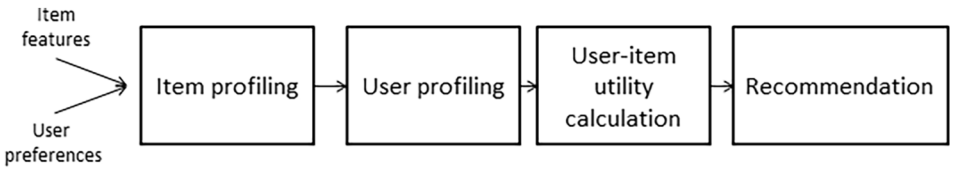

Fig. 1. General scheme of content-based recommendation.

where the score function is usually represented by an information retrieval-related function such as cosine or Jaccard measures (Adomavicius & Tuzhilin, 2005; Pazzani & Billsus, 2007; Bobadilla et al., 2013), as well as semantic similarity approaches (Pera & Ng, 2014) and tag-based similarities (Movahedian & Khayyambashi, 2014).

Finally, the top-n recommendation items for the user u (Gunawardana & Shani, 2015), are obtained by sorting all the available items downwardly according to their score values (Eq. 1).

Content-based recommendation is currently a very active research area (de Gemmis et al., 2015). In this research, we have followed the previous scheme in order to present a general framework for content-based group recommender systems.

2.2. Group recommender systems

Over the last few years, group recommender systems (GRSs) (Castro, Yera, & Martínez, 2018; De Pessemier et al., 2014) have started considering that there are groups of items, called social items, such as TV programs and touristic packages, that tend to be consumed by groups and not by individual users. In these situations, recommending items that satisfy group preference might be more difficult to achieve than the individual recommendation goal. Therefore, GRSs are focused on extending individual RS to recommend an item to a group of users, supported by the aggregation of the information associated with each individual user (Castro, Barranco, Rodríguez, & Martínez, 2018; Castro et al., 2017).

Specifically, GRSs extend RSs to target recommendations for groups of users ($G = \{g_1, \ldots, g_m\} \subseteq U$). Formally, GRSs focus on finding the item (or set of items) that maximizes the preferences predicted for the group of users:

$$Recommendation(I, G_a) = arg\max_{i_k \in I} Score(i_k, G_a) \quad (2)$$

i_k being an item in the item set I, and G_a a group composed of several users u.

There are two main approaches for group recommendation, which are built based on the individual user recommendation (De Pessemier et al., 2014):

- *Rating aggregation:* It creates a pseudo-user profile that combines the preference of the group. This profile is used as the final target user for generating recommendations (Fig. 2)
- *Recommendation aggregation:* It is based on the aggregation of the individual recommendation list associated with each member of the group (Fig. 3), and works toward a new recommendation list that is focused on the group.

2.3. Previous research on content-based group recommender systems

Even though the development of recommendation approaches for the group scenario has recently become a popular research goal, as far as we know there is a shortage of research that focus on exploring the role of content-based recommender systems (following the scheme in Fig. 1), in this scenario. This section covers the most relevant research that is at least partially focused on this topic, identifying the presence of research gaps that are addressed in the current paper. For a better analysis, we structure such research into three groups based on their characteristics: (1) Proposals that integrate multiple information sources, (2) Hybrid proposals for the tourism domain, and (3) Proposals integrating content-based and collaborative filtering.

Proposals that integrate multiple information sources. In this approach, several studies have incorporated the content-based recommendation paradigm into larger recommendation architectures that are composed by social networks-based and other similar sources of information. In this case, Quijano-Sanchez, Recio-Garcia, and Diaz-Agudo (2014) focused on incorporating social behavior knowledge into GRS. Here the individual predictions are enriched with social elements such as the assertiveness and cooperativeness dimensions. A large case study composed of real and synthetic datasets was developed to evaluate the proposal. In the same way, Yuan, Cong, and Lin (2011) outline the task of incorporating content information into group recommendation as a complementary component of their proposal. The use of a probabilistic approach that relates feature values with users is proposed, mainly focused on group formation across pre-established topics associated with the user. The evaluation is done using Precision and Recall measures, event-based and location-based social network datasets, and Movielens datasets. Zhang (2016) also considers a group-centric recommendation architecture that takes information from multiple sources such as behavior, user reviews, and preferences modeling. However, it is not focused on proposing an integrated model to manage such information.

Hybrid proposals for the tourism domain. In this group, Ardissono, Goy, Petrone, Segnan, and Torasso (2003) present one of the first proposals for GRS, centered on a system that suggests tourist attractions to groups. This system considers a content-based approach for modelling groups and items. However, the evaluation is only at system level and does not include accuracy measuring. Focused on the same tourism domain, De Pessemier et al. (2015) present a hybrid group recommender system with content-based recommendation as one of its dimensions, and that consider the individual recommendation aggregation of both the users' score and the users' personal ranking. Here the items (tourist destination) are modeled using tags. The experimentation used data gathered from a real world application developed as a result of this study. Khoshkangini, Pini, and Rossi (2016) propose a self-adaptive context-

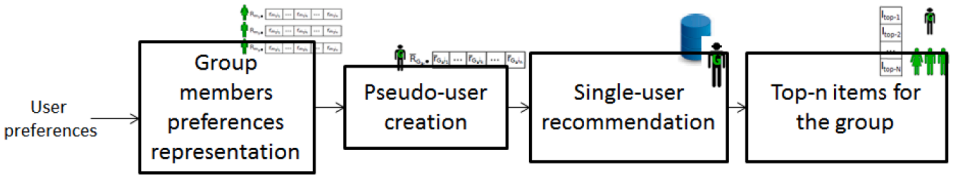

Fig. 2. Group recommendation based on rating aggregation.

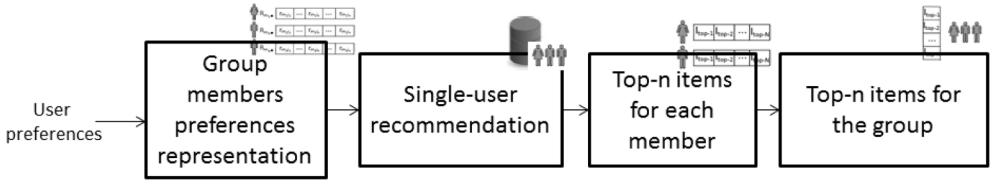

Fig. 3. Group recommendation based on recommendation aggregation.

aware recommendation that provides fair services to a group of users who have different importance levels within their group, considering simulated data over a dataset of restaurants. Furthermore, Nguyen and Ricci (2018) present a system that allows users to iteratively express and revise their preferences during the decision making process and supports such conversational process with a chat-based interface. Even though, items are characterized by their features, they only manage the presence or absence of a feature during the conversational, constraint-based process. The evaluation is also carried out by implementing a user study in the tourism domain.

Proposals integrating content-based and collaborative filtering. Furthermore, some related studies consider the integration of collaborative filtering-based and content-based recommendation in the group scenario. Seko, Yagi, Motegi, and Muto (2011) consider genres to be an important component of group recommendation, however this is not the main goal of the study, as it is focused on the balance of power between members. It is also focused on a dataset gathered from the TV domain and evaluating with a precision metric. Pera and Ng (2013) focus on proposing a group recommender system for movies, based on content similarity and popularity. The group is represented by an aggregated model that merges individual user models. Such models are represented by the tags assigned to the movies by the users. A movie is considered to be a candidate movie to be recommended if each of the personal tags in the group is highly similar to a tag in the tag cloud of the movie. The evaluation is carried out with experiments using the HetRec dataset.

Kaššák et al. (2016) more recently provide a hybrid group recommendation approach that combines individual collaborative filtering and content-based recommendation, aggregating in both cases the individually recommended list to compose group recommendations, and eventually combining the output of both recommendation paradigms. They also develop offline experiments using Movielens, and online experiments with real data. Finally, Pujahari and Sisodia (2020) focus on the use of the preference relation-based matrix factorization technique to obtain the predicted preference of the users, and then use graph aggregation to aggregate the preferences of the group members. They suggest the incorporation of the item features in the matrix factorization, but do not evaluate them individually. Here the evaluation is carried out using Precision and NDCG over Movielens and Netflix datasets.

In summary, the overall analysis of the related research identifies a tendency to cover specific domains, such as the tourism and social network domain (Ardissono et al., 2003; Quijano-Sanchez et al., 2014; Khoshkangini et al., 2016; De Pessemier et al., 2015; Nguyen & Ricci, 2018), as such proposals contain specific particularities associated with the data covering such domains, and therefore lack generality.

The analysis also suggests that most of the research presented is focused on presenting hybrid systems composed of a content-based dimension in order to enhance performance. An exception could be made concerning the research developed by Pera and Ng (2013) and Kaššák et al. (2016), who are more focused on analyzing the role of content-based recommendation in the group scenario. However, they do not follow the common scheme of content-based recommendation (Adomavicius & Tuzhilin, 2005), and the experimentation is limited.

This lack of research focused on content-based recommendation for the group scenario, has also been highlighted by the most recent surveys on group recommendation (Dara, Chowdary, & Kumar, 2020; De Pessemier et al., 2014; Kompan & Bieliková, 2014), in which this research task has only been mentioned occasionally. For example, Felfernig et al. (2018) recently made a brief reference to the possible structure of a content-based group recommendation system, but without the necessary formalization and experimental study.

This previous analysis highlights the necessity of the development of the current research, which is focused on presenting a general taxonomy and further extensions for content-based group recommender systems.

3. Introducing a taxonomy for content-based group recommendation

Section 2 demonstrates that previous research focused on incorporating the content-based paradigm into the group recommendation scenario, has either not followed a common research direction or tend to be adjusted to specific domains such as tourism and social networks. Furthermore, the related literature has not identified many studies that directly attempt to introduce the traditional content-based recommendation paradigm (Pazzani & Billsus, 2007) into the group recommendation scenario (Kaššák et al., 2016; Pera & Ng, 2013). However, such studies do not follow the common scheme of content-based recommendation (Adomavicius & Tuzhilin, 2005), and globally present hybrid systems that are also composed of collaborative filtering components such as popularity calculation (Pera & Ng, 2013) and user similarity calculation (Kaššák et al., 2016).

On the other hand and according to Section 2.2, group recommender systems have been clearly clustered into two categories: GRSs based on rating aggregation and GRSs based on recommendation aggregation (De Pessemier et al., 2014). Such a categorization has led to the development of several proposals for GRSs derived from these schemes (Castro et al., 2018; Castro, Lu, Zhang, Dong, & Martínez, 2018; Castro et al., 2017; Castro et al., 2018; Dara et al., 2020), in contrast to the shortage of developments (Section 2.3) that are specifically focused on content-based group recommendation approaches.

Previous issues have encouraged the development of a taxonomy for content-based group recommendation approaches, which are derived from the general taxonomy for group recommendation presented in Section 2.2, and the few developments in this area (Kaššák et al., 2016; Pera & Ng, 2013). The aim of this taxonomy is to serve as a starting point for further research in content-based group recommendation.

More specifically, considering the few developments in this area analyzed in Section 2.3, the proposed taxonomy (Fig. 4) revises three alternative models:

1. **Content-based group recommendation supported by recommendation aggregation and individual ranking (CB-GRS-Rank)**, which is built on the proposals of De Pessemier et al. (2015), where the personal rankings of individual users are merged into group recommendation using a ranking aggregation technique.
2. **Content-based group recommendation supported by recommendation aggregation and user-item matching values (CB-GRS-Match)**, which is built on the proposals of Kaššák et al. (2016), where the authors generate recommendations for each group member, and aggregate such individual recommendation into the group recommendation list by combining the individual scores. Furthermore, De Pessemier et al. (2015) also consider the aggregation of the

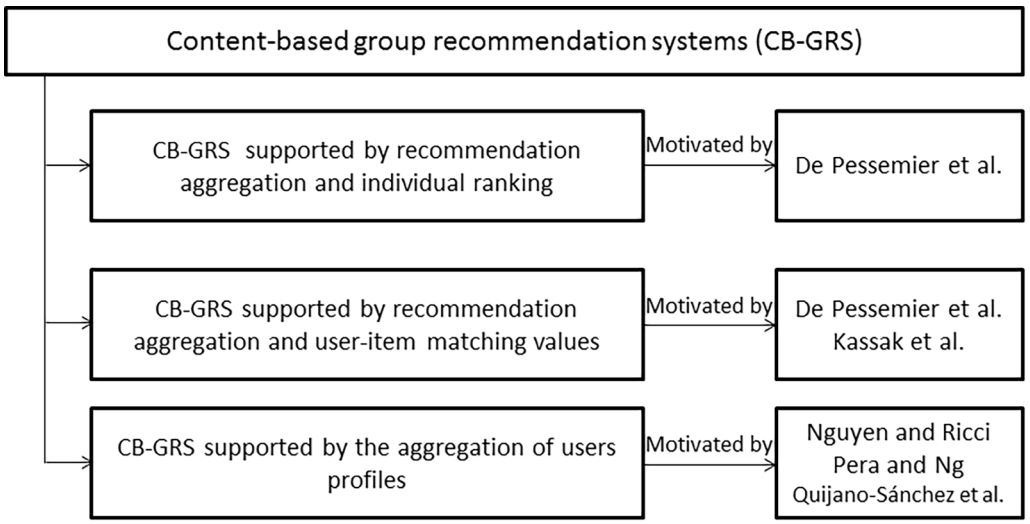

Fig. 4. Taxonomy for content-based group recommendation.

member's rating prediction in the users group, as a way to generate group recommendation.

3. **Content-based group recommendation supported by the aggregation of user profiles (CB-GRS-AggProf)**, which is built on some proposals already referred in Section 2.3, such as Quijano-Sanchez et al. (2014), Nguyen and Ricci (2018), Pera and Ng (2013), and so on, which all share the common feature of building a group profile that aggregates the individual preferences of the users in the group. This group profile is then used as input for the group recommendation generation.

Such models will be further detailed in the following subsections and will use the notation showed in Table 1.

3.1. Content-based group recommendation supported by recommendation aggregation and individual ranking (CB-GRS-Rank)

First, we revise a model that follows the recommendation aggregation paradigm and is supported by ranking aggregation to finally obtain the top n items to be recommended to the group (see Fig. 5). As previously mentioned at the beginning of Section 3, this model takes ideas from the proposal presented by De Pessemier et al. (2015), for group recommendation using recommendation aggregation and individual ranking. This model consists of four phases: (1) Item modeling, (2) User modeling, (3) Single user recommendation, and (4) Ranking aggregation to obtain the top k items for the group.

Table 1
Relevant notation

Term	Meaning
u	User
i	Item
G	Group
f_k^i	Value of the feature k for item i
f_k^u	Value of the feature k for user u
f_k^G	Value of the feature k for group G
$V^k = \{v_1^k, v_2^k, v_3^k, \ldots, v_p^k\}$	Possible values of the feature k in the item profile, for multivalued features
v^{ki}	Value of the feature k in the item i, for multivalued features
v_p^{ku}	pth value of the feature k for the user u, for multivalued features
top_u	List of top n recommendations for user u
S_{ui}	Matching value between user u and item i
S_i^G	Matching value between group G and item i
I_{top-k}^u	Top k items recommended to user u
I_G^u	Top k items recommended to the group G

3.1.1. Item modeling

Similar to traditional content-based recommendation, this model initially considers the modelling of the items to be recommended. These items are modeled in terms of a feature space used for characterizing the items (Fig. 5). Two basic approaches to item profiling will be considered in this study:

1. A basic approach that considers a binary profile that contains 1 if the item has the corresponding feature, and 0 if the item does not contain it. Formally, items are represented as the vector $i = (f_1^i, f_2^i, \ldots, f_m^i)$, where $f_k^i = 1$ if the feature k is associated to the item i, and $f_k^i = 0$ otherwise.
2. A more sophisticated approach that considers multivalued features (Castro, Rodríguez, & Barranco, 2014). In this case, items are also represented as the vector $i = (f_1^i, f_2^i, \ldots, f_m^i)$, but here f_k^i is associated with nominal or numeric values, in a domain associated to the feature k (Castro et al., 2014).

3.1.2. User modeling

The user modeling is represented by the same feature space that was previously used for item modeling. Two primary approaches can be considered to profile users in this study:

1. An approach based on TF-IDF (Aizawa, 2003), considering the preferred items. This approach assumes a binary item profile, and here users are represented by a vector $u = (f_1^u, f_2^u, \ldots, f_m^u)$. f_k^u is defined as:

$$f_k^u = FF(u,k) * IUF(k) \tag{3}$$

where $FF(u,k)$ is calculated as the number of items preferred by the user u, having $f_k^i = 1$. On the other hand, $IUF(k)$ is calculated as $IUF(k) = log\frac{|U|}{UF(k)}$, $UF(k)$ being the number of users that have preferred any item with the feature k, and $|U|$ the total number of users.
2. An approach that considers multivalued features (Castro et al., 2014), assuming the presence of nominal or numeric values in the item features, and therefore a new formulation of the user profile is necessary (Eq. 4).

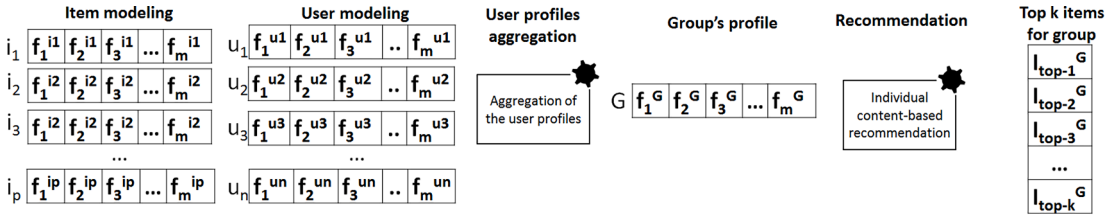

Fig. 5. Content-based group recommendation supported by aggregation of recommendations and individual ranking.

$$f_k^u = \begin{cases} \left\{(v_1^k, fr_{v_1^{ku}}), (v_2^k, fr_{v_2^{ku}}), (v_3^k, fr_{v_3^{ku}}), \ldots, (v_p^k, fr_{v_p^{ku}})\right\}, & \text{if } k \text{ is qualitative} \\ average(f_k^i) \text{for each item } i \text{ preferred by } u, & \text{if } k \text{ is quantitative} \end{cases} \tag{4}$$

In the case of qualitative features, f_k^u is formalized as a set of pairs (value, frequency) composed of each of the possible values v_p^k of the feature k, and the frequency $fr_{v_p^{ku}}$ of such value in the feature k in all the items preferred by the user u.

On the other hand, in the case of quantitative features, f_k^u will be the average of all the values associated with the corresponding feature, in all the items preferred by the user u.

3.1.3. Individual content-based recommendation

The model discussed in this section subsequently uses an individual content-based recommendation (Adomavicius & Tuzhilin, 2005; Bobadilla et al., 2013).

In this context, the former individual content-based recommendation approach will be used (Adomavicius & Tuzhilin, 2005), which is composed of the following steps:

1. For the active user, calculate the similarity degree between its profile and the profile of all the available items.
2. Sort the available items according to such degree in descending order.
3. Retrieve the top k items in the sorted list as recommendation.

The similarity degree will be calculated according to Eqs. (6) and (7), presented below.

3.1.4. Ranking aggregation to obtain the top k items for the group

CB-GRS-Rank depends on a ranking aggregation approach to aggregating the top k recommendation lists suggested to each individual member of the group.

In the context of this study focused on the screening of the three previously mentioned recommendation approaches, the Borda count (Marchant, 2001) will be used. It is a very popular social choice method that is actively used by the research community (Grandi, Loreggia, Rossi, & Saraswat, 2016; Abdrabbah, Ayadi, Ayachi, & Amor, 2017; Yera, Labella, Castro, & Martínez, 2018; Carballo-Cruz, Yera, Carballo-Ramos, & Betancourt, 2019).

As pos_i^u refers to the position of the item i in the top_u list with the individual recommendations provided to the user u, the Borda count assigns a weight a_i to each item that depends on the sum of the positions of such item in all the recommendation lists.

$$a_i = \sum_u (pos_i^u) \tag{5}$$

Finally, the aggregated ranking is composed by sorting all the recommended items in ascending order, according to their a_i values. This is the final recommendation list.

Other approaches to ranking aggregation can be integrated in this step (Baltrunas, Makcinskas, & Ricci, 2010). However, they are out of the scope of this paper.

3.2. Content-based group recommendation supported by recommendation aggregation and user-item matching values (CB-GRS-Match)

As an alternative to the previous model, this subsection revises a model that is based on the user-item matching values instead of the final ranking, and is also supported by the recommendation aggregation paradigm (Fig. 6). The criteria behind this model have been previously covered by Pera and Ng (2013) and De Pessemier et al. (2015), already described in Section 2.3 and at the beginning of Section 3.

This framework is composed of four stages: (1) Item modeling, (2) User modeling, (3) User-item matching value calculation, and (4) Matching value aggregation to obtain the top k items for the group.

3.2.1. Item and user modeling

In this model, the item and user modeling will be developed in a similar way to that previously presented in Sections 3.1.1 and 3.1.2.

3.2.2. User-item matching value calculation

CB-GRS-Match requires the value that indicates the matching degree between the corresponding user and item profile to be calculated (Castro et al., 2014). This matching value calculation is closely related to the nature of the data in the user and item profile. Therefore, it is necessary to formalize the matching value calculation in binary item profiles, and in item profiles with multivalued features.

1. In the case of binary item profiles, the cosine similarity function will be used directly between the user and item profiles u and i (Eq. 6). This similarity function has been widely used in previous research in recommendation systems (Adomavicius & Tuzhilin, 2005).

$$S_{ui} = \frac{\sum_{u,i} f_k^u * f_k^i}{\sqrt{(f_k^u)^2}\sqrt{(f_k^i)^2}} \tag{6}$$

2. In the case of the items with multivalued features, at first it is necessary to define the matching value between users and items, according to each individual feature k (Eq. 7). Specifically, for qualitative features, this value is calculated as fr_{v^k}, v being the associated key in the list of pairs in f_k^u, as well as the values in f_k^i. For quantitative features, this value is calculated as the inverse of the difference between f_k^u and f_k^i.

$$S_{ui}^{k*} = \begin{cases} fr_{v^k}, & \text{for } k \text{ qualitative} \\ \dfrac{1}{|f_k^u - f_k^i|}, & \text{for } k \text{ quantitative} \end{cases} \tag{7}$$

Furthermore, these values are normalized independently for qualitative and quantitative cases, finally reaching the matching values S_{ui}^{k*} according to each individual feature, which will be used in the next step.

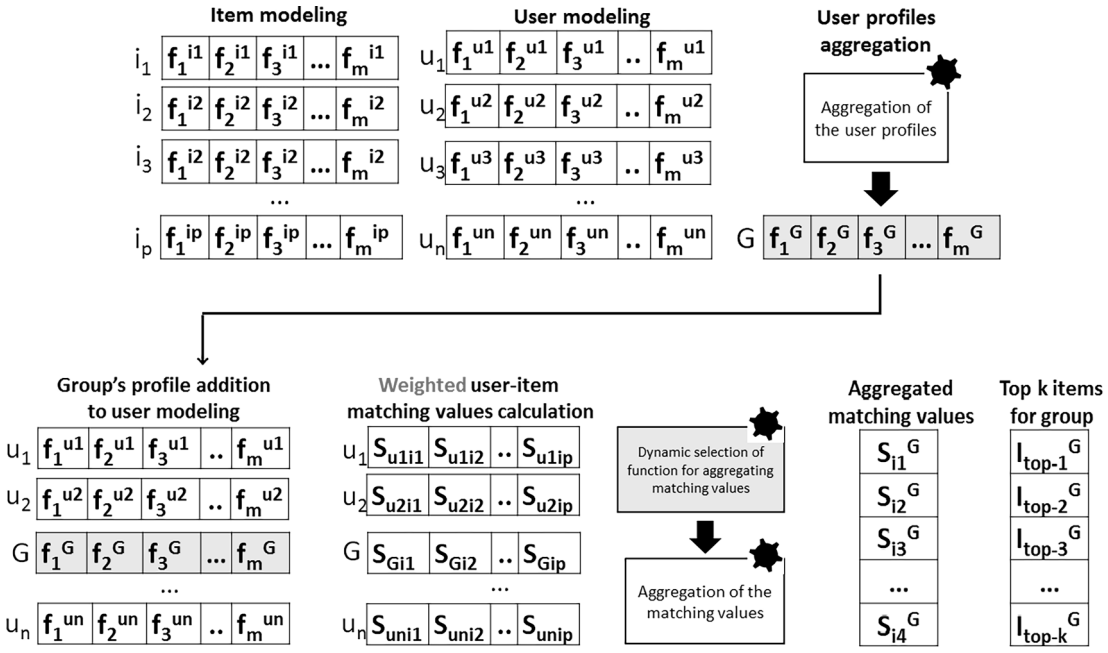

Fig. 6. Content-based group recommendation supported by aggregation of recommendations and user-item matching values.

3.2.3. Matching value aggregation to obtain the top k items for the group

CB-GRS-Match also requires the definition of an aggregation function to calculate the matching values associated with all users in the group, and each one of the items that can be recommended. To reach this goal, three aggregation schemes that are commonly used in group recommendation will be considered (De Pessemier et al., 2014).

1. Average: The average matching value for all the users in the group, n being the number of users.

$$S_i^G = \frac{\sum_u S_{ui}}{n} \tag{8}$$

2. Minimum: The aggregated matching value, as the matching value of the user with the lowest matching value in the group.

$$S_i^G = Min_u\ S_{ui} \tag{9}$$

3. Maximum: The aggregated matching value, as the matching value of the user with the highest matching value in the group.

$$S_i^G = Max_u\ S_{ui} \tag{10}$$

These aggregated matching values are used to generate the list of recommended items to the group, by sorting such values in descending order and retrieving their associated top k items.

3.3. Content-based group recommendation supported by the aggregation of user profiles (CB-GRS-AggProf)

In this subsection we revise a model based on rating aggregation for content-based group recommendation (Fig. 7). Here, once the item and user profiles are built, all the user profiles in the group are aggregated into a pseudo-user profile that represents the preference of the group. This profile is then used for recommendation generation by using an individual content-based recommendation approach. The criteria behind this model have been previously covered by some studies that were previously described in Section 2.3 and at the beginning of Section 3, such as Quijano-Sanchez et al. (2014), Nguyen and Ricci (2018), Pera and Ng (2013), and so on. Therefore, this framework is composed of four phases (Fig. 7): (1) Item modeling, (2) User modeling, (3) User profile aggregation, and (4) Individual content-based recommendation to obtain the top k items for the group.

3.3.1. Item and user modeling

Similarly, the item and user modeling in this framework will be developed according to the former approach presented in Sections 3.1.1 and 3.1.2.

3.3.2. User profile aggregation

CB-GRS-AggProf (Section 3.3), requires the definition of an aggregation approach to combine all the user profiles in the group into the pseudo-user profile $G = \{f_1^G, f_2^G, f_3^G, \ldots, f_m^G, \}$ that represents the group profile. Here three aggregation schemes that are commonly used in group recommendation will be also considered (De Pessemier et al., 2014), combined with the two user profiling approaches mentioned in

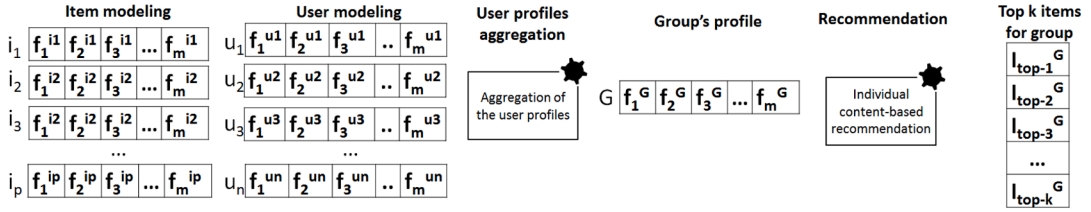

Fig. 7. Content-based group recommendation supported by the aggregation of user profiles.

Section 3.1.2.

1. For the approach based on TF-IDF (Section 3.1.2), the minimum aggregation (Eq. 11), the maximum aggregation (Eq. 12), and the average between all the feature values in the group (Eq. 13) will be considered to build each feature f_k^G in the group profile. Here n is the number of users in the group.

$$f_k^G = Min_u\, f_k^u \tag{11}$$

$$f_k^G = Max_u\, f_k^u \tag{12}$$

$$f_k^G = \frac{\sum_u f_k^u}{n} \tag{13}$$

2. For the approach that considers multivalued features, the aggregation of the user profiles will depend on the nature of each feature (quantitative or qualitative).

   Here the minimum aggregation scheme will be considered according to (Eq. 14), as a combination of the profiles formulated at Eq. (4). The maximum and the average aggregation scheme are formulated in the same way (Eqs. 15 and 16).

$$f_k^G = \begin{cases} \{(v_1^k, Min(fr_{v_1^{ku}})), (v_2^k, Min(fr_{v_1^{ku}})), (v_3^k, Min(fr_{v_1^{ku}})), \ldots, (v_p^k, Min(fr_{v_p^{ku}}))\}, & \text{for each user } u, k \text{ qualitative} \\ Min(average(f_{ku}^i)) & \text{for each item } i \text{ preferred by each } u, \text{ if } k \text{ is quantitative} \end{cases} \tag{14}$$

$$f_k^G = \begin{cases} \{(v_1^k, Max(fr_{v_1^{ku}})), (v_2^k, Max(fr_{v_1^{ku}})), (v_3^k, Max(fr_{v_1^{ku}})), \ldots, (v_p^k, Max(fr_{v_p^{ku}}))\}, & \text{for each user } u, k \text{ qualitative} \\ Max(average(f_{ku}^i)) & \text{for each item } i \text{ preferred by each } u, \text{ if } k \text{ is quantitative} \end{cases} \tag{15}$$

$$f_k^G = \begin{cases} \{(v_1^k, Avg(fr_{v_1^{ku}})), (v_2^k, Avg(fr_{v_1^{ku}})), (v_3^k, Avg(fr_{v_1^{ku}})), \ldots, (v_p^k, Avg(fr_{v_p^{ku}}))\}, & \text{for each user } u, k \text{ qualitative} \\ Avg(average(f_{ku}^i)) & \text{for each item } i \text{ preferred by each } u, \text{ if } k \text{ is quantitative} \end{cases} \tag{16}$$

3.3.3. Individual content-based recommendation for obtaining the top k items for the group

The last stage of this framework is the application of the individual content-based recommendation approach to the aggregated user profile obtained in the previous stage. To this end, the content-based recommendation approach previously presented in Section 3.1.3 is used.

The output of this approach is the final recommendation list that is retrieved as the output of this framework.

4. Further improvements in CB-GRSs beyond the state of the art

The previous section presented a taxonomy for content-based group recommender systems, based on the recent literature regarding group recommender systems, as well as a few previous studies concerning content-based group recommendation (refer to Sections 2.2 and 2.3). This taxonomy could serve as a basis for further developments in content-based group recommender systems.

With the aim of proposing more sophisticated content-based group recommendation approaches, this section is focused on presenting a novel content-based group recommendation model (CB-GRS-Hyb) supported by the hybridization of the CB-GRS-Match (Section 3.2) and the CB-GRS-AggProf (Section 3.3).

Fig. 8 shows the general scheme of this new proposal. First, the user and item modeling are performed in a similar way to the three former taxonomy models presented previously in Section 3. Afterwards, we aggregate the individual profiles of all the users in the group, to obtain an aggregated pseudo-user profile similarly to CB-GRS based on the aggregation of user profiles (Section 3.3). Furthermore, this newly obtained pseudo-user profile is added to the group's set of user profiles, and is considered to be a new group member. Subsequently, for each group member including the new pseudo-user, a weighted matching value with all the available items is calculated, built using the matching value calculation presented in Section 3.2.2. Once such matching values are calculated, a dynamic selection of the function for aggregating the calculated matching values for each available item is applied to the aggregation schemes presented in Section 3.2.3. After that, the selected aggregation function is used to reach a score that represents the aggregated matching value of each available item, which is used to finally obtain the top k for the group.

In the following subsections, these steps are explained in further detail.

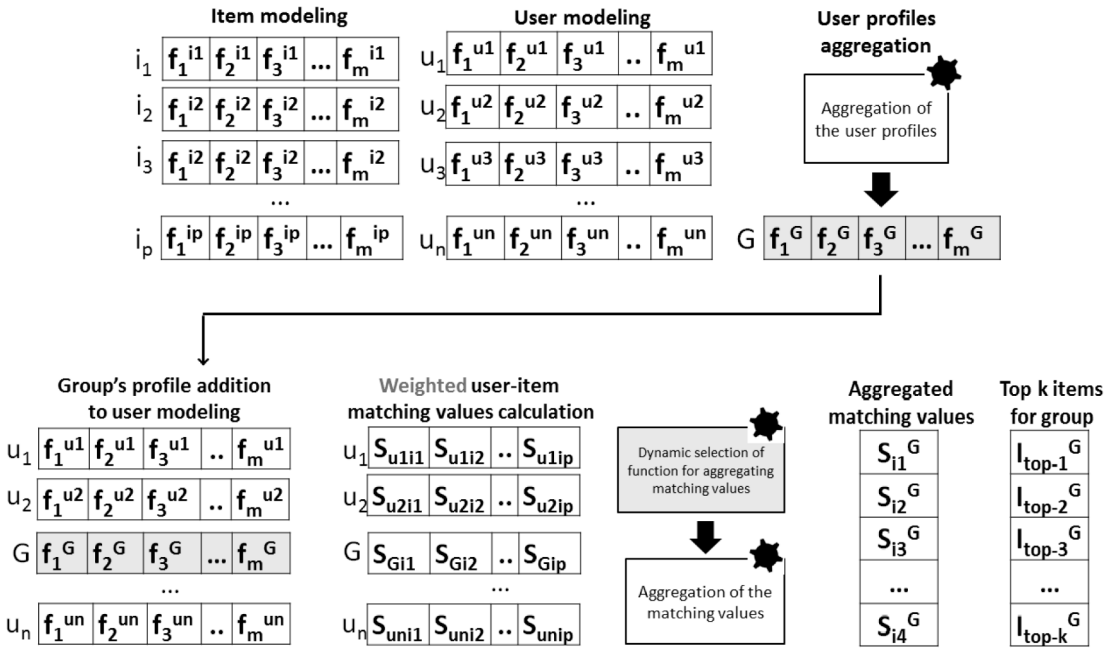

Fig. 8. The new hybrid content-based group recommendation method.

4.1. User and item modeling

This model uses the user and item profiling approaches presented previously in Section 3.1. Here we also consider the application of the binary or the multivalued feature representation scenario, depending on the context in which the method is applied.

4.2. User profiles aggregation

Once the user and item modeling has been developed, this model aggregates the individual profiles of all the users in the group in a similar way to CB-GRS based on the aggregation of the user profiles (Section 3.3), in order to create a pseudo-user profile that globally represents the preference of the group.

The aim of this new profile creation is to boost the clear preferences of the group, attenuating those that are unclear according to the possible user profile aggregation strategies already discussed in Section 3.3.2. Here, the average aggregation strategy will be considered for both the binary and multivalued feature scenario, because it tends to take into account all the users' preferences fairly, and avoids loss of information.

4.3. Group's profile addition to user modeling

The direct incorporation of pseudo-user or pseudo-item profiles into the group modeling has been previously addressed by the research literature (Domingues, Jorge, & Soares, 2013; Kagita, Pujari, & Padmanabhan, 2013; Kagita, Pujari, & Padmanabhan, 2015). The purpose of this strategy is to globally represent the group's preferences as well as the overall users' preferences, using these profiles to identify valuable information which finally leads to recommendation improvement (Domingues et al., 2013; Kagita et al., 2015).

At this stage, the pseudo-user profile created in the previous step is added to the list of users in the group, and is considered to be an individual user. Therefore, it is treated as if it were a standard user in the following steps of the recommendation process.

As it is assumed that this new profile boosts the clear preferences of the groups, this step considers that this boosting leads to better group preference representation, and therefore implies an improvement in recommendation performance.

4.4. Weighted user-item matching value calculation

Considering the user and item profiles, this section explores a weighted user-item matching value calculation, built on the matching value calculation schemes presented in Section 3.2.2.

Specifically, the use of feature weighting as a way to give more or less influence to some specific features in the final recommendation generation is explored. Therefore, the weighting value of a feature can be interpreted as the importance of such feature in the recommendation process. Some authors have previously considered the use of feature weighting in individual content-based recommendation (Castro et al., 2014; Cataltepe, Uluyağmur, & Tayfur, 2016).

In this research, the weighting scheme proposed by Castro et al. (2014), will be followed, defining the weight w_k^u for the feature k according to the user u, as:

$$w_k^u = DC(u,k) \tag{17}$$

where $DC(u,k)$ is the dependence coefficient between the ratings provided by the user u over a set of items and the values of the feature k for such items. It is formalized as:

$$DC(u,k) = \begin{cases} |PCC_{uk}|, & \text{if } k \text{ is quantitative} \\ VC_{uk}, & \text{if } k \text{ is qualitative} \end{cases} \tag{18}$$

Here, PCC_{uk} is the Pearson correlation coefficient according to the variables R_u and f_k^i, with n_u being the amount of ratings considered in this calculation:

$$PCC_{uk} = \frac{\sum_i r_{ui} v^{ki} - \frac{\sum_i r_{ui} \sum_i v^{ki}}{n_u}}{\sqrt{\sum_i (r_{ui}^2) - \frac{(\sum_i r_{ui})^2}{n_u}} \sqrt{\sum_i ((v^{ki})^2) - \frac{(\sum_i v^{ki})^2}{n_u}}} \tag{19}$$

Subsequently, VC_{uk} is the Cramer V contingency coefficient according to the same variables for qualitative features:

$$VC_{uk} = \sqrt{\frac{\sum_{k_u}\sum_{k_k \in V^k} \frac{(fr_{k_u,k_k} - \frac{fr_{k_u} fr_{k_k}}{n_u})^2}{\frac{fr_{k_u} fr_{k_k}}{n_u}}}{n_u min(|D_u|, |D_k|)}} \tag{20}$$

where k_u and $k_k \in V^k$ are the set of different values in r_{ui} and fr^i_k; fr_{k_u} and fr_{k_k} are the frequency of values indexed by k_u and k_k, and fr_{k_u,k_k} is the frequency of simultaneous co-occurrences of the two values indexed by k_u and k_k. $|D_u|$ and $|D_k|$ are the amount of different values associated with the rating r_u and the feature k. In this coefficient, frequency values equal to 0 are not considered.

Finally, the values obtained using Eq. (17) are finally normalized to reach the $(w^u_k)^*$ values.

Using these weighting values, the Eqs. (6) and (7) previously presented in Section 3.2.2 for the matching value calculation in the binary and multivalued profiles, are modified as follow to incorporate the calculated weights:

$$S_{ui} = \frac{\sum_{u,i} f^u_k * f^i_k * (w^u_k)^*}{\sqrt{(f^u_k)^2}\sqrt{(f^i_k)^2}} \tag{21}$$

$$S^{k*}_{ui} = \begin{cases} ((w^u_k)^*) * fr_{v^k}, & for\ k\ qualitative \\ (w^u_k)^* \frac{1}{|f^u_k - f^i_k|}, & for\ k\ quantitative \end{cases} \tag{22}$$

4.5. Aggregation of the matching values and final recommendation

Finally, this method requires the definition of an aggregation function to calculate the matching values associated with all users in the group, and each one of the items that are to be recommended. The former approach presented for CB-GRS, CB-GRS-Match (Section 3.2), analyzes the direct use of several aggregation approaches such as average, minimum, or maximum.

With this in mind, several authors have suggested that the proper aggregation approach in this context could depend on some group features such as group size and amount of rated movies (De Pessemier et al., 2014; Boratto, Carta, & Fenu, 2015). Based on such evidence, in this step we dynamically select the proper aggregation function according to the group characteristics.

Considering that a higher consensus degree is required where a higher amount of ratings are available, we have implemented this strategy using the *average* aggregation for those groups with a higher amount of ratings, and the *minimum* aggregation for groups with less ratings. These two aggregation strategies are selected because they perform better in relation to maximum aggregation (Section 6).

$$S^G_i = \begin{cases} \frac{\sum_u S_{ui}}{n}, if \sum_{u \in G} |R_u| \geqslant \alpha \\ Min_u\ S_{ui}, if \sum_{u \in G} |R_u| < \alpha \end{cases} \tag{23}$$

Eq. (23) formalizes this approach to aggregate the matching value for all the users in the group. With this in mind, the parameter α is defined, whose optimal value is experimentally reached in this paper. Where the global amount of ratings provided by the group is more than or equal to α, the average aggregation approach previously presented in Eq. (8) is used. However, if the global amount of ratings is under α, then the minimum aggregation approach previously presented in Eq. (9) is used. Possible future work could be focused on providing more sophisticated approaches to select the most appropriate aggregation strategies.

As in Section 3.2, the aggregated matching values are sorted in descending order and their associated top k items are retrieved to make a list of recommended items to the group.

Table 2
Comparison between the discussed approach.

	Advantages	Disadvantages
(1) CB-GRS-Rank	The ranking-based group recommendation allows for an easier understanding of the recommendation generation, considering it is closer to the real thinking.	The ranking generation introduces information loss that could affect accuracy.
(2) CB-GRS-Match	There is a less information loss than with the CB-GRS-Rank and CB-GRS-AggProf approaches, as it works with the numerical user-item matching values in its different stages.	It could be more difficult to understand the criteria for the final recommendation list based on the aggregation of individual recommendation.
(3) CB-GRS-AggProf	The recommendations are generated in a more direct way, considering it is involved only one profile: the unified group profile.	The nature of the method affects a possible consensus/refining of the recommendation list based on individual user preferences, considering they are formerly aggregated to compose the group profile. This generated group profile can also cause information loss, affecting accuracy.
(4) CB-GRS-Hyb	The use of the dynamic selection of the aggregation function introduces a more pertinent aggregation step tailored to the group's nature. The use of feature weighting gives importance to the more relevant features in the recommendation process.	A higher number of parameters for tuning.

5. Comparison between the discussed approaches

This section presents a comparison among the four approaches discussed previously in order to clarify their strengths and weaknesses. Such approaches are: (1) CB-GRS-Rank (supported by recommendation aggregation and individual ranking, Section 3.1), (2) CB-GRS-Match (supported by recommendation aggregation and user-item matching values, Section 3.2), (3) CB-GRS-AggProf (supported by the aggregation of user profiles, Section 3.3) and (4) CB-GRS-Hyb, the hybrid group recommendation approach proposed at Section 4.

Table 2 succinctly presents this comparison, which analyzes different criteria such as the possible effect of the preference aggregation or the recommendation aggregation steps in the final recommendation accuracy due to the aggregation-related information loss. Here, it should be pointed out that the preference aggregation (carried out in the CB-GRS-AggProf approach), can produce greater information loss, considering that it fuses the preferences of all the users into a unified group profile, in contrast to the CB-GRS-Rank and CB-GRS-Match approaches that first generate individual recommendations for each user. The CB-GRS-Rank approach, however, can also be affected by information loss in the ranking generation step (see Table 2).

It should also be noted, however, that in this comparison detailed in Table 2 we can observe despite the previously mentioned drawback of CB-GRS-Rank, that it also presents a relevant advantage as the ranking aggregation tends to be more transparent to users, in contrast to the other approaches that generate the recommendations solely based on numerical information (Loyola-González, 2019).

Another relevant issue covered by Table 2 is the fact that CB-GRS-AggProf only needs to generate recommendations for one profile (the group profile), while the other methods generate recommendations or at

least calculate the user-item matching values for each individual profile.

Finally, the comparison highlights the strengths of CB-GRS-Hyb as compared with the other approaches, by introducing feature weighting and dynamic aggregation function selection. However, it is also pointed out that these new steps use further parameters (such as α at Eq. (23)) that need to be adjusted, and whose values depend on the specific context in which the method is applied.

The goal of this section is to clarify strengths and weaknesses of each approach discussed. The following sections are focused on the experimental evaluation that complements this analysis.

6. Experiments

This section is focused on developing the experimental analysis of the content-based group recommendation frameworks presented previously. Specifically, it has three main goals: (1) to individually study the performances of the discussed frameworks and each design alternative, (2) to compare this framework with a collaborative filtering-based group recommendation framework, in order to measure the value of the current proposal regarding previous schemes, and (3) to evaluate the hybrid content-based group recommendation approach. Therefore, this section first presents details of the study, including datasets, evaluation metrics, and evaluation protocol. Subsequently, the performance of the proposals is presented and discussed. Then, the evaluation protocol and the results of the comparison with the aforementioned related research are both presented. Finally, the evaluation of the hybrid approach is presented. Final discussion and future challenges are also pointed out.

6.1. Datasets

This study uses two widely-used datasets in RS research, that, in addition, contain features that are directly associated to the available items. This is a mandatory requirement in this study in order to evaluate the content-based approaches. These datasets are:

- **Movielens 100 K,** composed of 100.000 ratings provided by 943 users for 1682 movies, in the 1 to 5 stars domain (Harper & Konstan, 2015). The genres of each movie are also available. Overall, the dataset assumes 19 genres, including Action, Sci-Fi, Comedy, Adventure, etc. Regarding such additional information, we represent the item profile as a binary vector composed of 19 dimensions, 1 if the corresponding genre is associated to the movie, and 0 otherwise. This dataset will be used to evaluate the approaches that consider the binary item profile.
- **HetRec,** which is an extension of the Movielens10M dataset, published by GroupLens research group (http://www.grouplens.org), which links the movies from the Movielens dataset with their corresponding web pages in the Internet Movie Database (IMDb) (http://www.imdb.com) and Rotten Tomatoes movie review systems (http://www.rottentomatoes.com) (Cantador, Brusilovsky, & Kuflik, 2011). The items in this dataset have associated several attributes such as year, genre, director, country, and the audience score. These five attributes are used to represent the item profile as a multivalued vector. The audience score is represented as a quantitative value. The genre, director, and country are treated as qualitative values. Even though the year is a number, here operations like average or mode do not make any sense, and therefore the year is represented as a qualitative value. Furthermore, in this case we have executed our evaluation with the first 300 users in the dataset.

6.2. Evaluation metrics

In order to evaluate the proposals, the two metrics considered have been previously used to evaluate the top n item recommendation tasks in group recommender systems (Baltrunas et al., 2010; Kaššák et al., 2016). Such metrics are Precision and NDCG, also commonly used in previously mentioned related research (see Section 2.3).

- **Precision (**Gunawardana & Shani, 2009**).** For each list of top n recommended items, Precision is defined as the ratio between the number of recommended items that were actually preferred by the current user, and the overall number of recommended items (in this case n).

$$Precision = \frac{|recommended\ items \cap preferred\ items|}{|recommended\ items|} \tag{24}$$

- **NDCG (**Jarvelin & Kekalainen, 2002**),** which depends on the Discounted Cumulative Gain (DCG) and is based on the premise that highly relevant items that appear at the end of a search result list should be penalized, as the graded relevance value is reduced logarithmically in proportion to the position of the result. DCG is formalized as:

$$DCG_u = \sum_{k=1}^{N} \frac{r_{u,recom_{u,k}}}{log_2(k+1)} \tag{25}$$

where $recom_{u,k} \in I$ is the item recommended to user u in k position.

To obtain NDCG, this DCG value should be normalized by dividing it by the maximum DCG value, $DCG_{perfect}$, that can be reached (Jarvelin & Kekalainen, 2002). $DCG_{perfect}$ is a perfect recommendation list, i.e., the most preferred items appear first on the list.

The NDCG values for each user are calculated as:

$$NDCG = \frac{DCG}{DCG_{perfect}} \tag{26}$$

Finally, such user-associated NDCG values are averaged to obtain the final NDCG reported value.

6.3. Experimental protocol

We have evaluated the proposal by implementing the following steps (Castro et al., 2017; Gunawardana & Shani, 2009):

- Taking as input the user profiles represented by the items that they prefer (i.e. evaluate) as well as the item features, we randomly split each user profile into two sets (i.e. training and test sets). The overall training and test sets are composed by combining the training and test set of each user.
- We build user groups of different sizes, following a group formation criterion that will be explained below.
- For each group, we apply each proposed framework to the users training data, obtaining the top n recommended items for the group.
- The accuracy of the top n recommendation list is evaluated using Precision and NDCG metrics, by comparing it with the items currently associated with each user in the test set. The average Precision and NDCG values for all groups are finally computed.

Table 3
CB-GRS-Rank, with group sizes {3, 4, 5} and top n {1,2,3,4,5,10,15,20}. Binary user profile. Precision metric

	3	4	5
1	**0.5483**	0.5354	0.5409
2	0.5397	0.54	**0.5416**
3	0.5411	0.5412	**0.5444**
4	0.5439	0.5436	**0.5455**
5	0.5454	0.5448	**0.5463**
10	0.5467	0.5462	**0.547**
15	0.5474	0.5472	**0.5478**
20	0.5482	0.548	**0.5486**

Table 4
CB-GRS-Rank, with group sizes {3, 4, 5} and top n {2, 3, 4, 5, 10, 15, 20}. Binary user profile. NDCG metric

	3	4	5
2	**0.9906**	0.9848	0.9807
3	**0.9757**	0.9723	0.9691
4	**0.9659**	0.9633	0.9607
5	**0.9584**	0.9565	0.9546
10	**0.9522**	0.95	0.9481
15	**0.9462**	0.9445	0.943
20	**0.9416**	0.9403	0.9391

To complete the Precision and NDCG calculation task, all the items in each user's test set are taken into account for recommendation generation, similarly to previous studies that were focused on the same recommendation task (Baltrunas et al., 2010; De Pessemier et al., 2014; Park, Kim, Oh, & Yu, 2016). In addition, in Precision calculation a preference threshold *pref* is used, that considers preferred items to be those that verify $r_{ui} \geq 4$, which is a common criterion for this threshold selection (Ricci et al., 2011).

Furthermore, several approaches have been considered for group composition in the literature. They include random group composition (Castro et al., 2017; Castro et al., 2018), as well as the use of some criteria that assure that groups' members have some characteristics in common (Baltrunas et al., 2010; Kaššák et al., 2016). Here, we use this second approach and more specifically, we follow the criteria referred to in Kaššák et al. (2016), that consider groups composed of individuals that have commonly rated the same set of items. Here, we consider item sets of size 3 located in the ratings test set. This size is selected because it is difficult to build larger groups, considering the nature of the datasets used. On the other hand, a size below 3 (e.g. size 2) does not represent a real matching between the users.

For each evaluation we build 20 groups that guarantee the fulfilment of this criteria (Kaššák et al., 2016). We consider groups with 3, 4, and 5 members. Furthermore, for each of these groups, we generate the top n recommendation list varying n in the range [1;5] with step 1 and in the range [5;20] with step 5. Such ranges assure the evaluation with both small and large recommendation lists. In the case of NDCG evaluation, we do not consider lists with size 1 because this measure is focused on evaluating the accuracy of the ordered list.

Overall, the protocol presented in this section is repeated 10 times, to avoid any bias in the evaluation. The 10 results are averaged, reporting such average as the evaluation value associated with the corresponding proposal.

6.4. Results

This section is focused on presenting and commenting on the results of the evaluation of the presented proposals. Considering that the datasets used have different characteristics, we have discussed their associated results in different sections.

6.4.1. Movielens

This subsection presents the results associated with Movielens 100 K, where binary item profile representation was used (see Section 6.1). Therefore, the proposals in this subsection will be implemented (Section 3) according to such item profile representation.

Main parameters study: Here, Tables 3,4 present the evaluation results of the content-based group recommendation model CB-GRS-Rank. Specifically, the tables present the values for group of 3, 4, and 5 (in columns), and sizes of the recommendation list with values 1, 2, 3, 4, 5, 10, 15, and 20 (in rows). For the Precision metric it is observed that the best performance tends to be obtained with more group members and a longer list of recommended items. However, the exception was reached for the top 1 recommendation and groups of 3, where in addition the best result from all the evaluations is obtained (0.5483). On the other hand, for NDCG the best results were obtained with shorter recommendation lists and small groups.

Furthermore, Tables 5,6 present the evaluation results of the content-based group recommendation model CB-GRS-Match. Here the sizes of groups and of the top n recommendation lists are evaluated similarly to the previous tables. Transversally, the three presented schemes for matching value aggregation (i.e. average, minimum, and maximum) are considered. In this evaluation, the best precision results were also obtained with larger group sizes and longer recommendation lists. However, in contrast to the approach focused on ranking aggregation, these precision values for the top 5 and longer recommendation lists tend to be constant and only vary slightly in several scenarios such as the average and the maximum aggregation. Overall, the best precision is associated with the average aggregation approach. In the NDCG case, the performance clearly decreases as the length of the recommendation list increases, and as with the precision measures, the best results are associated with the average aggregation.

Finally, Tables 7,8 present the results associated with CB-GRS-AggProf, using the same parameter values as the previous methods. These tables also used the three user profile aggregation approaches, average, minimum, and maximum, shown in Section 3.3.2. In contrast to the previous approach, the best results were obtained in several cases using a minimum aggregation approach, in this case the minimum approach for aggregating the user profiles. Such approach leads to Precision values over 0.5366 for recommendation lists longer or equal to 5, and in NDCG it leads to the best performance in all the scenarios. Overall, the better results tend to be obtained for larger groups in Precision, and smaller groups in NDCG.

Comparison across the proposals: Once the evaluation of each proposal is performed, including the study of their main parameters (i.e. the size of groups and of the recommendation list), a direct comparison of each proposal must be performed. In this comparison seven recently analyzed proposals will be considered. The captions of Figs. 9,10 detail such proposals according to Precision and NDCG. These figures compare the proposals considering a short recommendation list (1 for precision, 2 for NDCG), a medium-sized list (5), and a longer recommendation list (20) as well as considering the three groups sizes regarded in the experiments (i.e. 3, 4 and 5) (see caption in X axis). Overall, the figures show a clear

Table 5
CB-GRS-Match, with group sizes {3, 4, 5}, top n {1, 2, 3, 4, 5, 10, 15, 20}, and considering the three aggregation approaches avg, min, and max. Binary user profile. Precision metric

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
1	**0.5167**	0.5083	0.5133	**0.524**	0.5198	0.5067	**0.5336**	0.5229	0.5258
2	**0.5313**	0.5194	0.526	**0.532**	0.521	0.5234	**0.5365**	0.5278	0.5286
3	**0.5359**	0.5266	0.5291	**0.536**	0.5261	0.5289	**0.5378**	0.5301	0.5308
4	**0.5372**	0.53	0.5312	**0.5365**	0.53	0.5302	**0.538**	0.5326	0.5315
5	**0.5373**	0.5326	0.5319	**0.5363**	0.5323	0.5316	**0.5374**	0.5344	0.5326
10	**0.5374**	0.5346	0.5326	**0.5366**	0.5345	0.5324	**0.5372**	0.5357	0.5332
15	**0.5371**	0.536	0.5334	**0.5366**	0.5358	0.5331	**0.5372**	0.5367	0.5337
20	**0.5373**	0.5371	0.5339	0.5368	**0.5369**	0.5337	0.5372	**0.5376**	0.5341

Table 6
CB-GRS-Match, with group sizes {3,4,5}, top n {2,3,4,5,10,15,20}, and considering the three aggregation approaches avg, min, and max. Binary user profile. NDCG metric.

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
2	0.9908	**0.9911**	0.991	**0.9859**	0.9855	0.9851	**0.9822**	0.9813	0.9818
3	0.9769	0.9764	**0.977**	**0.9735**	0.9725	0.9728	**0.9705**	0.9694	0.9699
4	**0.967**	0.966	0.9667	**0.9643**	0.9632	0.9637	**0.962**	0.9607	0.9616
5	**0.9595**	0.9582	0.9591	**0.9575**	0.9561	0.9569	**0.9556**	0.9542	0.9551
10	**0.953**	0.9517	0.9527	**0.9509**	0.9495	0.9504	**0.9489**	0.9477	0.9486
15	**0.947**	0.9457	0.9467	**0.9453**	0.944	0.9448	**0.9437**	0.9425	0.9434
20	**0.9422**	0.941	0.9419	**0.9408**	0.9397	0.9405	**0.9396**	0.9385	0.9393

Table 7
CB-GRS-AggProf, with group sizes {3,4,5}, top n {1,2,3,4,5,10,15,20}, and considering the three aggregation approaches avg, min, and max. Binary user profile. Precision metric.

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
1	**0.505**	0.4933	0.495	**0.5144**	0.5117	0.5062	0.5259	**0.5261**	0.5192
2	**0.5265**	0.5217	0.5173	**0.53**	0.5253	0.5181	**0.5351**	0.5324	0.5235
3	**0.5341**	0.5309	0.5237	**0.5343**	0.5312	0.5241	**0.5367**	0.5356	0.5267
4	**0.5356**	0.5355	0.5272	0.5348	**0.5351**	0.527	0.5364	**0.5374**	0.5285
5	0.5355	**0.537**	0.529	0.5345	**0.5366**	0.5287	0.5357	**0.5381**	0.5299
10	0.5359	**0.5382**	0.5302	0.5351	**0.5379**	0.5302	0.5357	**0.5385**	0.531
15	0.5359	**0.5386**	0.5314	0.5353	**0.5384**	0.5313	0.5359	**0.5388**	0.532
20	0.5361	**0.539**	0.5323	0.5357	**0.5389**	0.532	0.536	**0.5391**	0.5325

Table 8
CB-GRS-AggProf, with group sizes {3,4,5}, top n {2,3,4,5,10,15,20}, and considering the three aggregation approaches avg, min, and max. Binary user profile. NDCG metric.

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
2	0.9909	**0.9914**	0.9906	0.9857	**0.9861**	0.985	0.9815	**0.9826**	0.9813
3	0.9766	**0.9774**	0.9761	0.9732	**0.9739**	0.9721	0.9702	**0.971**	0.9692
4	0.9668	**0.9675**	0.9656	0.9643	**0.9649**	0.9629	0.962	**0.9628**	0.9607
5	0.9595	**0.9602**	0.9581	0.9575	**0.9581**	0.956	0.9557	**0.9565**	0.9542
10	0.9531	**0.9538**	0.9516	0.951	**0.9516**	0.9494	0.9491	**0.9499**	0.9476
15	0.9471	**0.9479**	0.9456	0.9454	**0.9461**	0.9439	0.9438	**0.9446**	0.9424
20	0.9423	**0.9431**	0.9409	0.941	**0.9417**	0.9396	0.9398	**0.9406**	0.9384

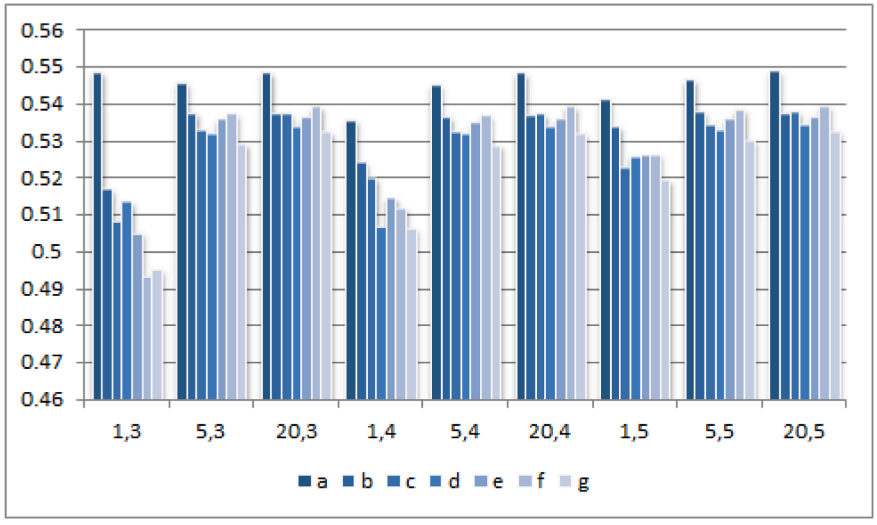

Fig. 9. Comparison between the proposals for 3, 4 and 5 group members, and 1, 5 and 20 top recommendations (Precision values and Movielens). (a) the CB-GRS based on recommendation aggregation and individual ranking, (b) the CB-GRS based on recommendation aggregation and user-item matching with average aggregation, (c) the same approach with minimum aggregation, (d) the same approach with maximum aggregation, (e) the CB-GRS supported by the aggregation of user profiles with average aggregation, (f) this approach with minimum aggregation, and (g) this approach with maximum aggregation.

correlation between the results associated with both measures across different group and recommendation list sizes.

In terms of precision, Fig. 9 shows the clear superiority of CB-GRS-Rank (a), which achieves the best performance in all scenarios. Furthermore, it is worth noting that CB-GRS-Match with average aggregation (b), also performs very well in all scenarios. However, for

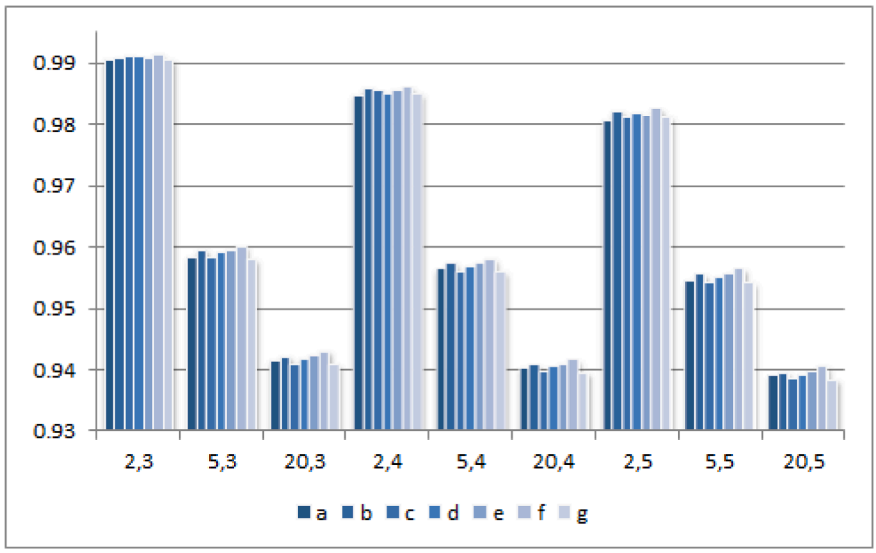

Fig. 10. Comparison between the proposals for 3, 4 and 5 group members, and 2, 5 and 20 top recommendations (NDCG values and Movielens). (a) the CB-GRS based on recommendation aggregation and individual ranking, (b) the CB-GRS based on recommendation aggregation and user-item matching with average aggregation, (c) the same approach with minimum aggregation, (d) the same approach with maximum aggregation, (e) the CB-GRS supported by the aggregation of user profiles with average aggregation, (f) this approach with minimum aggregation, and (g) this approach with maximum aggregation.

Table 9
Comparison against a collaborative filtering approach. Binary profile. Precision. pref = 3.

	3		4		5	
	CB-GRS	CF-GRS	CB-GRS	CF-GRS	CB-GRS	CF-GRS
1	**0.84**	0.8067	**0.8444**	0.8127	**0.8509**	0.8201
2	**0.8476**	0.8195	**0.8494**	0.8225	**0.8517**	0.8262
3	**0.8511**	0.8266	**0.8517**	0.8283	**0.852**	0.8306
4	**0.8513**	0.8312	**0.8515**	0.8328	**0.8518**	0.8344
5	**0.851**	0.8351	**0.8508**	0.836	**0.851**	0.8374
10	**0.8504**	0.8387	**0.8503**	0.8403	**0.8502**	0.842
15	**0.8496**	0.8433	**0.8494**	0.8447	**0.8492**	0.8462
20	**0.8486**	0.8473	**0.8485**	0.8485	0.8483	**0.8496**

Table 10
Comparison against a collaborative filtering approach. Binary profile. Precision. pref = 4.

	3		4		5	
	CB-GRS	CF-GRS	CB-GRS	CF-GRS	CB-GRS	CF-GRS
1	0.5583	**0.59**	0.5717	**0.5925**	0.5794	**0.594**
2	0.5765	**0.5963**	0.5789	**0.5982**	0.5812	**0.5992**
3	0.58	**0.6004**	0.582	**0.603**	0.5826	**0.6036**
4	0.5818	**0.6053**	0.5832	**0.6083**	0.5837	**0.6093**
5	0.5828	**0.6111**	0.5836	**0.6128**	0.5839	**0.6139**
10	0.5834	**0.6157**	0.5837	**0.6174**	0.5832	**0.6185**
15	0.5828	**0.6202**	0.5826	**0.6215**	0.5823	**0.6225**
20	0.5819	**0.6238**	0.5818	**0.6246**	0.5814	**0.6253**

Table 11
Comparison against a collaborative filtering approach. Binary profile. NDCG.

	3		4		5	
	CB-GRS	CF-GRS	CB-GRS	CF-GRS	CB-GRS	CF-GRS
2	**0.9915**	0.99	**0.9867**	0.9842	**0.9836**	0.9809
3	**0.9786**	0.9753	**0.9752**	0.9714	**0.973**	0.9687
4	**0.9696**	0.9648	**0.967**	0.9619	**0.9652**	0.9598
5	**0.9627**	0.9571	**0.9608**	0.9549	**0.9592**	0.9533
10	**0.9567**	0.9509	**0.9547**	0.9489	**0.953**	0.9473
15	**0.951**	0.9455	**0.9494**	0.9441	**0.948**	0.9429
20	**0.9464**	0.9416	**0.9452**	0.9406	**0.9442**	0.9397

longer recommendation lists, these results tend to be similar to the results achieved by other methods, such as CB-GRS-AggProf with minimum aggregation (f). Furthermore, both CB-GRS-Match with maximum aggregation, and CB-GRS-AggProf with maximum aggregation, achieve the poorest results in all cases.

For the NDCG metric, Fig. 10 shows that all the approaches tend to behave similarly, performing worse with longer recommendation lists and larger group sizes. However, it could be observed that CB-GRS-Match with average aggregation (b) and CB-GRS-AggProf with minimum aggregation (f) usually outperform the remaining approaches.

Comparison with the baseline: In order to measure the value of the content-based paradigm in the group recommendation scenario proposed in this paper, we have also compared it with a group recommender system based on collaborative filtering (Castro et al., 2017; De Pessemier et al., 2014). Here we will use a classical collaborative filtering approach, as we are also considering classical content-based approaches applied to the group scenario.

Several authors have shown that collaborative filtering is a successful paradigm for implementing recommender systems (Adomavicius & Tuzhilin, 2005; Bobadilla et al., 2013; Pilászy & Tikk, 2009). However, as previously mentioned in the Introduction section, the current research is derived from the fact that content-based recommendation can outperform collaborative filtering in very sparse scenarios, as it only depends on the information associated with the active user that requests the recommendation, and does not depend on rating co-occurrences across the users, as collaborative filtering does.

In order to verify this issue in group recommendation, in this section we have compared the presented methods with a collaborative filtering approach in a very sparse scenario. With this aim in mind, we have

Table 12
CB-GRS-Rank, with group sizes $\{3, 4, 5\}$, top n $\{1, 2, 3, 4, 5, 10, 15, 20\}$. Multi-valued profile. Precision metric.

	3	4	5
1	0.4717	0.4671	**0.4791**
2	0.478	0.4779	**0.4824**
3	0.4824	0.481	**0.4833**
4	0.4841	0.4828	**0.4849**
5	0.486	0.4849	**0.4864**
10	0.4866	0.4856	**0.4868**
15	**0.487**	0.4859	0.4868
20	**0.487**	0.4861	0.4868

Table 13
CB-GRS-Rank, with group sizes {3,4,5}, top n {2,3,4,5,10,15,20}. Multivalued profile. NDCG metric.

	3	4	5
2	**0.9917**	0.9875	0.9846
3	**0.9808**	0.978	0.9753
4	**0.9725**	0.9702	0.9683
5	**0.9661**	0.9645	0.963
10	**0.961**	0.9592	0.9576
15	**0.956**	0.9545	0.9532
20	**0.9521**	0.9509	0.9499

followed the same experimental protocol presented in Section 6.3, but have only considered three randomly selected ratings per user to build each user profile. In addition, it is verified that each item is evaluated by no more than five different users, thus addressing both the user cold-start and the item cold-start problem (Son, 2015). For each user, the remaining ratings are used as a test set in the evaluation. Finally, to develop a more in-depth assessment, $pref = 3$ and $pref = 4$ are taken as the preference threshold for the precision calculation.

We use a collaborative filtering-based group recommender based on recommendation aggregation as they have achieved the best results in previous studies (Castro et al., 2017). Furthermore, for individual prediction we use the Resnick's User K-nearest neighbors approach with Pearson similarity, which is the former approach in RS (Resnick, Iacovou, Suchak, Bergstrom, & Riedl, 1994), as we are also using former approaches in individual prediction (Adomavicius & Tuzhilin, 2005). We consider $k = 80$ for the amount of nearest neighbors and the average as aggregation scheme, having been selected for their optimal performance after several experiments.

Tables 9–11 present the comparison between a relevant approach in our proposal (i.e. CB-GRS-Match with average aggregation), tagged in the table as CB-GRS; and the mentioned collaborative filtering approach, tagged as CF-GRS. Table 9 shows the precision value for the preference threshold $pref = 3$, verifying that our current proposal is able to outperform collaborative filtering for all the experimental scenarios (with the exception of the top 20 recommendation task with 5 members). Specifically, the improvement percentage is higher for smaller recommendation lists, while for longer recommendation lists, CF-GRS obtains closely similar results to CB-GRS. The NDCG metric also achieves similar results, clearly outperforming CB-GRS to CF-GRS in the entire experimental setting. In contrast, for $pref = 4$ (Table 10), the CF-GRS method performed the best.

Overall, the results verify our initial assumption that CB-GRS is able to outperform CF-GRS in a highly sparse scenario, such as the scenario analyzed in this section. This is due to the fact that CB-GRS depends on those item features that are always available; while CF-GRS depends on a large amount of rating values being available, as well as their co-occurrence across several items and users, often unavailable to many users.

6.4.2. HetRec

This subsection presents the results associated with the HetRec dataset, in which the multivalued item profile representation was used (see Section 6.1). Therefore, the proposals will thus be implemented (Section 3) according to such item profile representation.

Main parameters study: Tables 12,13 present the evaluation results of the content-based group recommendation model CB-GRS-Rank. Here, the group and recommendation list sizes are the same as those used in the previous dataset. For the Precision metric, the better results tend to be obtained for groups with 5 members, and longer recommendation lists. On the other hand, the best NDCG is achieved in groups with 3 members. As observed in the previous dataset, a longer recommendation list will help achieve a lower NDCG value.

Furthermore, Tables 14,15 present the evaluation results of the content-based group recommendation model CB-GRS-Match, using the aforementioned protocol. As in the previous dataset, the three aforementioned approaches to matching value aggregation (i.e. average, minimum, and maximum) have been considered. Here, it is interesting to note that in contrast to previous evaluations, increasing the number of group members and items in the recommendation list usually leads to worse performance in Precision values. We think that this could be related to the presence of very dissimilar and possibly imprecise user-item matching values obtained in the proposal. In this scenario, such values, aggregated in larger groups and item recommendation lists, can lead to a more discrete performance as compared with smaller groups and shorter recommendation lists. Overall, in this case the best precision value was associated with the top 1 recommendation task, with 3 group members, in the minimum aggregation approach. Furthermore, regarding the Precision value, the minimum aggregation approach outperforms the average and the maximum aggregation in all evaluation scenarios. The best results for the NDCG evaluation criteria (Table 15) were obtained globally with small group sizes and short top n recommendation lists, and with the minimum aggregation scheme.

Finally, Tables 16,17 present the results associated with CB-GRS-AggProf, using the same parameter values as the previous methods and the three user profile aggregation approaches in Section 3.3.2. Here, the best precision and NDCG result was obtained with the average aggregation approach in most cases, achieving the best results with longer recommendation lists. We think that these findings are related to the fact that the minimum and the maximum approaches are more sensitive to imprecise data, which could affect larger groups and longer recommendation lists in this method based on the aggregation of the user profiles, leading therefore such aggregation approaches to a worse performance.

Comparison across the proposals: Figs. 11,12 show the comparisons made between the proposals, following the same methodology used in the other dataset (Section 6.4.1).

In this case the best precision values are obtained by CB-GRS-Match with minimum aggregation (c), and for larger groups CB-GRS-AggProf with average aggregation (e) also performed well in comparison with the other proposals. As was noted in the Movielens dataset, the approaches supported by the maximum aggregation scheme (d and g) also performed poorly. In contrast to such dataset, here CB-GRS-AggRank (a)

Table 14
CB-GRS-Match, with group sizes {3, 4, 5}, top n {1, 2, 3, 4, 5, 10, 15, 20}, and considering the three aggregation approaches avg, min, and max. Multivalued profile. Precision metric.

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
1	0.4583	**0.5267**	0.47	0.471	**0.4996**	0.4706	0.483	**0.5074**	0.4818
2	0.4819	**0.5116**	0.4834	0.4787	**0.5078**	0.4802	0.482	**0.5099**	0.4838
3	0.4832	**0.5103**	0.4842	0.4804	**0.5071**	0.4811	0.482	**0.5084**	0.4824
4	0.4828	**0.5083**	0.4827	0.4808	**0.506**	0.4811	0.4817	**0.5068**	0.482
5	0.482	**0.5064**	0.482	0.4804	**0.5044**	0.4807	0.481	**0.5049**	0.4811
10	0.4811	**0.504**	0.4809	0.4794	**0.5021**	0.4792	0.4799	**0.5023**	0.4796
15	0.48	**0.5015**	0.4795	0.4788	**0.5**	0.478	0.4793	**0.5003**	0.4786
20	0.4793	**0.4995**	0.4786	0.4782	**0.4982**	0.4776	0.4788	**0.4983**	0.4781

Table 15
CB-GRS-Match, with group sizes {3,4,5}, top n {2,3,4,5,10,15,20}, and considering the three aggregation approaches avg, min, and max. Multivalued profile. NDCG metric.

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
2	0.9918	**0.9923**	0.9918	**0.9879**	0.9876	0.9874	**0.9852**	0.9851	0.9848
3	0.9804	**0.9812**	0.9804	0.9776	**0.9778**	0.9773	0.9755	**0.9757**	0.9752
4	0.9724	**0.9732**	0.9724	0.9702	**0.9707**	0.97	0.9684	**0.9692**	0.9683
5	0.9663	**0.9674**	0.9662	0.9646	**0.9655**	0.9644	0.9633	**0.9642**	0.9631
10	0.9611	**0.9623**	0.9611	0.9593	**0.9604**	0.9592	0.9577	**0.959**	0.9576
15	0.956	**0.9575**	0.9559	0.9544	**0.9559**	0.9544	0.9531	**0.9547**	0.9531
20	0.9518	**0.9535**	0.9518	0.9506	**0.9523**	0.9506	0.9496	**0.9513**	0.9496

Table 16
CB-GRS-AggProf, with group sizes {3,4,5}, top n {1,2,3,4,5,10,15,20}, and considering the three aggregation approaches avg, min, and max. Multivalued profile. Precision metric.

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
1	0.4867	**0.4917**	0.46	**0.4833**	0.4646	0.455	**0.4939**	0.4747	0.462
2	**0.494**	0.4792	0.4632	**0.4905**	0.47	0.4617	**0.4948**	0.4731	0.466
3	**0.4939**	0.4747	0.4658	**0.493**	0.4702	0.4639	**0.4952**	0.4718	0.4665
4	**0.4959**	0.4724	0.4669	**0.4954**	0.4698	0.4652	**0.4974**	0.4709	0.4671
5	**0.4985**	0.4715	0.4672	**0.498**	0.4696	0.4659	**0.4995**	0.4704	0.4676
10	**0.4999**	0.4704	0.4686	**0.4998**	0.4694	0.4672	**0.5015**	0.47	0.4689
15	**0.5015**	0.4704	0.4696	**0.5012**	0.4694	0.4687	**0.5025**	0.4701	0.4701
20	**0.5023**	0.4702	0.4706	**0.5016**	0.4695	0.4699	**0.5024**	0.47	0.4709

Table 17
CB-GRS-AggProf, with group sizes {3,4,5}, top n {2,3,4,5,10,15,20}, and considering the three aggregation approaches avg, min, and max. Multivalued profile. NDCG metric.

	3			4			5		
	avg	min	max	avg	min	max	avg	min	max
2	0.9917	**0.9925**	0.9915	**0.9872**	0.9872	0.9869	**0.9843**	0.984	0.9841
3	0.9799	**0.9802**	0.9796	**0.9766**	0.9764	0.9764	**0.9745**	0.974	0.9742
4	**0.9716**	0.9715	0.9712	**0.9692**	0.9687	0.969	**0.9675**	0.9669	0.9673
5	**0.9652**	0.9651	0.9652	**0.9634**	0.963	0.9633	**0.9621**	0.9615	0.962
10	**0.96**	0.9596	0.9598	**0.9582**	0.9575	0.958	**0.9568**	0.956	0.9564
15	**0.9553**	0.9545	0.9547	**0.9538**	0.9528	0.9532	**0.9527**	0.9515	0.9519
20	**0.9515**	0.9504	0.9506	**0.9505**	0.949	0.9494	**0.9496**	0.948	0.9484

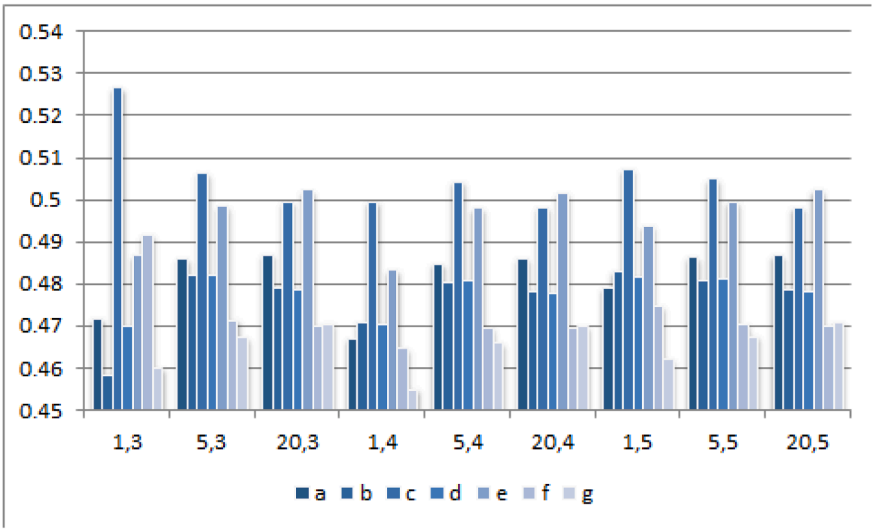

Fig. 11. Comparison between the proposals for 3, 4 and 5 group members, and 1, 5 and 20 top recommendations (Precision values and HetRec). (a) the CB-GRS based on recommendation aggregation and individual ranking, (b) the CB-GRS based on recommendation aggregation and user-item matching with average aggregation, (c) the same approach with minimum aggregation, (d) the same approach with maximum aggregation, (e) the CB-GRS supported by the aggregation of user profiles with average aggregation, (f) this approach with minimum aggregation, and (g) this approach with maximum aggregation.

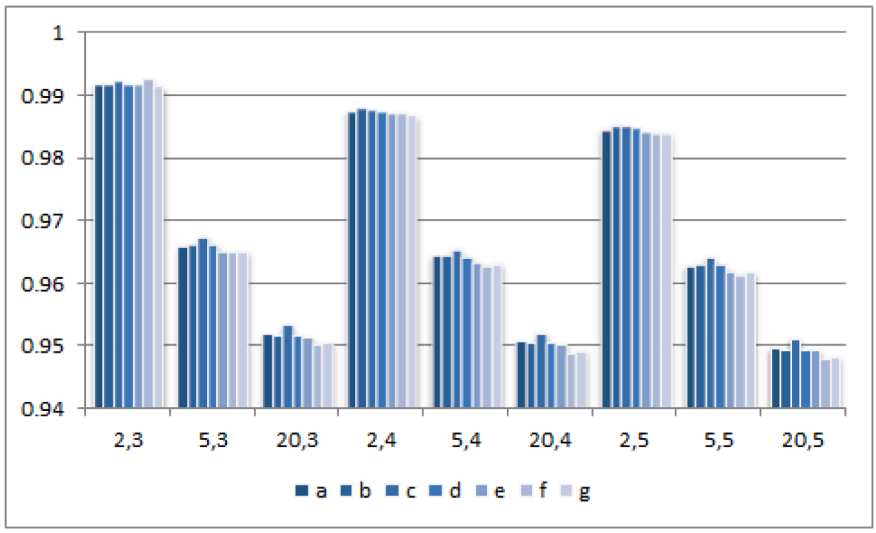

Fig. 12. Comparison between the proposals for 3, 4 and 5 group members, and 2, 5 and 20 top recommendations (NDCG values and HetRec). (a) the CB-GRS based on recommendation aggregation and individual ranking, (b) the CB-GRS based on recommendation aggregation and user-item matching with average aggregation, (c) the same approach with minimum aggregation, (d) the same approach with maximum aggregation, (e) the CB-GRS supported by the aggregation of user profiles with average aggregation, (f) this approach with minimum aggregation, and (g) this approach with maximum aggregation.

Table 18
Comparison against a collaborative filtering approach. Multivalued profile. Precision. pref = 3.

	3		4		5	
	CB-GRS	CF-GRS	CB-GRS	CF-GRS	CB-GRS	CF-GRS
1	**0.855**	0.7717	**0.8488**	0.779	**0.8502**	0.78
2	**0.8476**	0.7798	**0.8476**	0.7797	**0.8482**	0.7828
3	**0.8462**	0.7839	**0.8463**	0.7849	**0.8457**	0.7878
4	**0.8447**	0.7884	**0.8447**	0.7894	**0.8442**	0.7919
5	**0.8433**	0.7924	**0.8434**	0.7935	**0.8433**	0.7956
10	**0.8426**	0.7967	**0.8423**	0.7982	**0.8423**	0.8001
15	**0.8417**	0.8014	**0.8417**	0.8028	**0.8417**	0.8045
20	**0.8412**	0.8056	**0.841**	0.8067	**0.8409**	0.8083

Table 19
Comparison against a collaborative filtering approach. Multivalued profile. Precision. pref = 4.

	3		4		5	
	CB-GRS	CF-GRS	CB-GRS	CF-GRS	CB-GRS	CF-GRS
1	**0.5333**	0.4617	**0.5304**	0.4815	**0.5189**	0.4953
2	**0.5146**	0.489	**0.5177**	0.4904	**0.5151**	0.4954
3	**0.5132**	0.4945	**0.5151**	0.4956	**0.5127**	0.4998
4	**0.5112**	0.4993	**0.512**	0.5002	**0.5105**	0.5032
5	**0.5098**	0.5033	**0.5108**	0.5042	**0.5099**	0.507
10	**0.5101**	0.5077	**0.5107**	0.5096	0.5101	**0.5121**
15	0.5098	**0.5129**	0.5104	**0.5146**	0.5101	**0.5166**
20	0.5098	**0.5172**	0.5102	**0.5186**	0.51	**0.5204**

Table 20
Comparison against a collaborative filtering approach. Multivalued profile. NDCG.

	3		4		5	
	CB-GRS	CF-GRS	CB-GRS	CF-GRS	CB-GRS	CF-GRS
2	**0.993**	0.9911	**0.988**	0.9866	**0.9852**	0.9829
3	**0.9813**	0.9775	**0.978**	0.9742	**0.976**	0.9713
4	**0.9732**	0.9676	**0.9708**	0.9649	**0.9692**	0.9629
5	**0.9672**	0.9601	**0.9654**	0.9581	**0.964**	0.9566
10	**0.9619**	0.9541	**0.9601**	0.9521	**0.9585**	0.9506
15	**0.9569**	0.9487	**0.9554**	0.9472	**0.9541**	0.9461
20	**0.9528**	0.9446	**0.9517**	0.9435	**0.9506**	0.9426

also performed poorly in precision.

For the NDCG metric, Fig. 12 shows that all the approaches tend to behave similarly, with performance decreasing for longer recommendation lists and group sizes. However, it can be observed that CB-GRS-AggMatch with minimum aggregation (c) outperforms the other approaches in some of the scenarios.

Comparison with the baseline: For this dataset we also perform a comparison with the baseline represented by a GRS scenario with collaborative filtering, similar to that performed in Section 6.4.1. Here we have also used the CB-GRS supported by recommendation aggregation and user-item matching with minimum aggregation, as it performs better in this dataset (see Figs. 11,12).

Tables 18,19 show the precision results of this comparison. For the preference threshold *pref* = 3, the CB-GRS method outperforms CF-GRS in all the experimental settings. Furthermore, for *pref* = 4, CB-GRS also outperforms CF-GRS for $n \leqslant 10$. (with the exception of the top 10 recommendations with groups of 5). Finally, with NDCG our proposal obtains the best results in all the experimental scenarios (Table 20).

These results demonstrate the superiority of our proposals in relation to collaborative filtering for group recommendation in a very sparse scenario.

6.5. Evaluating the hybrid content-based group recommendation method

This subsection is focused on evaluating the hybrid content-based group recommendation method presented in Section 4.

Here we use the same experimental protocol presented in Section 6.3. The optimal value for the parameter α (Eq. 23) was empirically determined after several evaluations, reaching $\alpha = 200$ for Movielens, and $\alpha = 450$ for HetRec. Furthermore, the minimum scheme was used for the aggregation of the user profiles (see Fig. 8) because it reached the best result in the current experimental scenario.

Tables 21–24 present the results of this evaluation (hyb column), as well as its comparison against the average and the minimum aggregation approach presented in Section 3.2.3 (avg and min columns, respectively, in the Tables 21–24). The comparison between the average and the minimum aggregation approach is developed considering that the proposed hybrid approach directly includes components from the average and the minimum aggregation schemes associated with CB-GRS-Match, also proposed in this paper in Section 3.

Overall, the results suggest that, for almost all scenarios in the case of Movielens (i.e. the binary profile), and for short recommendation lists in

Table 21
Evaluation of the hybrid approach and comparison with former proposals, with group sizes $\{3, 4, 5\}$ and top n $\{1, 2, 3, 4, 5, 10, 15, 20\}$. Binary user profile. Precision metric.

	3			4			5		
	avg	min	hyb	avg	min	hyb	avg	min	hyb
1	0.5167	0.5083	**0.5183**	**0.524**	0.5198	0.5154	0.5336	0.5229	**0.5419**
2	0.5313	0.5194	**0.5398**	0.532	0.521	**0.5367**	0.5365	0.5278	**0.5453**
3	0.5359	0.5266	**0.5438**	0.536	0.5261	**0.5428**	0.5378	0.5301	**0.5474**
4	0.5372	0.53	**0.5459**	0.5365	0.53	**0.5446**	0.538	0.5326	**0.5474**
5	0.5373	0.5326	**0.5462**	0.5363	0.5323	**0.5448**	0.5374	0.5344	**0.5468**
10	0.5374	0.5346	**0.5462**	0.5366	0.5345	**0.5451**	0.5372	0.5357	**0.5458**
15	0.5371	0.536	**0.5456**	0.5366	0.5358	**0.5447**	0.5372	0.5367	**0.5452**
20	0.5373	0.5371	**0.545**	0.5368	0.5369	**0.5443**	0.5372	0.5376	**0.5449**

Table 22
Evaluation of the hybrid approach and comparison with former proposals, with group sizes $\{3, 4, 5\}$ and top n $\{2, 3, 4, 5, 10, 15, 20\}$. Binary user profile. NDCG metric

	3			4			5		
	avg	min	hyb	avg	min	hyb	avg	min	hyb
2	0.9908	**0.9911**	0.9906	**0.9859**	0.9855	0.9850	**0.9822**	0.9813	0.9818
3	**0.9769**	0.9764	0.9767	**0.9735**	0.9725	0.9727	**0.9705**	0.9694	0.9699
4	**0.967**	0.966	0.9667	**0.9643**	0.9632	0.9639	**0.962**	0.9607	0.9616
5	**0.9595**	0.9582	0.9593	**0.9575**	0.9561	0.9571	**0.9556**	0.9542	**0.9556**
10	**0.953**	0.9517	**0.953**	**0.9509**	0.9495	0.9507	0.9489	0.9477	**0.9490**
15	**0.9470**	0.9457	**0.9470**	**0.9453**	0.944	**0.9453**	0.9437	0.9425	**0.9438**
20	0.9422	0.941	**0.9423**	0.9408	0.9397	**0.9409**	0.9396	0.9385	**0.9398**

Table 23
Evaluation of the hybrid approach and comparison with former proposals, with group sizes $\{3, 4, 5\}$ and top n $\{1,2,3,4,5,10,15,20\}$. Multivalued user profile. Precision metric.

	3			4			5		
	avg	min	hyb	avg	min	hyb	avg	min	hyb
1	0.4583	0.5267	**0.5317**	0.471	0.4996	**0.5033**	0.483	0.5074	**0.5116**
2	0.4819	0.5116	**0.5118**	0.4787	**0.5078**	0.5041	0.482	**0.5099**	0.5056
3	0.4832	**0.5103**	0.5056	0.4804	**0.5071**	0.5016	0.482	**0.5084**	0.5028
4	0.4828	**0.5083**	0.5029	0.4808	**0.506**	0.5005	0.4817	**0.5068**	0.5012
5	0.482	**0.5064**	0.501	0.4804	**0.5044**	0.4989	0.481	**0.5049**	0.4995
10	0.4811	**0.504**	0.499	0.4794	**0.5021**	0.4967	0.4799	**0.5023**	0.4971
15	0.48	**0.5015**	0.4966	0.4788	**0.5**	0.4949	0.4793	**0.5003**	0.4951
20	0.4793	**0.4995**	0.4945	0.4782	**0.4982**	0.4932	0.4788	**0.4983**	0.4935

Table 24
Evaluation of the hybrid approach and comparison with former proposals, with group sizes $\{3, 4, 5\}$ and top n $\{2, 3, 4, 5, 10, 15, 20\}$. Multivalued user profile. NDCG metric.

	3			4			5		
	avg	min	hyb	avg	min	hyb	avg	min	hyb
2	0.9918	**0.9923**	**0.9923**	**0.9879**	0.9876	0.9875	**0.9852**	0.9851	0.9848
3	0.9804	**0.9812**	0.9805	0.9776	**0.9778**	0.9773	0.9755	**0.9757**	0.9750
4	0.9724	**0.9732**	0.9724	0.9702	**0.9707**	0.97	0.9684	**0.9692**	0.9682
5	0.9663	**0.9674**	0.9664	0.9646	**0.9655**	0.9645	0.9633	**0.9642**	0.9631
10	0.9611	**0.9623**	0.9612	0.9593	**0.9604**	0.9593	0.9577	**0.959**	0.9578
15	0.956	**0.9575**	0.9562	0.9544	**0.9559**	0.9547	0.9531	**0.9547**	0.9535
20	0.9518	**0.9535**	0.9523	0.9506	**0.9523**	0.9510	0.9496	**0.9513**	0.95

the case of HetRec (i.e. the multivalued profile), the hybrid approach performs the best results.

In the specific case of the Movielens dataset (the binary profile), the hybrid method obtains the best precision results in almost all scenarios, with the exception of the top 1 recommendation task with groups of 4, where the average aggregation obtains the best results. In the case of the NDCG metric, the three compared methods perform similarly, but the hybrid method obtains the best results or at least performs as well as the best method, when $n > 4$.

On the other hand, in HetRec the hybrid approach tends to obtain the best precision results for the top 1 and the top 2 recommendation tasks. However, for larger values of n, the minimum aggregation approach (also discussed in this contribution in Section 3.2.3), outperforms the hybrid approach. This analysis also applies for the NDCG results, which are equal to or close to the best results for short recommendation lists, but do not outperform the methods presented in Section 3.2.3 for longer lists. This poor performance of the hybrid approach with longer recommendation lists in HetRec, in relation to the previously presented methods (Section 3), might be related to the nature of the item features in such dataset. In fact, an initial analysis of its data reflects an important imbalance in some feature values (e.g. most of films are from United States of America, there are some years with too many films, and so on),

and that imbalance could affect the performance of the feature weighting process, which is a part of this hybrid method, and therefore, its accuracy. Feature engineering and new feature weighting schemes need to be further developed beyond those presented in Section 4.4, to try and address this issue. This is, however, out of the scope of this paper, which one of its objectives is the introduction of the hybrid content-based group recommendation scheme (Section 4) as well as proving that it could be effective in real scenarios. It will be then considered for future research.

Summarizing, the results show that this hybrid CB-GRS approach can be used as starting point for further developments in the creation of new CB-GRS approaches.

6.6. Final discussion

The global analysis of the evaluation results of the content-based group recommendation approaches presented in this study has lead us to the following findings:

- The recommendation aggregation paradigm in content-based group recommendation usually guarantees high performance, as the results associated with such methods are among the best in all the evaluation scenarios. The analysis of Tables 3–8 and Tables 12–17 identifies that for the HetRec dataset (the multivalued profile case), the paradigms based on the recommendation aggregation get the best results in 39 of the 45 evaluation scenarios (considering 2 evaluation metrics, 3 group sizes, and 8 different sizes of the recommendation list, and excluding the top 1 recommendation task in the NDCG metric). In the case of the Movielens dataset (the binary profile), such paradigms always achieve the best results with the Precision metric. Finally, for Movielens in the NDCG case, both CB-GRS-Match with average aggregation, and CB-GRS-AggProf with minimum aggregation perform similarly.
- There is a trend for the proposed content-based group recommendation approaches to globally improve their precision performance for larger group sizes. Analyzing Tables 3–8 and Tables 12–17, in 77 of 112 experimental scenarios, considering this proposal in the precision case (2 datasets, 8 different recommendation list sizes, and 7 recommendation approaches), a tendency to increase the precision value for larger group sizes can be observed. Other authors have also reported an analogous performance in previous experimental evaluations (Kaššák et al., 2016). However, there also some scenarios, such as those supported by the minimum aggregation approach in the multivalued profile, which do not particularly match this tendency and thus require further analysis. It is also worth mentioning, however, that with the NDCG metric the associated values decrease for larger group sizes, in all the scenarios.
- The obtained results empirically suggest that it makes sense the proposal of two different approaches to implement the recommendation aggregation paradigm (i.e. based on individual rankings and based on user-item matching, see Section 3), because they have different performance values for each experimental scenario (Tables 3–6). In this context, it was initially expected that the approach based on individual rankings (Section 3.1) and based on user-item matching (Section 3.2) would get similar results, as both are based on recommendation aggregation. However, the experimental results show that both methods perform differently, and that in several cases the individual ranking-based approach leads to the better results. Such facts, combined with the strengths of this approach shown in Table 2, make it a good alternative to use in real scenarios.
- An improved CB-GRS approach that integrates feature weighting, a hybridization between CB-GRS based on recommendation aggregation and user-item matching and CB-GRS based on the aggregation of the user profiles, and a switching strategy between the average and minimum matching value aggregation (Section 4); achieves promising results by obtaining the best performance in several scenarios in comparison with other methods, previously presented in Section 3. Specifically, for the binary profile it achieves or is equal to the best results for 32 of 45 experimental scenarios (considering 2 evaluation metrics, 8 recommendation list sizes, and 3 group sizes, and excluding the top 1 recommendation task in the NDCG metric). However, for the multivalued profile dataset, it achieves the best results for shorter recommendation lists. The reasons for these results have already been discussed in Section 6.5.
- Overall, the proposed content-based group recommendation paradigm outperforms the collaborative filtering-based group recommendation framework in a top n recommendation task with sparse data in many scenarios, verifying the initial assumption that content-based recommendation could play a relevant role in group recommendation because it mainly depends on item features and does not depend on the rating co-occurrences across the users. Here, for each dataset 69 experimental scenarios were considered, composed of 3 evaluation metrics (precision for $pref = 3$, precision for $pref = 4$, and NDCG), 8 recommendation list sizes, and 3 group sizes, and the top 1 recommendation task in the NDCG metric was excluded. We can observe that in the Movielens dataset, Tables 9–11 in 44 of 69 experimental scenarios, and in the HetRec dataset Tables 18–20 in 62 of 69 experimental scenarios, the proposed paradigm outperforms a collaborative filtering-based group recommendation framework. Further discussion on these tables has already been presented in Sections 6.4.1 and 6.4.2.

6.7. Future challenges

The methods presented as well as the evaluation results obtained and detailed in this paper, open several research challenges for the near future. Some of these challenges are:

- The development of semantic-aware CB-GRSs, due to the fact that the integration of semantics has recently played a relevant role in content-based recommendation.To this end, ontological knowledge, including domain-specific ontologies, has been incorporated into several semantic-aware content-based RSs; an active use has been made of unstructured or semi-structured encyclopedic knowledge sources such as Wikipedia; and the use of the Linked Open Data cloud (de Gemmis et al., 2015) has also been explored.Some of these lines of research might be considered to be of interest in future in order to be able to further enrich the proposals presented in this paper.
- The use of machine learning algorithms to determine the best switching strategy among the matching functions and the aggregation strategies used in the proposals.Section 4.5 shows that the use of a simple switching strategy to select one of two aggregation strategies can lead to an improvement in the recommendation accuracy in several scenarios.Therefore more research in this area is required in order to develop it further, for example the training of machine learning models, such as decision trees (Rokach, 2007) and support vector machines (Scholkopf & Smola, 2002), in order to match users' characteristics and the more suitable aggregation functions.
- The use of matrix factorization methods for user and item profiling in CB-GRSs.Matrix factorization methods have been successfully used in several RSs scenarios (Koren & Bell, 2015), including content-based recommendation approaches (Lin, Kuo, & Lin, 2014; Nguyen & Zhu, 2013) and in group recommendation (Ortega et al., 2016). Therefore, matrix factorization methods, specifically focused on CB-GRSs, need to be developed.
- The evaluation of recommendation diversity in the proposed models. Recommendation diversity has recently become an important research topic (Aytekin & Karakaya, 2014).It is therefore necessary to evaluate recommendation diversity associated with the results of the proposed methods.Furthermore, new CB-GRS methods can be proposed to prioritize the diversity criteria.

- The proposal of approaches for natural noise management in the CB-GRS scenario (Martínez, Castro, & Yera, 2016;Yera, Barranco, Alzahrani, & Martínez, 2019). Natural noise management approaches have been successfully proposed for the GRS scenario (Castro et al., 2017; Castro et al., 2018). It is therefore necessary to define the concept of natural noise in the CB-GRS, as well as proposing approaches to mitigate it, in order to improve the recommendation performance.

7. Conclusions

This paper presents a general taxonomy for the development of content-based group recommender systems (CB-GRS). In this way, even though the development of group recommender systems is a problem that has been covered by several studies, the analysis of the related literature concludes that most of such studies use the collaborative filtering-based paradigm as the main approach when building the proposal. Therefore, due to the lack of rating values and rating co-occurrences as issues that affect collaborative filtering, they are quite limited in terms of performance when applied to highly sparse scenarios.

Taking this shortcoming into account, the current study explores a general taxonomy for CB-GRS, built on the traditional content-based recommendation paradigm. Specifically, three models have been discussed in this paper that can be used to build CB-GRSs, which are (1) CB-GRSs supported by recommendation aggregation and individual ranking (CB-GRS-Rank), (2) CB-GRSs supported by recommendation aggregation and user-item matching (CB-GRS-Match), and (3) CB-GRSs supported by the aggregation of user profiles (CB-GRS-AggProf). Furthermore, we have presented a further improvement to such models, by proposing a hybrid CB-GRS (CB-GRS-Hyb) that uses feature weighting, aggregation function switching, and the combination of two models previously discussed in Section 3.

In order to evaluate the proposals, an experimental protocol using well-known datasets has been developed. Such experimentation demonstrates that the recommendation aggregation paradigm performs the best in this scenario, and that the proposals tend to improve the precision performance with longer recommendation lists and larger groups. Furthermore, the proposals outperform a collaborative filtering-based group recommendation framework in the top n recommendation task with very sparse data. Finally, the hybrid proposal is able to outperform the formerly proposed models.

Considering a practical viewpoint for applications in real scenarios, the experimental results suggest that the CB-GRS-Hyb approach (Section 4) could be a good election across binary and multivalued user profiles, even though the more relevant results were reached for the binary case. Beyond such approach, experiments also evidence that CB-GRS-Rank (Section 3.1) obtains good results for the binary profile. On the other hand, CB-GRS-Match (Section 3.2), specifically with minimum aggregation, reaches good results for the sparser, multivalued user profile.

The current study could serve as a starting point for future research in the area of content-based group recommender systems, ideally addressing the challenges previously pointed out in Section 6.7. Our future research in this area will be focused on: (1) exploring the use of machine learning for the proper aggregation functions selection (Rokach, 2007), (2) developing CB-GRS models based on matrix factorization techniques (Koren & Bell, 2015; Ortega et al., 2016), and (3) exploring natural noise management in the CB-GRS scenario (Castro et al., 2018; Martínez et al., 2016).

CRediT authorship contribution statement

Yilena Pérez-Almaguer: Software, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Visualization. **Raciel Yera:** Software, Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Writing - review & editing. **Ahmad A. Alzahrani:** Conceptualization, Methodology, Supervision, Project administration, Funding acquisition. **Luis Martínez:** Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Writing - review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under Grant FP-122-42.

References

Abdrabbah, S. B., Ayadi, M., Ayachi, R., & Amor, N. B. (2017). Aggregating top-k lists in group recommendation using borda rule. In *Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems* (pp. 325–334). Springer.

Adomavicius, G., & Tuzhilin, A. T. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering, 17*, 734–749.

Aizawa, A. (2003). An information-theoretic perspective of tf-idf measures. *Information Processing and Management, 39*, 45–65.

Ardissono, L., Goy, A., Petrone, G., Segnan, M., & Torasso, P. (2003). Intrigue: personalized recommendation of tourist attractions for desktop and hand held devices. *Applied Artificial Intelligence, 17*, 687–714.

Aytekin, T., & Karakaya, M. (2014). Clustering-based diversity improvement in top-n recommendation. *Journal of Intelligent Information Systems, 42*, 1–18.

Baltrunas, L., Makcinskas, T., & Ricci, F. (2010). Group recommendations with rank aggregation and collaborative filtering. In *Proceedings of the 4th ACM Conference on Recommender Systems RecSys '10* (pp. 119–126). New York, NY, USA: ACM.

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems, 46*, 109–132.

Boratto, L., Carta, S., & Fenu, G. (2015). Discovery and representation of the preferences of automatically detected groups: Exploiting the link between group modeling and clustering. *Future Generation Computer Systems, 24*, 833–848.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modelling and User-Adapted Interaction, 12*, 331–370.

Cantador, I., Brusilovsky, P., & Kuflik, T. (2011). 2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011). In *Proceedings of the 5th ACM conference on Recommender systems RecSys 2011*. New York, NY, USA: ACM.

Carballo-Cruz, E., Yera, R., Carballo-Ramos, E., & Betancourt, M. E. (2019). An intelligent system for sequencing product innovation activities in hotels. *IEEE Latin America Transactions, 17*, 305–315.

Castro, J., Barranco, M. J., Rodríguez, R. M., & Martínez, L. (2018). Group recommendations based on hesitant fuzzy sets. *International Journal of Intelligent Systems, 33*, 2058–2077.

Castro, J., Lu, J., Zhang, G., Dong, Y., & Martínez, L. (2018). Opinion dynamics-based group recommender systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48*, 2394–2406.

Castro, J., Quesada, F. J., Palomares, I., & Martínez, L. (2015). A consensus-driven group recommender system. *International Journal of Intelligent Systems, 30*, 887–906.

Castro, J., Rodríguez, R. M., & Barranco, M. J. (2014). Weighting of features in content-based filtering with entropy and dependence measures. *International Journal of Computational Intelligence Systems, 7*, 80–89.

Castro, J., Yera, R., Alzahrani, A. A., Sanchez, P., Barranco, M., & Martínez, L. (2019). A big data semantic driven context aware recommendation method for question-answer items. *IEEE Access, 7*, 182664–182678.

Castro, J., Yera, R., & Martínez, L. (2017). An empirical study of natural noise management in group recommendation systems. *Decision Support Systems, 94*, 1–11.

Castro, J., Yera, R., & Martínez, L. (2018). A fuzzy approach for natural noise management in group recommender systems. *Expert Systems with Applications, 94*, 237–249.

Cataltepe, Z., Uluyağmur, M., & Tayfur, E. (2016). Feature selection for movie recommendation. *Turkish Journal of Electrical Engineering & Computer Sciences, 24*, 833–848.

Dara, S., Chowdary, C., & Kumar, C. (2020). A survey on group recommender systems. *Journal of Intelligent Information Systems, 54*, 271–295.

De Pessemier, T., Dhondt, J., Vanhecke, K., & Martens, L. (2015). Travelwithfriends: a hybrid group recommender system for travel destinations. In *Workshop on tourism recommender systems (touRS15) in conjunction with the 9th ACM conference on recommender systems (recsys 2015)* (pp. 51–60).

De Pessemier, T., Dooms, S., & Martens, L. (2014). Comparison of group recommendation algorithms. *Multimedia Tools and Applications, 72*, 2497–2541.

Domingues, M. A., Jorge, A. M., & Soares, C. (2013). Dimensions as virtual items: Improving the predictive ability of top-n recommender systems. *Information Processing & Management, 49*, 698–720.

Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends in Human-Computer Interaction, 4*, 81–173.
Felfernig, A., Boratto, L., Stettinger, M., & Tkalčič, M. (2018). *Group recommender systems: An introduction*. Springer.
Gedikli, F., & Jannach, D. (2013). Improving recommendation accuracy based on item-specific tag preferences. *ACM Transactions on Intelligent Systems and Technology, 4*, 11.
de Gemmis, M., Lops, P., Musto, C., Narducci, F., & Semeraro, G. (2015). Semantics-aware content-based recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 119–159). US: Springer.
Ghazarian, S., & Nematbakhsh, M. A. (2015). Enhancing memory-based collaborative filtering for group recommender systems. *Expert Systems with Applications, 42*, 3801–3812.
Grandi, U., Loreggia, A., Rossi, F., & Saraswat, V. (2016). A borda count for collective sentiment analysis. *Annals of Mathematics and Artificial Intelligence, 77*, 281–302.
Gunawardana, A., & Shani, G. (2009). A Survey of Accuracy Evaluation Metrics of Recommendation Tasks. *Journal of Machine Learning Research, 10*, 2935–2962.
Gunawardana, A., & Shani, G. (2015). Evaluating recommender systems. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 265–308). US: Springer.
Harper, F. M., & Konstan, J. A. (2015). The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems, 5*, 19:1–19:19.
Jarvelin, K., & Kekalainen, J. (2002). Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems, 20*, 422–446.
Kagita, V., Pujari, A., & Padmanabhan, V. (2013). Group recommender systems: a virtual user approach based on precedence mining. In *Australasian Conference on Artificial Intelligence* (pp. 434–440).
Kagita, V. R., Pujari, A. K., & Padmanabhan, V. (2015). Virtual user approach for group recommender systems using precedence relations. *Information Sciences, 294*, 15–30.
Kaminskas, M., & Bridge, D. (2016). Diversity, serendipity, novelty, and coverage: A survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Transactions on Interactive Intelligent Systems, 7*, 2:1–2:42.
Kaššák, O., Kompan, M., & Bieliková, M. (2016). Personalized hybrid recommendation for group of users: Top-n multimedia recommender. *Information Processing & Management, 52*, 459–477.
Khoshkangini, R., Pini, M. S., & Rossi, F. (2016). A self-adaptive context-aware group recommender system. In *Conference of the Italian Association for Artificial Intelligence* (pp. 250–265). Springer.
Kirshenbaum, E., Forman, G., & Dugan, M. (2012). A live comparison of methods for personalized article recommendation at forbes.com. In *Machine Learning and Knowledge Discovery in Databases* (pp. 51–61). Springer.
Kompan, M., & Bieliková, M. (2014). Group recommendations:survey and perspectives. *Computing and Informatics, 33*, 1–13.
Koren, Y., & Bell, R. (2015). Advances in collaborative filtering. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 77–118). Boston, MA: Springer, US.
Lin, C.-J., Kuo, T.-T., & Lin, S.-D. (2014). A content-based matrix factorization model for recipe recommendation. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining* (pp. 560–571). Springer.
Lops, P., Gemmis, M., & Semeraro, G. (2011). Content-based recommender systems: State of the art and trends. In Recommender Systems Handbook chapter 3. (pp. 73–105). Springer, US.
Lops, P., Jannach, D., Musto, C., Bogers, T., & Koolen, M. (2019). Trends in content-based recommendation. *User Modeling and User-Adapted Interaction, 29*, 239–249.
Loyola-González, O. (2019). Black-box vs white-box: understanding their advantages and weaknesses from a practical point of view. *IEEE Access, 7*, 154096–154113.
Lu, J., Wu, D., Mao, M., Wang, W., & Zhang, G. (2015). Recommender system application developments: A survey. *Decision Support Systems, 74*, 12–32.
Mahyar, H., Ghalebi K, E., Morshedi, S.M., Khalili, S., Grosu, R., & Movaghar, A. (2017). Centrality-based group formation in group recommender systems. In Proceedings of the 26th International Conference on World Wide Web Companion (pp. 1187–1196). International World Wide Web Conferences Steering Committee.
Marchant, T. (2001). The borda rule and pareto stability: a further comment. *Fuzzy Sets and Systems, 120*, 423–428.
Martínez, L., Castro, J., & Yera, R. (2016). Managing natural noise in recommender systems. In C. Martín-Vide, T. Mizuki, & M. A. Vega-Rodríguez (Eds.), *Theory and Practice of Natural Computing 5th International Conference, TPNC 2016* (pp. 3–17). Springer International Publishing.
Movahedian, H., & Khayyambashi, M. R. (2014). Folksonomy-based user interest and disinterest profiling for improved recommendations: An ontological approach. *Journal of Information Science, 40*, 594–610.
Nguyen, J., & Zhu, M. (2013). Content-boosted matrix factorization techniques for recommender systems. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 6*, 286–301.
Nguyen, T. N., & Ricci, F. (2018). A chat-based group recommender system for tourism. *Information Technology & Tourism, 18*, 5–28.
Nilashi, M., bin Ibrahim, O., Ithnin, N., & Sarmin, N.H. (2015). A multi-criteria collaborative filtering recommender system for the tourism domain using expectation maximization (em) and pca–anfis. Electronic Commerce Research and Applications, 14, 542–562.
Ortega, F., Hernando, A., Bobadilla, J., & Kang, J. H. (2016). Recommending items to group of users using matrix factorization based collaborative filtering. *Information Sciences, 345*, 313–324.
Park, C., Kim, D., Oh, J., & Yu, H. (2016). Using user trust network to improve top-k recommendation. *Information Science, 374*, 100–114.
Pazzani, M., & Billsus, D. (2007). Content-Based Recommendation Systems. *The Adaptive Web, 4321*, 325–341.
Pazzani, M. J. (1999). A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review, 13*, 393–408.
Pera, M., & Ng, Y. (2013). A group recommender for movies based on content similarity and popularity. *Information Processing and Management, 49*, 673–687.
Pera, M., & Ng, Y.K. (2014). Automating readers advisory to make book recommendations for k-12 readers. In Proceedings of the ACM Conference on Recommender Systems RecSys '14 (pp. 9–16). ACM.
Pilászy, I., & Tikk, D. (2009). Recommending new movies: even a few ratings are more valuable than metadata. In Proceedings of the third ACM conference on Recommender systems (pp. 93–100). ACM.
Pujahari, A., & Sisodia, D.S. (2020). Aggregation of preference relations to enhance the ranking quality of collaborative filtering based group recommender system. Expert Systems with Applications, (p. 113476).
Quijano-Sanchez, L., Recio-Garcia, J. A., & Diaz-Agudo, B. (2014). An architecture and functional description to integrate social behaviour knowledge into group recommender systems. *Applied Intelligence, 40*, 732–748.
Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work* (pp. 175–186). New York, USA: ACM.
Ricci, F., Rokach, L., & Shapira, B. (2011). Recommender systems handbook. In F. Ricci, L. Rokach, B. Shapira, & P. B. Kantor (Eds.), *Recommender Systems Handbook, chapter 1* pp. 1–35). Springer.
Rokach, L. (2007). *Data Mining with Decision Trees: Theory and Applications*. World Scientific.
Scholkopf, B., & Smola, A. (2002). *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT Press.
Seko, S., Yagi, T., Motegi, M., & Muto, S. (2011). Group recommendation using feature space representing behavioral tendency and power balance among members. In Proceedings of the fifth ACM conference on Recommender systems (pp. 101–108). ACM.
Seo, Y.-D., Kim, Y.-G., Lee, E., Seol, K.-S., & Baik, D.-K. (2018). An enhanced aggregation method considering deviations for a group recommendation. *Expert Systems with Applications, 93*, 299–312.
Son, L. H. (2015). Hu-fcf++: A novel hybrid method for the new user cold-start problem in recommender systems. *Engineering Applications of Artificial Intelligence, 41*, 207–222.
Wang, J., Jiang, Y., Sun, J., Liu, Y., & Liu, X. (2018). Group recommendation based on a bidirectional tensor factorization model. *World Wide Web, 21*, 961–984.
Wang, W., Zhang, G., & Lu, J. (2016). Member contribution-based group recommender system. *Decision Support Systems, 87*, 80–93.
Yera, R., Alzahrani, A. A., & Martínez, L. (2019). A food recommender system considering nutritional information and user preferences. *IEEE Access, 7*, 96695–96711.
Yera, R., Barranco, M., Alzahrani, A. A., & Martínez, L. (2019). Exploring fuzzy rating regularities for managing natural noise in collaborative recommendation. *International Journal of Computational Intelligence Systems, 12*, 1382–1392.
Yera, R., Caballero Mota, Y., & Martínez, L. (2018). A recommender system for programming online judges using fuzzy information modeling. *Informatics, 5*, 17.
Yera, R., Labella, Á., Castro, J., & Martínez, L. (2018). On group recommendation supported by a minimum cost consensus model. In Data Science and Knowledge Engineering for Sensing Decision Support: Proceedings of the 13th International FLINS Conference (FLINS 2018) (p. 227). World Scientific volume 11.
Yera, R., & Martínez, L. (2017). Fuzzy tools in recommender systems: A survey. *International Journal of Computational Intelligence Systems, 10*, 776–803.
Yera, R., & Martínez, L. (2017). A recommendation approach for programming online judges supported by data preprocessing techniques. *Applied Intelligence, 47*, 277–290.
Yuan, Q., Cong, G., & Lin, C.-Y. (2011). Com: a generative model for group recommendation. In Proceedings of ACM International Conference on Knowledge Discovery and Data Mining (KDD 2014) (pp. 163–172).
Zhang, Y. (2016). Grorec: a group-centric intelligent recommender system integrating social, mobile and big data technologies. *IEEE Transactions on Services Computing, 9*, 786–795.