# A Generic Framework for Collaborative Filtering Based on Social Collective Recommendation

Leschek Homann[1], Bernadetta Maleszka[2](✉), Denis Mayr Lima Martins[1], and Gottfried Vossen[1]

[1] ERCIS, University of Münster, Leonardo-Campus 3, 48149 Münster, Germany
{leschek.homann,denis.martins,vossen}@wi.uni-muenster.de

[2] Faculty of Computer Science and Management, Wroclaw University of Science and Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
Bernadetta.Maleszka@pwr.edu.pl

**Abstract.** Collaborative filtering has been considered the most used approach for recommender systems in both practice and research. Unfortunately, traditional collaborative filtering suffers from the so-called cold-start problem, which is the challenge to recommend items for an unknown user. In this paper, we introduce a generic framework for social collective recommendations targeting to support and complement traditional recommender systems to achieve better results. Our framework is composed of three modules, namely, a User Clustering module, a Representative module, and an Adaption module. The User Clustering module aims to find groups of users, the Representative module is responsible for determining a representative of each group, and the Adaption module handles new users and assigns them appropriately. By the composition of the framework, the cold-start problem is alleviated.

**Keywords:** Recommender system · Cold-start problem · User profile Social networks · Collaborative filtering

## 1 Introduction

Recommender systems are basically divided into content-based, knowledge-based, collaborative-based, and hybrid approaches [7]. Content-based recommender systems predict recommendations for a customer's interest in an item by considering its characteristics. Collaborative filtering approaches are either user-based or item-based. The former first identifies similar users by calculating their ratings for items rated by both of them and then considering their ratings to the item to predict. The latter first identifies similar items by their ratings and then by considering the users rating for those similar items. As the name suggests, hybrid systems aim to combine different approaches to improve the achieved recommendations. Although substantial research has been conducted on collaborative filtering conducted in recent years, the problem of generating suitable


N. T. Nguyen et al. (Eds.): ICCCI 2018, LNAI 11055, pp. 238–247, 2018.
https://doi.org/10.1007/978-3-319-98443-8_22

recommendations for an unknown user is still open. This paper presents a fresh approach based on a generic framework for social collective recommendations.

Current approaches try to tackle the cold start problem by considering social characteristics of a user as part of the recommendations. Nevertheless, these approaches are limited, as they basically only consider one aspect of the problem. Therefore, we separate the problem into three different modules: a Clustering module, a Representative module, and an Adaption module, which compose a generic framework to solve the cold-start problem. At the same time, we introduce our idea of how a social collective can be represented. The most obvious source for social collectives nowadays are social networks [4,6]. As an example, we propose to use a clique [11], known from graph theory, to describe a social collective, as it provides a very intuitive and easily understandable representation of groups. Furthermore, we introduce an evaluation concept for our generic framework.

The remainder of this paper is structured as follows: In Sect. 2 we describe related work to recommender systems. In Sect. 3 the problem definition is stated. Afterwards, we specify our framework and its modules in Sect. 4. Section 5 describes the evaluation idea of the framework. We conclude our paper in Sect. 6 by describing limitations and giving an outlook of future work.

## 2 Related Work

Recommender systems become more and more important in everyday life. When a user wants to find or buy an item, he or she usually first checks information about the product, e.g., in the Internet (content-based approach), asks some specialists about the quality of the product, or simply asks his or her friends for an opinion on the product (collaborative filtering approach). Information about similar people or tags is often used in tagged recommendation systems [16].

The increasing interest in recommender systems research is confirmed by the substantial literature produced in this area. Beel et al. [2] noticed that 55% of over 200 papers from the last few years are focusing on content-based approaches, while almost 35% are connected to collaborative filtering or graph-based recommendation. This demonstrates that content (description) of products, documents or items are really important, but also the opinion of different users about items should be taken into account. A similar result is presented by Anjumol et al. [1]. They claim that psychological and sociological studies have shown dependencies between individual preferences and interpersonal influences: Users' decisions and absorptions of information are affected by other users.

With growing interest in social networks, information about the community of a user can be obtained easily. Social recommender systems were introduced by Ma et al. [12]. A social network is considered as a set of users (e.g., nodes of the network) and links (e.g., the ties) by one or more relations [3]. One can differentiate weak and strong ties which influence user opinions [17]. Among different types of social networks we can distinguish the following: dedicated networks (e.g., networks of business, friends, graduates, etc.); indirect (e.g., online communities) or networks connected with common activities (e.g., co-authors, co-organizers); local networks (e.g., neighborhood, family, employees). Kywe et al. [9] differentiate two types of social recommendation approaches: The first one is connected with other users' recommendation based on their social network and the collaborative or content-based approach, while the second one uses information about social connections and the collaborative approach to recommend new items [18]. In our paper, we focus on the second approach.

The most important challenges are determined by several disadvantages and weaknesses of existing approaches [15]. The first aspect is correlated with data sparsity – determining complete sets of similar users is a problem in NP. Quality of recommendations is mostly biased by the quality of members of the group. The second key issue is connected with a new user (cold-start problem) – the system cannot classify him or her into a proper group until he or she rates some items. The next problem is to overfit the profile and recommending items from a very narrow scope of user interests. The last problem is a more sociological problem: People trust recommendations which were generated in a more transparent way, e.g., by his or her friends (not by anonymous people). Another approach for recommender systems is deep collaborative filtering recommendation [5,19]. Deep learning techniques become more popular in recommendation systems due to the problem of boosting scalability of the system and improving activity among users. The method could be effectively used to enrich collaborative filtering based recommender systems.

In this paper, we develop a framework for a social recommendation system which aims to avoid or fix the above-mentioned problems by combining the collaborative filtering approach with the social collective of users.

## 3 Problem Description

Traditional recommender systems produce recommendations by analyzing historical data about items and/or users. More specifically, a recommender system tries to estimate the relevance of an item $i$ from an itemset $I$ for a user $u$ from a set $U$ of users using a predictive method. In collaborative filtering, this prediction method exploits patterns in previous ratings $r_{u,i}$, $u \in U = \{u_l : l = 1, 2, \dots, n_u\}$ and $i \in I = \{i_k : k = 1, 2, \dots, n_i\}$, where $n_u$ is the number of users and $n_i$ is the numbers of items in the system, respectively. Nevertheless, these traditional approaches usually produce poor recommendations for users that are new to the system. As no historical data about these users is available, it is not possible to find users with alike interests in the past. Hence, the predictive method outputs inappropriate estimations. In the literature, this problem is known as the *cold-start problem*, which we try to tackle with the design of a generic, adaptable framework that considers social-collective knowledge.

## 4 Generic Framework: Social Recommendation System

In this section, we introduce a generic framework for alleviating the cold-start problem by considering social-based recommendations.

A general idea of our approach is to consider a set of registered users. By processing the available knowledge about them it is possible to discover groups of users with similar interests and to determine a representative of each group. Based on only little information about a new user, we can classify him or her into a proper group and recommend relevant items directly at first use of the system. In this paper, we consider social aspects as a source of knowledge about the users.

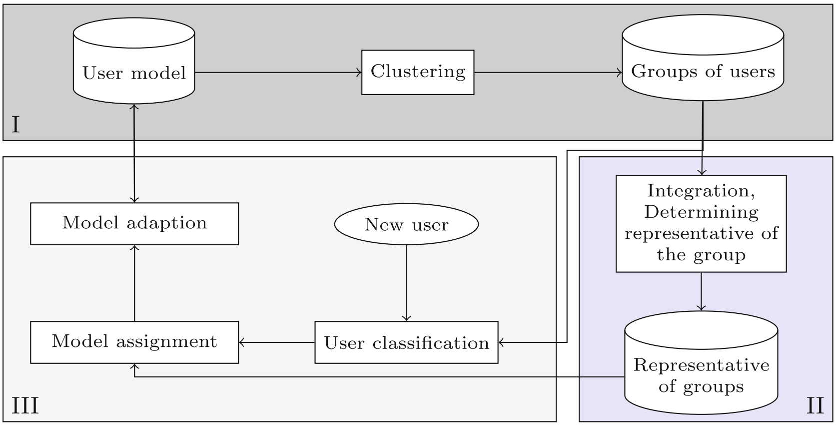


**Fig. 1.** Conceptual schema of the proposed generic framework.

Our framework which is depicted in Fig. 1 consists of three modules, namely, User clustering, Representative, and Adaptive module [13]. An overall idea of the collaborative social recommendation method is presented in Algorithm 1. The first module is connected with a base of users models. A single user model should contain information about users preferences, demographic data, usage

---

**Algorithm 1.** Idea of method for collective social recommendation.

---

**Input**: A set of $N$ users profiles in social community, a new user $u$.
**Output**: A profile for the new user.

1. Cluster users according to social and usage data stored in profiles.
2. Determine a representative profile for each group.
3. Determine significant demographic data for each group of users.
4. **if** *A new user u join the system* **then**
    - (a) Ask him or her about significant demographic data.
    - (b) Collect information about social data.
    - (c) Classify him or her to the most appropriate group.
    - (d) Assign him or her the representative profile of the proper group.
    - (e) Adapt his or her profile according to his or her current activities.

---

data, or social connections to other users in the system. Based on this user model database it is possible to cluster similar user models into groups taking different aspects into account, where each group is stored separately. These groups are updated regularly to account with recent changes into user preferences.

The second module aims to determine a representative model (e.g., characteristics) of a group. When users in the groups change (i.e., when the set of users in the group or the models of users changes) it is necessary to recalculate the representative model.

The last module is correlated with a new user that enters the system. Based on some additional information about the user it is possible to assign him or her to one of the existing groups. The main aim of such an approach is to avoid the cold-start problem – it is possible to assign him or her a representative model of the group. The activities of the user are observed and his or her model is tuned according to new information about the user. In the following subsections, we describe each module of our framework in more detail.

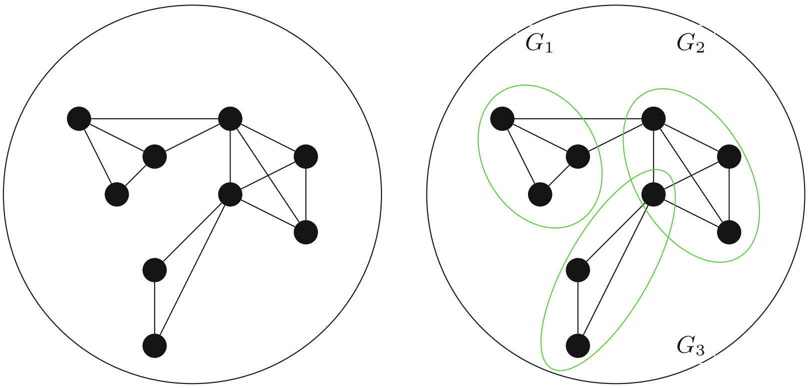


**Fig. 2.** A community in which user relations are represented as edges (left). A community with three groups $G_1$, $G_2$ and $G_3$ given as cliques (right).

### 4.1 User Clustering Module

The user clustering module aims to divide the entire set of users (i.e., user model) into groups so that user with similar interests are part of the same group. In the context of social-based recommendations, the user model is represented as a graph where each node portrays a user and each edge depicts any kind of connection or relation between two users. For instance, relations can describe friendship, collaboration (e.g., coauthors of papers), or interactions in social media channels (e.g., commenting on Facebook or Twitter posts). Given this social representation, user groups can be considered as communities (i.e., a subgraph) within the graph, as Fig. 2 illustrates.

A possible strategy to find communities is to determine cliques in the social graph. A clique is a special graph in which each node is connected to every other node in the graph. In other words, each user in a clique is directly related to all other users of the clique. Although the problem to find cliques within a given graph is NP-complete, this strategy is suitable for finding close related users, and therefore, helps to clarify the idea of a social collective approach. Nevertheless, heuristic algorithms exist to calculate cliques based on a given graph in a more acceptable runtime. Additionally, other graph algorithm approaches to find communities can also be included such as betweenness-oriented [11] or density-oriented algorithms. Figure 2 shows possible cliques and indicates the possibility of assigning a given user to multiple cliques or, more generally, to multiple groups.

### 4.2 Representative Module

The task of the representative module is to determine the representatives of the clustered user groups from the user clustering module in the first place and to store the representatives in a database for further classifications. Here, multiple aspects could be considered to determine a representative. To determine a model of the entire group, it is necessary to check similarity between each pair of users; in particular, some classical methods to determine the centroid or median object can be used.

As we are considering the social collective among users for our recommendation system, our main focus is to identify possible representatives by their importance within the community. Note that the importance could be measured by the number of possible relations each user in a group has to other users in the community, i.e., the number of incoming and outgoing edges of the user. Another possible approach is to weight the edges between users within a clique by the number of communications among them. Thereby, the most communicative user is considered as representative of the group. Finally, a representative could be chosen by its (node) betweenness, which indicates that he or she is important for the interaction between separate groups in the community.

### 4.3 Adaptation Module

The adaption module tends to be the core module to solve the well-known cold-start problem of recommender systems. The cold-start problem describes the situation in which a so-far unknown new user interacts with the recommender system. In this case, a collaborative recommender system fails to recommend items, as no similarities to other users can be determined. Some solutions for this problem involve interactions of the user, e.g., selecting or rating specific items after setting up an account.

More sophisticated methods for the user classification problem were presented in [14]. As we can gather demographic data about users, it is possible to analyze which attributes are significant for the classification process. In the first step, users are clustered into groups of similar interests (based on usage data). Next, in each group, we can determine a subset of demographic data that represents the group in the most suitable way. When a new user is entering our system, it is sufficient to ask him or her only about a few demographic features to classify him or her into a proper group.

In a social collective recommendation system, this situation can be simplified by the fact that in social networks most users directly define relations to their friends and interests without the influence of a system. For this reason, our approach is as follows: A new user is classified by the already calculated groups of users in the user clustering module. After the user has been classified, the model assignment takes place based on the representatives of the different groups. In the last step, the model is adapted or updated with the new user to recalculate the user model and clustering, as they might have changed.

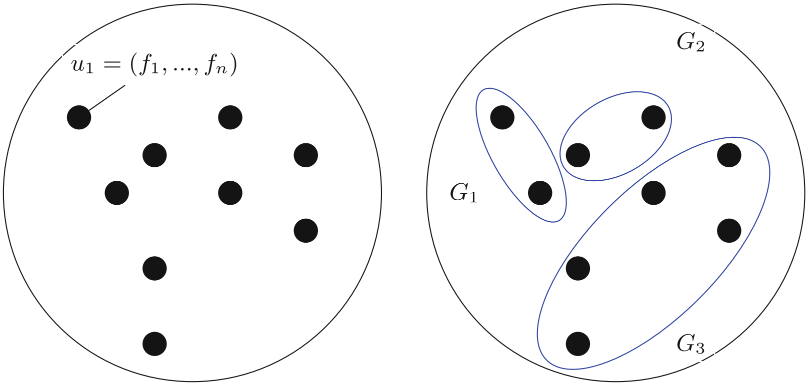


**Fig. 3.** A set of users in which each user is described by a user profile with features (left). Three user groups $G_1$, $G_2$ and $G_3$ discovered by some clustering algorithm based on user profile information (right).

## 5 Proposal of an Experimental Evaluation

In this section, we describe how to evaluate our approach based on comparing user profiles and relations among users for the clustering process. Since the introduced framework comprises only the theoretical aspects of a social-based recommender system, we highlight a concrete experimental scenario is outside the scope of this paper. Nevertheless, a description of future empirical analysis and experimentation is provided in the following.

An empirical evaluation of the benefits brought by our social-based recommendation approach can be assessed by measuring the quality of the returned recommendations via several evaluation metrics such as root mean squared error (RMSE), mean absolute error (MAE), precision, recall, and F-measure. For instance, a baseline approach composed by a k-Means algorithm may be applied to cluster the entire set of user considering only user-related attributes such as age, gender, and location, all stored in a user profile. Thereby, a user profile $u$ is characterized by a number of $n$ attributes so that $u = (f_1, \ldots, f_n)$ with

$u \in U$, as illustrated in Fig. 3. Additionally, in this approach, representatives of the clusters may be determined by, for instance, the centroid of a cluster.

On the other hand, our social-based approach exploits a different data source (i.e., a social database) to find user groups (i.e., communities) and afterward select representatives to alleviate the cold-start problem, as previously shown in Fig. 2. A social database is represented as a graph in which users are nodes and edges are relations among users. For instance, relations may depict users that rate or interact with a particular item (e.g., movie, picture, article). Since a social-based grouping may not depend on any other information besides user-user relations, an empirical evaluation may examine the amount of data needed to generate appropriate results.

We highlight that different clustering or community-based algorithms can be applied to calculate groups of users and find appropriate representatives for the new user. These groups or communities are afterward used to calculate either item-item matrices in the case of item-based collaborative filtering or user-user matrices in the case of user-based collaborative filtering. Especially in the latter case, the storage complexity can be reduced since the size of a user-user matrix is smaller for each group of users – either representing a cluster or community – found by the algorithms. In this sense, an empirical analysis may focus on the reduction of computation runtime due to the fact that groups of users can be employed to speed up similarity calculations during the rating prediction process, for instance, by considering only users within the current representative's group.

Another important analysis consists of evaluating the different representative selection strategies. Since finding appropriate representatives has a substantial impact system's performance, distance or similarity metrics between user profiles or member of a community may be applied as proxies to evaluate the quality of produced recommendations. Finally, we argue that, in order to conduct a complete analysis of a social-based recommender system designed from the principles described in this work, the following data is required:

- User data: Age, gender, date of birth, address, height, weight etc.
- User relations: Relations between users (e.g., friend, family, colleague).
- Usage data: User ratings, comments, reviews, and interactions with items.
- Geographic data: Location of his or her interactions, places that he or she would like to visit etc.

## 6 Summary and Future Work

The importance of recommender systems is still ubiquitous. Social-based recommender systems – especially those exploiting social relationships among users (e.g., communities) to improve personal recommendations – are a new approach in this field. In this paper, we have introduced a generic framework to support the design of collaborative filtering based recommender systems. The proposed framework describes how social collective information can be incorporated in traditional collaborative filtering approaches to alleviate the cold-start problem.

In this sense, our framework considers simultaneously the clustering, the representative determination, and adaption to handle new users. Additionally, we presented our idea how to evaluate the framework's performance compared to a collaborative filtering approach using user profiles.

As our framework is a theoretical one, the main future task is to evaluate its efficiency using real-world datasets. Therefore, it is important to define a specific application for which user profiles, items, ratings and graph representations of users are either already given or can be created from the data at hand. As the quality of user groups have a huge impact on the prediction results, different graph algorithm approaches to determine distinct groups have to be investigated, e.g., based on betweenness, density and so on. Furthermore, the complexity of these algorithms has to be taken into account, as many algorithms are computation-intensive and approximation have to be used. Additionally, the framework should be implemented in a big-data infrastructure, considering the storage of the different models in an appropriate graph database.

We also plan to investigate how the introduced framework can alleviate other traditional collaborative filtering problems such as grey-sheep users (i.e., users who present rating patterns mismatching any other users, and hence, receiving poor recommendations) [8] or shilling attacks (i.e., when fake user profiles are created by malicious users to manipulate item ratings and distrust recommendations) [10].

**Acknowledgment.** This research was partially supported by the Wrocław University of Science and Technology under Polish-German cooperation program between the Ministry of Science and Higher Education and the German Academic Exchange Service, Project No. 0401/0115/18; by statute research grant of Ministry of Science and Higher Education, Project No. 0402/0071/17; by DAAD under grant PPP 57391625; and by the Brazilian National Council for Scientific and Technological Development (CNPq) - Science without Borders Program.

## References

1. Anjumol, M., Ancy, K.S.: A survey on semantic based social recommendation. Int. Res. J. Eng. Technol. (IRJET) **3**(6), 415–420 (2016)
2. Beel, J., Gipp, B., Langer, S., Breitinger, C.: Research-paper recommender systems: a literature survey. Int. J. Digit. Libr. **17**(4), 305–338 (2016)
3. Ben-Shimon, D., Tsikinovsky, A., Rokach, L., Meisles, A., Shani, G., Naamani, L.: Recommender system from personal social networks. In: Wegrzyn-Wolska, K.M., Szczepaniak, P.S. (eds.) Advances in Intelligent Web Mastering. AINSC, vol. 43, pp. 47–55. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-72575-6_8
4. Bernardes, D., Diaby, M., Fournier, R.: A social formalism and survey for recommender systems. SIGKDD Explor. **16**(2), 20–36 (2014)
5. Betru, B.T., Onana, Ch.A, Batchakui, B.: Deep learning methods on recommender system: a survey of state-of-the-art. Int. J. Comput. Appl. **162**(10), 17–22 (2017)
6. Gorripati, S.K., Vatsavayi, V.K.: A community based content recommender systems. Int. J. Appl. Eng. Res. **12**(22), 12989–12996 (2017). ISSN 0973–4562

7. Jannach, D., Zanker, M., Felfernig, A., Friedrich, G.: Recommender Systems: An Introduction. Cambridge University Press, Cambridge (2011)
8. Khusro, S., Ali, Z., Ullah, I.: Recommender systems: issues, challenges, and research opportunities. Information Science and Applications (ICISA) 2016. LNEE, vol. 376, pp. 1179–1189. Springer, Singapore (2016). https://doi.org/10.1007/978-981-10-0557-2_112
9. Kywe, S.M., Lim, E.-P., Zhu, F.: A survey of recommender systems in Twitter. In: Aberer, K., Flache, A., Jager, W., Liu, L., Tang, J., Guéret, C. (eds.) SocInfo 2012. LNCS, vol. 7710, pp. 420–433. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35386-4_31
10. Lam, S.K., Riedl, J.: Shilling recommender systems for fun and profit. In: Proceedings of the 13th ACM International Conference on World Wide Web, pp. 393–402 (2004)
11. Leskovec, J., Rajaraman, A., Ullman, J.D.: Mining of Massive Datasets (2014)
12. Ma, H., Yang, H., Lyu, M.R., King, I.: SoRec: social recommendation using probabilistic matrix factorization. In: International Conference on Information and Knowledge Management (CIKM) (2008)
13. Maleszka, M., Mianowska, B., Nguyen, N.T.: A method for collaborative recommendation using knowledge integration tools and hierarchical structure of user profiles. Knowl. Based Syst. **47**, 1–13 (2013)
14. Mianowska, B., Nguyen, N.T.: A method for collaborative recommendation in document retrieval systems. In: Selamat, A., Nguyen, N.T., Haron, H. (eds.) ACIIDS 2013. LNCS (LNAI), vol. 7803, pp. 168–177. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-36543-0_18
15. Nagwekar, K.: A survey on recommendation systems based on online social communities. Int. J. Innov. Res. Comput. Commun. Eng. **4**(12), 21857–21863 (2016)
16. Ricci, F., Rokach, L., Shapira, B., Kantor, P.B.: Recommender Systems Handbook. Springer, Boston (2011). https://doi.org/10.1007/978-0-387-85820-3
17. Wang, X., Hoi, S.C.H., Ester, M., Bu, J., Chen, Ch.: Learning personalized preference of strong and weak ties for social recommendation. In: Proceedings of International World Wide Web Conference, pp. 1601–1610 (2017)
18. Yang, X., Guo, Y., Liu, Y., Steck, H.: A survey of collaborative filtering based social recommender systems. Comput. Commun. **41**, 1–10 (2014)
19. Zhang, S., Yao, L., Sun, A.: Deep learning based recommender system: a survey and new perspectives. ACM J. Comput. Cult. Herit. **1**(1), Article 35 (2017). https://dblp.uni-trier.de/rec/bibtex/journals/corr/ZhangYS17aa