Future Generation Computer Systems 64 (2016) 165–174

Contents lists available at ScienceDirect

Future Generation Computer Systems

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

Discovery and representation of the preferences of automatically detected groups: Exploiting the link between group modeling and clustering

Ludovico Boratto *, Salvatore Carta, Gianni Fenu

Dipartimento di Matematica e Informatica, Università di Cagliari, Via Ospedale 72 - 09124, Cagliari, Italy

HIGHLIGHTS

- We devise algorithms to model the preferences of groups automatically detected by clustering.
- We compare different modeling strategies for automatically detected groups.
- We find the most performing strategy for automatically detected groups.
- The group modeling strategy affects the accuracy of a group recommender system.
- We show that there is a link between clustering and group modeling.

ARTICLE INFO

Article history:
Received 20 March 2015
Received in revised form
4 October 2015
Accepted 14 October 2015
Available online 12 November 2015

Keywords:
Group modeling
Group recommendation
Clustering
Collaborative filtering

ABSTRACT

There are types of information systems, like those that produce group recommendations or a market segmentation, in which it is necessary to aggregate big amounts of data about a group of users in order to filter the data. *Group modeling* is the process that combines multiple user models into a single model that represents the knowledge available about the preferences of the users in a group. In group recommendation, group modeling allows a system to derive a group preference for each item. Different strategies lead to completely different group models, so the strategy used to model a group has to be evaluated in the domain in which the group recommender system operates. This paper evaluates group modeling strategies in a group recommendation scenario in which groups are detected by clustering the users. Once users are clustered and groups are formed, different strategies are tested, in order to find the one that allows a group recommender system to get the best accuracy. Experimental results show that the strategy used to build the group models strongly affects the performance of a group recommender system. An interesting property derived by our study is that clustering and group modeling are strongly connected. Indeed, the modeling strategy takes the same role that the centroid has when users are clustered, by producing group preferences that are equally distant from the preferences of every user. This "continuity" among the two tasks is essential in order to build accurate group recommendations.

1. Introduction

Information systems are widely used to collect and process data in order to support important business processes. In case of the information systems used to produce *market segmentations* [1,2] and *group recommendations* [3,4], the data related to large groups of users and millions of items are aggregated in order to produce suggestions for the users in terms of ads or items. It is essential for these systems to deal with big data and to be able to filter large amounts of information. In this paper, we focus on the knowledge discovery and representation of the preferences of the users in a group recommender system.

Group recommendation [3] is designed for contexts in which more than one person is involved in the recommendation process [4]. While the objective of a classic recommender system is to produce personalized content for users in the form of suggestion of items that the users might like [5], group recommender systems suggest items to a group, by combining individual models that contain a user's preferences [6]. In the last few years, group recommendation has been highlighted as a challenge in the recommendation research [4,7].

* Corresponding author.
E-mail addresses: ludovico.boratto@unica.it (L. Boratto), salvatore@unica.it (S. Carta), fenu@unica.it (G. Fenu).

http://dx.doi.org/10.1016/j.future.2015.10.007
0167-739X/

In order to derive a group preference for the items, *group modeling* strategies, which combine individual user models, are employed in the development of a group recommender system. J. Masthoff presented several studies [6,8–12] that adopt group modeling strategies in a scenario where a group of viewers has to select a sequence of items to watch. Pizzutilo et al. [13] highlighted that there is no strategy useful in every context independently from the environment, and that the choice of the strategy that best models a group should be made after a deep analysis of the context in which the group is modeled.

An interesting application domain in which group recommendation can be used, is when the number of recommendation lists that can be produced is limited (i.e., it is not possible to produce a list of recommended items for each user). The two examples below present scenarios in which there is such a constraint.

Application scenario 1. A company decides to print recommendation flyers that present suggested products. Even if the data to produce a flyer that contains individual recommendations for each customer is available, the process of printing a different flyer for everyone would be technically too hard to accomplish and costs would be too high. A possible solution would be to set a number of different flyers to print, such that the printing process could be affordable in terms of costs and the recipients of the same flyer would be interested by its content.

Application scenario 2. A political party is participating to an electoral campaign. The campaign advertising might involve the distribution of flyers. Since a campaign usually presents the program of the party for different issues (e.g., education, economy, health) and candidates, preparing a single flyer to present the purposes of the party might not be useful. Indeed, users would not be targeted and a single flyer to present all the aims of the party would certainly be confusing. Given the results of polls about parties and issues, a possible solution would be to set a number of different flyers to print that can be affordable in terms of costs. Each flyer could contain a subset of issues and could be addressed to voters interested in the same issues, in order to maximize the voters interest in the party.

With respect to a classic group recommender system, this type of systems adds the complexity of defining groups, in order to respect the constraint on the number of recommendation lists that can be produced and maximize users' satisfaction. The presence of a classification task inside the group recommendation process, which allows a system to detect the groups, is a novel aspect that is not present in the existing systems in the literature. Studying this research area becomes even more challenging with respect to classic group recommendation, because we deal both with the classification and recommendation aspects, in order to be able produce the recommendations.

This paper analyzes the problem of modeling the preferences of a group that has been detected by clustering users. Since no group recommendation approach in the literature works with automatically detected groups, this is the first time that group modeling is studied for groups created by a clustering algorithm. In this application scenario, groups are big because of the limited amount of recommendation lists that can be produced. In other words, for each item the preferences of a large set of users has to be accurately combined. Moreover, typically the knowledge base (e.g., an e-commerce website) involves millions of items.

This form of recommendation is strongly related to cyber–physical–socio intelligence [14]. Indeed, our approach is able to link the physical space (the real world), in which the users explicitly express preferences for items they had an experience with, to a cyber and socio space in which the interaction among users is made possible by computers and machine learning algorithms [15]. From the cyber–physical–socio intelligence point of view, the objective of our proposal is to find complex and previously unknown links among the users and automatically detect groups, in order to produce recommendations to them. This creates a so-called *knowledge grid* (i.e., a virtual socio grid), in which users enjoy the services (i.e., the group recommendations) made available to them thanks to these algorithms. Inferring and explaining the social existence leads to the definition of the so-called *social semantics* [16], which can also be employed to detect communities [17]. Focusing on the topic of this paper, i.e., a group modeling task that allows a system to infer the preferences of an automatically detected group for a set of items, it lends itself well to a processing through grid computing and to the concept of knowledge grid. Indeed, the individual preferences of each item might be aggregated on different computers (e.g., a possible optimized solution is to use a single computer for the computation of the group rating for a subset of items, so that the computation of the final group model is distributed over different computers, by employing large scale distributed computing models like MapReduce). Moreover, very recently Donohoe et al. [18] highlighted that context-awareness is a challenging research area in grid computing, because of the large amounts of data that might flow in the systems and the adaptivity aspects that characterize them. The approach we are proposing exactly reflects this trend and challenge in the literature, since we are dealing with adaptive group recommender systems that filter large amounts of data and, in particular, we focus on group modeling algorithms, whose computation can be split on multiple computers.

More specifically, the problem statement is the following:

Problem 1. We are given a set of users $U = \{u_1, u_2, \dots, u_n\}$, a set of items $I = \{i_1, i_2, \dots, i_m\}$, and a set V of values used to express the user preferences (e.g., $V = [1, 5]$ or $V = \{like, dislike\}$). The set of all possible ratings expressed by the users is a ternary relation $R \subseteq U \times I \times V$. We also consider a function $f : U \times I \to V$ that, for each element $(u, i, v) \notin R$, predicts a value $p_{ui} = f(u, i)$. Given a value k, which denotes the maximum number of recommendation lists that can be generated, we split the set of users U in a partition of k groups, so that for each group $g_q \subseteq U$ $(q \in \{1, \dots, k\})$ every user $u \in g_q$ receives the same recommendations.

This paper explores a set of strategies to combine into a model m_q the predicted ratings for each item i for the users in a group g_q. Each model m_q contains a score for each item, which predicts the interest of that group for the item by combining the individual predictions. *Our goal is to find the strategy that allows a system to create a model m_q that maximizes the accuracy of the group recommender system.*

Throughout this work, we are going to give an answer to a set of questions that arise when approaching the group modeling problem in a new scenario:

Question 1. Can all the existing group modeling strategies be applied to a context in which groups are detected by clustering users?
Question 2. Is there any strategy that works in some conditions, but cannot be used under some circumstances?
Question 3. Is there a strategy that works better than the others?
Question 4. Is there a significant difference in the models produced by the strategies?

The scientific contributions coming from this paper are now presented:

- We devise algorithms to implement the modeling strategies in a group recommendation context in which groups are automatically detected.
- By analyzing group modeling in a novel group recommendation context, this paper is the first to evaluate, compare, and analyze