
A New Collaborative Filtering Approach Based on Game Theory for Recommendation Systems

Selma Benkessirat*, Narhimene Boustia and Rezoug Nachida

SIIR/LRDSI, Blida 1 University, Blida, Algeria

E-mail: selmagsi@hotmail.fr; nboustia@gmail.com; n_rezoug@esi.dz

**Corresponding Author*

Received 13 January 2020; Accepted 01 December 2020;
Publication 09 March 2021

Abstract

Recommendation systems can help internet users to find interesting things that match more with their profile. With the development of the digital age, recommendation systems have become indispensable in our lives. On the one hand, most of recommendation systems of the actual generation are based on Collaborative Filtering (CF) and their effectiveness is proved in several real applications. The main objective of this paper is to improve the recommendations provided by collaborative filtering using clustering. Nevertheless, taking into account the intrinsic relationship between users can enhance the recommendations performances. On the other hand, cooperative game theory techniques such as Shapley Value, take into consideration the intrinsic relationship among users when creating communities. With that in mind, we have used SV for the creation of user communities. Indeed, our proposed algorithm preforms into two steps, the first one consists to generate communities user based on Shapley Value, all taking into account the intrinsic properties between users. It applies in the second step a classical collaborative filtering process on each community to provide the Top-N recommendation. Experimental results show that the proposed approach significantly enhances

Journal of Web Engineering, Vol. 20.2, 303–326.

doi: 10.13052/jwe1540-9589.2024

© 2021 River Publishers

the recommendation compared to the classical collaborative filtering and k-means based collaborative filtering. The cooperative game theory contributes to the improvement of the clustering based CF process because the quality of the users communities obtained is better.

Keywords: Recommendation systems, collaborative filtering, cooperative game theory, Shapley Value.

1 Introduction

Over the past decade, the World Wide Web has surpassed all other technological developments in our history by its ubiquity. As a consequence, we are rapidly dependent on a variety of web applications. These applications should be robust and perform well. Such needs are addressed by the emerging field of Web Engineering. Recommender systems are directly related to website customization. A well-designed recommendation model contributes to solve web engineering approach for the majority of emerging domains in the web such as: e-commerce, e-tourism, e-education, among others.

Recommendation systems (RSs) are in fact a form of information filtering [23], which aims at presenting the information contents that are likely to interest the user. Recently, several techniques for filtering irrelevant resources were proposed in the literature, which are classified into three categories: Collaborative Filtering (CF), Content-Based filtering (CB) and Hybrid filtering.

Among all the available RSs, CF is the most common technique [1, 2, 3]. CF algorithm aims to offer recommendations based on similarities between an active user and other users. It consists of two steps [4]: (1) Create a community of neighbors that contains k most similar users (2) producing a prediction on the information provided by such neighbors. In fact, CF is based on similarity measures, such as Cosine distance, Euclidean distance, etc. The performances of recommendation to a user strongly depend on the choice of similarity measure [5]. The community creation consists of looking for the users who have the same interests as the active user u_a . However, a mean limitation of CF is that the relationship between users into a same community is not taken into account.

Clustering is considered as an unsupervised machine learning approach that involves the grouping points of data. It aims to create groups of objects (named also clusters) based on their common characteristics and features with

the intend to make similar objects in the same groups. The existing clustering techniques aims to achieve optimal clusters with respect of some objectives, for instance, k-means algorithm try to minimize the average squared distance between each point and its closest cluster center. Many studies used clustering in recommender system. In [9] k-means algorithm was used as a pre-processing step to form neighborhood. This later use the distance from the user to different centroids as a pre-selection step for the neighbors. The distance user-user is not taken into account. In this work, we apply clustering in CF process as a pre-processing step of neighbors creation. We focus on two key objectives: (1) minimize the average distance between each user and the cluster centroid. (2) minimize the average intracluster user-to-user distance.

Unfortunately, most of the existing clustering algorithms do not reflect completely the intrinsic notion of a cluster, they aim to minimize distance of every object from its closest representative alone without taking into account the importance of other objects in the same cluster [7]. Our work is powerfully motivated by what has been claimed in [6]; authors have claimed that: “Clustering should be done not just on the basis of distance between a pair of points, A and B , but also on the relationship of A and B to other data points. Therefore, there is a need for an algorithm that gives an optimal solution in keeping both the point-to-center and point-to-point distances, within a cluster, to a minimum.”

On the other hand, “Shapley Value” (SV) [7] is a fair solution concept, from cooperative game theory, in that it divides the collective or total value of the game among the players according to their marginal contributions in achieving that collective value. Further, as described later, there are certain intrinsic properties that are crucial in the context of clustering. The Shapley value is an important solution concept, from cooperative game theory, that satisfies these properties and thereby captures key natural criteria for clustering.

In a perspective of collaborative filtering, intrinsic properties should be taken into consideration when creating communities, as they try to minimize the distance of each user from his only representative in the nearest community, while considering the importance of other users in the same community [7] in order to improve the performance of the recommendations provided. To meet this requirement, we propose in this paper a novel method for Collaborative Filtering based on Game Theory (CF-GT). Our idea is to exploit Game Theory notably Shapley Value algorithm in order to create communities of similar users. We strive to make best use of this intrinsic

property of marginal contribution-based fairness for efficient clustering. Then, apply a classical collaborative filtering on each community. The contributions of this paper can be summarized as follows.

- We formulate the problem of community creation in CF as a game theory problem.
- We propose a novel approach, Collaborative Filtering based on Game Theory for recommendation systems, which uses Shapley Value to create communities.
- We demonstrate the efficacy of our approach through an extensive experimentation.

The remainder of the paper is organized into six sections: Section 2 outlines a brief overview of *CF* based recommendation approaches. Section 3 presents preliminaries on which our proposed approach is based. The proposed framework is discussed in Section 4. The experimental results are presented in Section 5; Section 6 concludes the paper.

2 Related Work

In this section, we give an overview of the most recent and referenced works on recommendation systems. We follow the SLR methodology [24] for identifying, explaining and evaluating the research works. This methodology is mainly composed by three phases namely, planning, performing and comparing the review as depicted in Figure 1 and detailed in the following sections.

2.1 Planning the Review

This step aims to propose a protocol review, which highlights the main hypothesis and objectives of the review and identify clearly the main research questions. In fact, in our study, we present and discuss the recent published works through some well-studied research questions. We mainly focus on:

- RQ1: How recommendation systems can impact different Web engineering systems?
- RQ2: Which domains have benefited from the advantages of recommendation systems?
- RQ3: Which models have been adapted in each system?
- RQ4: How do the authors evaluate their system?

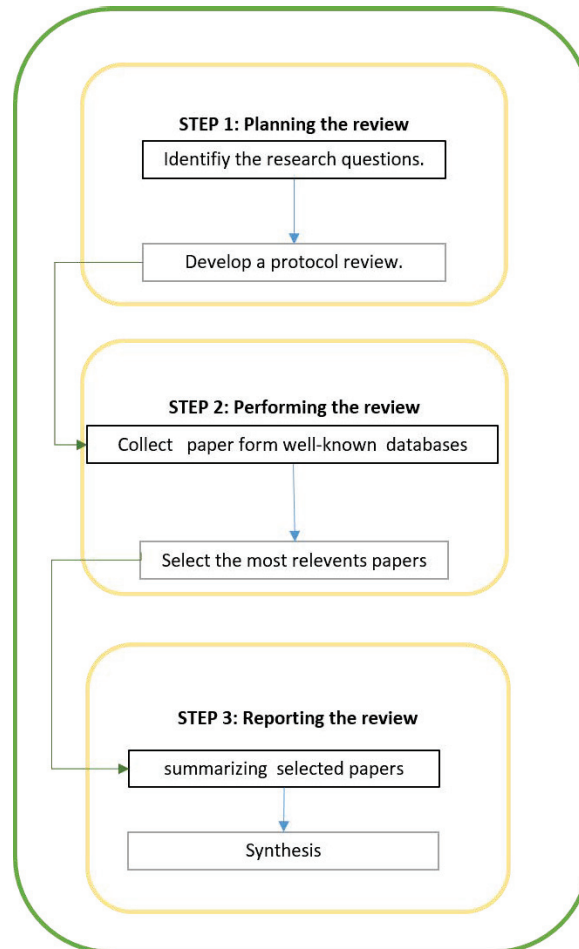


Figure 1 Systematic literature review process.

2.2 Performing the Review

The ultimate goal of this phase is to select a set of papers that meet our objectives. In fact, in our selection strategy, several data bases and research engines were used such as ACM Digital Library, Google Scholar, IEEE Xplore Digita Library, Science Direct, Web of Science and Elsevier so as to find a high-level published studies. Our research strategy is based on a set of key words “recommendation systems” “collaborative filtering”, “content

based filtering”, “k-means based recommendations systems”... etc. After collecting a set of papers based on our research strategy, the next step is to filter this set considering only the relevant and the recent contributions. Indeed, we only consider peer-reviewed journals, conferences and book chapters dated from **2005 to 2019** in order to make a critical analysis of the selected papers.

2.3 Reporting the Review

This step aims at summarizing selected papers to give answers to our research questions and discuss them in depth in order to identify research gaps.

2.3.1 Summarizing selected papers

Several recommendation systems were proposed in the literature. Some of these systems were proposed as general systems, which can be adapted in any domain of application. Other systems have been designed and incorporated in a several aspect of Web engineering. Indeed, RS Can improve extensively the quality of Web services for example to retain customers of a given brand, promote tourism in such a country, improve the quality of distance education, etc. (response of RQ1). Consequently, lots of Web based applications were developed notably in e-commerce, e-learning, e-tourism, etc. (response of RQ2). These applications were developed using a set of techniques like Collaborative Filtering, Content-Based filtering, fuzzy logic and especially machine learning techniques such as clustering based approaches (response of RQ3). With this in mind, we classify the studied approaches into two categories, Clustering based approaches that exploit clustering and the approaches have been proposed as a solution for a specific domain. In the following, we present this two categories starting by Clustering based approaches. It is worth noting that each presented approach is discussed based on the exploited models and its validation (response of RQ4).

Clustering based approaches

Clustering based approaches take into account the similarity between users to more than one level. Clustering makes it possible to have similar user groups; it is a preselecting that takes into account the similarity between the users. After obtaining user clusters, a recommendation process is applied on each cluster independent of others. Many researchers integrated the clustering to the recommendation model [1].

In [11], authors have proposed clustering with regards to recommender system by employing the k-means algorithm as a pre-processing step in order

to aid in neighbourhood formation. As a pre-selection step for neighbours, they employed the distance from the user to various cancroids. Experimental outcomes indicate that the recommended structure can considerably enhance the precision of predication and scalability issue. The datasets utilised for the experimentation are minor to evaluate the scalability problem.

In Linqi et al. [12], an algorithm has been designed that separately explores the context space as well as the item space, and creates an algorithm that integrates clustering of the items with that of information aggregation pertaining to the context space. The proposed framework lacks a test and validation.

Son et al. [13] have put forward an empirical study that integrates Dimension Reduction as well as User Clustering in Collaborative Filtering. A pre-processing phase is employed as the available algorithms. Post that, it employs traditional approaches to deal with these issues. The quality of the model has not been evaluated. An evaluation and comparison with conventional approaches would be necessary to validate it. The results of processing time have shown that the performance of the developed system is significantly enhanced.

Zarzour et al. [14] have developed a new collaborative filtering recommendation algorithm by considering the dimensionality reduction as well as clustering techniques. In order to reap benefits from the advantages pertaining to each algorithm, in this solution, the Singular Value Decomposition (SVD) and k-means algorithm were both used. The experimental results showed that the proposed method improved significantly the performance of the recommendations and remained the lowest values in the RMSE curve in the whole neighbors range.

The presented framework [15] is based on Hierarchical Clustering. The clusters are formed by employing the Chameleon Hierarchical clustering algorithm. As per the experimental results, the put forward framework was seen to yield less error versus K-means based Recommender System. However, lower running time complexity was associated with K-means based RS versus the put forward method.

General approaches

Designed by Beel et al. [8], Docear is regarded as a new system that encompasses different scientific applications, especially an RS pertaining to scientific literature. Docear's suggested module employs CF and CB algorithms. This latter can determine the interests pertaining to the users, thereby offering recommendations that are highly pertinent. The work presents a

limit, the system has not been evaluated or compared. An evaluation would be required to estimate the reliability of the model.

In the domain of e-tourism, a new approach of recommendations on the basis of fuzzy logic and associative classification was proposed by Lucas et al. [9]. The put forward approach considers the collaborative filtering pertaining to the content. This new method can be applied in the tourism system. The analysis pertaining to a realised simulation of real critical situations demonstrated that the filtering methods employed in the method, the fuzzy logic, and the associative classification can be used for the RSs. Thus, issues pertaining to sparsity, scalability and grey sheep have been lightened considerably. However, the system has been evaluated manually with a limited number of users, this must be reviewed to validate the real capabilities of the model.

For the field of Technology-Enhanced Learning (TEL) a recommender system has been proposed in [23]. The authors aimed to cope with grey-sheep problem and to improve the suitability of recommending learning objects. The developed system investigate the possibility of accommodating the computed user's learning style and web page features in order to deliver the best educational resources to every user. The system was evaluated by a group of student and validated in the basis of some hypotheses. TEL is very interesting field. A reliable validation of the developed system would be interesting to move to an efficient use in this field.

For the educational domain, Salehi et al. [10] recommended an RS for learning resources that is based on a multidimensional information model and genetic algorithm. The system includes two key recommendation modules: a module that accounts for explicitly collected attributes and a preference matrix put forward to control the interests of the learner with regards to the explicit characteristics of learning resources for a multidimensional space. The second module is centred on the attributes gathered tacitly and a genetic algorithm is utilised for mining these embedded attributes. The CF recommendation technique has been used to derive both modules. A list of resources is suggested by each module and then a linear combination is used pertaining to collaborative filtering by considering explicit characteristics and collaborative filtering by considering implicit attributes, which are employed for the final recommendation. This system was assessed, and the results showed that the proposed approach surpasses the precision pertaining to conventional algorithms as well as moderates cold start and sparsity issues. However, in the field of educational, it is crucial to integrate knowledge of the learners in order to offer a customised recommendation as well as adequateness in

terms of his educational level. This information is found to be lacking in the suggested recommendation procedure.

2.4 Synthesis

After reviewing the selected approaches, we suggest a comparative study according to several criteria that are related to the proposed research questions.

Table 1 gives a summary of the above-discussed related work, according to several criteria. In Table 1:

- **Domain of Application:** This criterion indicates if the approaches are designed as generic approaches or for a specific domain of application.
- **Used Models:** This criterion determines the used models and techniques in the proposed approaches.

Table 1 Comparison of existing approaches

Reference	Domain Application	Used Models	Clustering	Intrinsic Property	Validation
[8]	Scientific applications	CF CB	×	×	No validation made
[9]	E-tourism	CF, fuzzy logic Associative classification	×	×	Not satisfactory
[23]	TEL	Formal approach	×	×	Not satisfactory
[10]	Educational domain	CF, genetic algorithm model Multidimensional information model	×	×	Very satisfactory
[11]	generic approach	k-means	✓	×	Satisfactory
[12]	generic approach	Contextual bandit algorithm	✓	×	No validation made
[13]	generic approach	CF, Dimensionality reduction	✓	×	No validation made
[14]	generic approach	CF, Dimensionality, SVD, k-means	✓	×	Very satisfactory
[15]	generic approach	Hierarchical clustering algorithm	✓	×	Satisfactory

✓: criterion treated ×: criterion not treated

- **Clustering:** This criterion verifies whether the proposed approaches use a clustering technique as pre-processing step for regrouping similar users.
- **Intrinsic property:** This criterion determines if the proposed approaches take into account the intrinsic notion of a cluster.
- **Validation:** Within this criterion, we aim to study the quality of the validation in each proposed approach.

From this comparative table, we notice some remarks and we distinguish some gaps, which are described below.

- We notice that most of the proposed approaches use the CF algorithm.
- We remark also that most of these approaches take advantage of clustering for regrouping similar users before applying the CF algorithm.
- The similarity is computed only with respect to a user pair for all the approaches.
- The relationship between a user pair and other users in the same group is not taken into account.
- For most approaches, the validation is not satisfactory.

From this perspective, our main goal is to address these gaps by proposing a new approach based on game theory and CF, which takes into consideration the intrinsic notion of a cluster (i.e. Take into account the relationship between a user pair and other users in the same group).

3 Background Material

In this section, we briefly recall some necessary background knowledge, which will be used in the rest of paper including Cooperative Game and Shapley Value algorithm.

3.1 Cooperative Game

A cooperative game can be defined as a pair (N, v) in which N signifies a set of players and $v: 2^N \rightarrow R$ represents the characteristic function. It defines the value $v(S)$ pertaining to any coalition as $S \subseteq N$ [16].

3.2 The Shapley Value

In the cooperative game theory, the Shapley value can be defined as a solution concept. It defines an effective approach that can allocate gains obtained by

cooperation fairly amongst players who are associated with a cooperative game. As certain players may have higher contribution towards the total value versus others, distributing the gains fairly among players is significant. The Shapley value concept regards the relative importance pertaining to each player in the game to decide the payoff that has to be allocated amongst the players. The Shapley value is given as follows:

$$\phi_i(N, v) = \sum_{C \subseteq N-i} \frac{|C|!(n - |C| - 1)!}{n!} \{v(C \cup i) - v(C)\} \quad (1)$$

where $\phi_i(N, v)$ is expected payoff to player i and $N - i$ denotes $N \setminus \{i\}$.

The Shapley value is regarded as a unique mapping that fulfils three fundamental attributes: symmetry, linearity, and carrier property [22]. A player's Shapley value reflects precisely the marginal value that the player brings to the game and the bargaining potential of the player.

4 The Proposed Method: Game Theory based algorithm for Collaborative Filtering

In this section, we introduce a game theory based algorithm for collaborative filtering. Our algorithm performs in two steps, in the first step, it creates groups of similar users. After that, it applies a collaborative filtering algorithm to generate Top-N recommendations. Indeed, the proposed algorithm is based on the weaknesses of the existing related works and it has the following outstanding features:

- Unlike the previous cited works, our proposed algorithm takes into account the intrinsic notion of a cluster.
- To the best of our knowledge, game theory, in particular SV has never been exploited for the benefit of the recommendations systems.
- Is designed as a generic approach and has been implemented as a prototype, which can be used in many domain of application.

In the following, we present our proper modelling of Shapley value for regrouping similar users and cooperative game model. After that, we describe our new algorithm for Collaborative Filtering based on Game Theory.

4.1 Shapley Value for Regrouping Similar Users

The performances of CF algorithm are very sensitive to its first step, which is communities creation. Indeed, the quality results of CF depend on the quality

of the created communities. In fact, the problem of creation of communities can be seen as clustering task. Therefore, the choice of an efficient algorithm to create clusters of similar users is one of the most important issues in CF algorithm. In this paper we adopt the solution concept SV that allows creating clusters of similar users and taking into account some intrinsic properties, which are not considered by conventional algorithms of clustering. The main idea here is to map the communities creation to the formation of coalitions. The usage of the SV for regrouping similar users is justified by its axiomatic qualities:

Axiom 1 (Permutation invariance or symmetry). For any game (N, v) and permutation π on N , the following is applicable:

$\sum_{i \in N} \phi_i(N, v) = \sum_{i \in N} \phi_{\pi(i)}(N, \pi v)$. This axiom indicates that the value is not changed through arbitrary reorder or renaming of the users. Here, we mention about symmetry. This trait is considerably important in achieving order independence, which is another sought-after attribute of regrouping.

Axiom 2 (Preservation of carrier or dummy-property). For any game (N, v) it holds that $v(S \cup \{i\}) = v(S) \forall S \subseteq N$. This axiom indicates a dummy user that does not have any impact on the overall value of a group of similar users. There is no incentive in becoming a member of that group. This addresses the issues related to density, as the points in sparse regions or the outliers are distinctly separated from the points occurring in the high density regions.

Axiom 3 (Additivity or aggregation). For any two games (N, v) and (N, w) it applies that $\phi_i(N, v + w) = \phi_i(N, v) + \phi_i(N, w)$ where $(v + w)(S) = v(S) + w(S)$. If a real number α is used to scale the payoff function v , then the same factor is also used to scale the Shapley value. That indicates the property of linearity. It is necessary for attaining scale invariance with regards to the value function.

Axiom 4 (Normalization or Pareto optimality). For any game it applies that $\sum_{i \in N} \phi_i(N, v) = v(N)$. With regards to the regrouping of the same users, this axiom indicates that the overall value derived due to every user in the dataset has a distribution over all the data points.

In reality, the Shapley value offers a compelling approach to achieve effective clustering. These axioms are satisfied simultaneously by the unique solution concept [22].

4.2 Cooperative Game Model

Let us assume that $U = \{u_1, u_2, \dots, u_n\}$ is a dataset of n users and $d(u_i, u_j) \forall u_i, u_j \in U$ is a function that represents a similarity between u_i and u_j . d refer to any similarity measurement function. Also, let $f : R^+ \cup \{0\} \rightarrow [0, 1)$ imply a monotonically non-decreasing similarity function over d such that $f(0) = 0$. Now, the problem can be observed as grouping those similar users, as represented by f . The input data points are then modelled as a cooperative game (N, v) , where every n user corresponds to a player in the game, so that $|N| = n$. Each player has an interaction with other players and attempts to forge an alliance or group with them, to amplify its value. For our game, for a coalition S , the characteristic function is chosen as in [7]

$$v(S) = \frac{1}{2} \sum_{\substack{u_i, u_j \in S \\ u_i \neq u_j}} f(d(u_i, u_j)) \tag{2}$$

On the whole, the estimation of Shapley values for n players is computationally a difficult problem as it takes into account the average over all the $n!$ permutation orderings. It is shown in [7] that when Equation (2) is used as characteristic function, Shapley value of player i can be computed in polynomial time and is given by:

$$\phi_i = \frac{1}{2} \sum_{\substack{u_j \in U \\ i \neq j}} f(d(u_i, u_j)) \tag{3}$$

4.3 The CF-GT Algorithm Semantics

Our recommendation model goes through two steps, the first one serves to regroup similar users following a cooperative game model, using SV . As for the second step, it is an application of the collaborative filtering algorithm on each group of users for a given active user u_a . Subsequently, we detail the steps of our proposed algorithm.

- **Step 1: (Create Communities User)** in this step, CF-GT takes as input a data set of users to be regrouped and σ a threshold parameter of similarity.

The algorithm starts by computing the Shapley value of each player. Then, in an iterative manner, the algorithm chooses the unallocated point (u_m) with the maximum Shapley value and assigns all points whose similarity to u_m are greater than or equal to σ . σ ensure that nearby

points, which tend to have almost equal SV , are affected to the same group and the starting point of each group are reasonably far. Finally, it outputs a set G containing all similar user groups g_i . The creation of communities user step is summarized in Algorithm 1:

Algorithm 1: Create communities users

```

1  $U = \{u_1, u_2, \dots, u_n\}$  set of users.
2  $\sigma$  threshold similarity.  $G = \{g_1, g_2, \dots, g_k\}$  set of similar users groups.
3 Compute  $SV$  for each user in  $U$ 
4 for  $i = 1$  to  $n$  do
5    $\phi_i = \frac{1}{2} \sum_{\substack{u_j \in U \\ i \neq j}} f(d(u_i, u_j))$ 
6 Form groups of similar users following  $\sigma$ 
7  $C = U$ ;
8 while  $C \neq \emptyset$  do
   •  $m = \text{argmax}(\phi_i)$ ;
   •  $G_m = \{u_i \in C : f(d(u_m, u_i)) \geq \sigma\}$ ;
   •  $C = C \setminus G_m$ ;

```

- **Step 2: (Generating Top- N recommendations)** in this step, CF-GT takes as input the set G . This step aims to provide the top- N recommendation items for each user based on the computation of the users/items matrix and the prediction of the ratings. Generating Top- N recommendations step is formulated in Algorithm 2:

Algorithm 2: Top- N recommendations generating

```

1  $G = \{g_1, g_2, \dots, g_k\}$  set of similar users groups. top- $N$  recommendations
2 foreach  $g_i \in G$  do
   • Compute the user \ item matrix  $M$ 
   • Compute the similarity between users following  $M$ 
3 foreach user  $u_i$  do
   • Give the  $k$ -neighbors of  $u_i$  /* the  $k$ -similar users */
   • Predict rating that  $u_i$  will give to each item following the  $k$ -neighbors
   • Provide the top- $N$  recommendations

```

5 Experiments and Results

In this section, the proposed recommender system is evaluated. Accordingly, we present the results and compare our method with state-of-the-art recommendation methods. Section 5.1 starts by describing the dataset used in our study. Next, Section 5.2 discusses the evaluation measures for recommender

systems. After that, Section 5.3 presents the experimental results. Finally, Section 5.4 discusses our notable findings.

5.1 Simulation Platform

To implement and test the proposed method, we used the Python language on a PC with 2.4 GHz Intel Core i3 CPU, 4 GB RAM, 500 GB Hard Disk and 64-bit Windows 7 Integral Edition Operating System.

5.2 Dataset

To evaluate the proposed model, we consider the MovieLens dataset. This dataset (<http://www.movieLens.org>) is one of the most widely used dataset in the literature for testing recommendation systems. With MovieLens, the results remain relevant because of the density and quality of the data. It allows us to rapidly prototype our systems and compare our approach with other state-of-the-art methods. MovieLens dataset contains 1682 movies and 943 persons. The movies are rated by users (on a 1 – 5 scale). Every user has rated, as a minimum, 20 movies. We divide the data into training and test sets data.

5.3 Evaluation

There is several evaluations metric to evaluate the quality of recommender systems. According to [18, 21], evaluation measures for recommender systems are separated into three categories:

- **Predictive Accuracy Measures:** this is used to evaluate the accuracy of the prediction made by the system by comparing the predictions with the choices the user would have provided in a real case. The Absolute Mean Error (*MAE*) was used, which measures the deviation of the predicted recommendations from the actual choices made by the users. The lower the mean absolute error, the better the prediction.
- **Classification Accuracy Measures:** to determine the effectiveness of the system, we have calculated the accuracy (to determine the probability that a recommended item is relevant) and recall (to determine the probability that a relevant item is recommended). These measure the frequency with which a recommender system makes correct or incorrect decisions about whether an item is good.
- **Rank Accuracy Measures:** These measures assess the significance of the ordering of items conducted by the recommendation system.

Mean Absolute Error (MAE) is a statistical accuracy metrics that evaluate the accuracy of a recommendation model. It shows the deviation between predictions and real user specified values. It is computed as follows:

$$MAE = \frac{1}{N} \sum_{i=1}^N |p_{i,j} - q_{i,j}| \quad (4)$$

Where $p_{i,j}$ is the prediction rating of user i on item j , $q_{i,j}$ is the actual rating and n . is the number of ratings over all the test cases. A lower MAE indicates better performance of the system.

Precision: The ratio of the number of relevant items selected to the number of overall items selected is referred to as precision. The precision is computed as follows [21]:

$$Precision = \frac{TR}{(TR + FR)} \quad (5)$$

FR is the number of false relevant predictions, while TR is the number of true relevant predictions. Precision is defined as the probability that shows the relevance of a selected item.

Recall On the other hand, the ratio of relevant items selected to the total number of relevant items available is called recall. Equation (6) calculates the recall [21].

$$Recall = \frac{TR}{(TR + FN)} \quad (6)$$

Here, FR is the number of false relevant predictions, TR is the number of true relevant predictions, and FN is the number of false non-relevant predictions. Recall signifies the likelihood that a pertinent item would be chosen.

5.4 Experimental Results

In this section, performing various experiments, the evaluation results of CFGT approach are presented. In addition, the proposed approach is compared with state-of-the-art recommendation methods, Collaborative filtering method (CF) and the K-means based Collaborative Filtering ($K - means - CF$). To show the effectiveness of the proposed method in improving the precision of the provided recommendations, we evaluate our method on MovieLens and dataset. For any experiment carried out, we consider three values of the threshold similarity : $\sigma = 0,79, 0,86$ and $0,87$ which gives respectively two, three and four community of users.

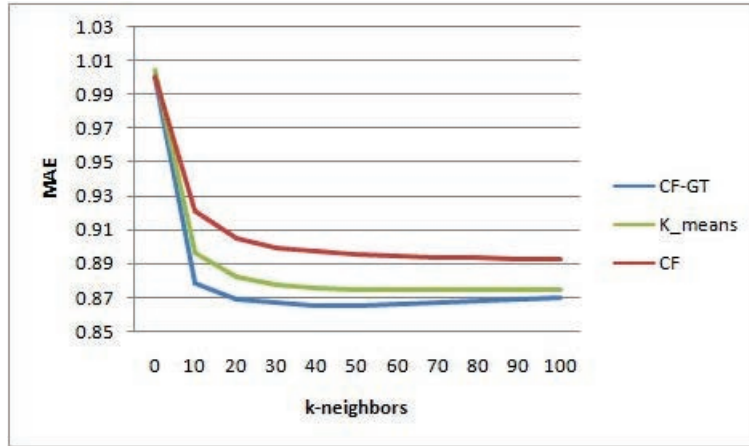


Figure 2 MAE for $\sigma = 0.79$.

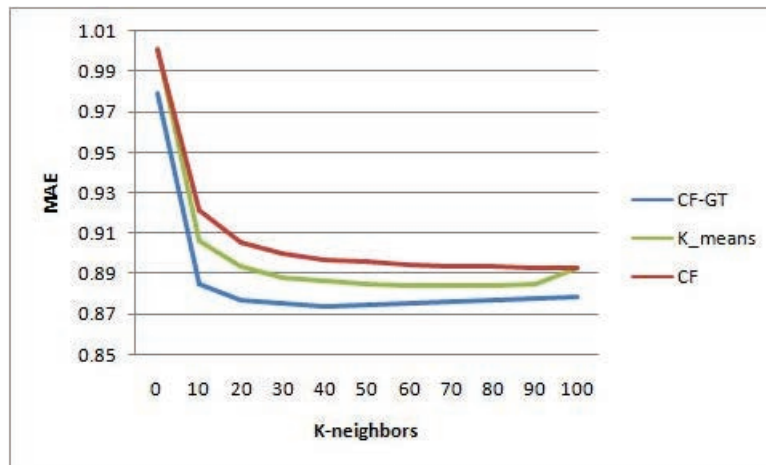


Figure 3 MAE for $\sigma = 0.86$.

- Experiment one:** For predictive accuracy we measure the MAE between the predicted and the actual ratings and compared it with the previous cited approach. Similar to previous studies [19, 2] we consider different numbers of neighbors (k) for this evaluation ($k = 10; 20; 30; 40; 50; 60; 70; 80; 90$ and 100). Figures 2–4 shows the prediction accuracy for different neighborhood size k on MovieLens datasets. It can be found that in the considered

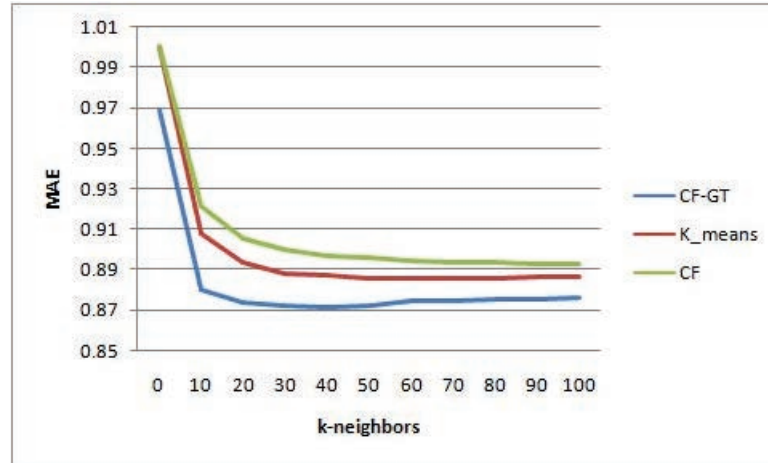


Figure 4 MAE for $\sigma = 0.87$.

neighbor sizes $CF - GT$ help to improve remarkably the prediction accuracy compared with CF and $K - means - CF$. All methods give the most accurate predictions when the number of neighbors is around 40. Without exception, $CF - GT$ outperforms the tests algorithms at any number of neighborhoods. The superiority of the proposed method can be explained by the fact that $CF - GT$ exploits the solution concept SV to form a group of similar users as preselection for CF algorithm. We recall that takes SV takes into account relationship between a user pair and other users in the same group.

- **Experiment two:** to determine the effectiveness of the system, we have calculated the accuracy (to determine the probability that a recommended item is relevant) and recall (to determine the probability that a relevant item is recommended).

With regard to classification accuracy measures the precision and recall were calculated on different numbers of top- N . In our study, we consider ($N = 5, 10, 15, 20, 25, 30, 35, 40, 45$ and 50) which means that we evaluate the method when recommending the top- N movies by the proposed recommender system. Tables 2–4 show the precision and recall values for different top- N .

From these tables we can observe that for different top- N the precision and recall obtained by our proposed method are high in relation to the evaluation methods considered.

Table 2 Precision and recall for $\sigma = 0.79$

Top	CF-GT		CF		K_means	
	Precision	Recall	Precision	Recall	Precision	Recall
5	25%	25%	20%	20%	21%	21%
10	37%	38%	33%	33%	32%	32%
15	45%	45%	40%	41%	39%	40%
20	50%	51%	46%	48%	45%	46%
25	54%	55%	50%	52%	49%	50%
30	57%	58%	53%	55%	52%	53%
35	59%	61%	56%	58%	55%	56%
40	61%	63%	58%	60%	56%	58%
45	62%	65%	59%	62%	58%	60%
50	63%	66%	61%	64%	60%	62%

Table 3 Precision and recall for $\sigma = 0.86$

Top	CF-GT		CF		K_means	
	Precision	Recall	Precision	Recall	Precision	Recall
5	25%	25%	20%	20%	17%	17%
10	37%	37%	33%	33%	27%	27%
15	45%	46%	40%	41%	34%	34%
20	51%	51%	46%	48%	40%	40%
25	54%	55%	50%	52%	44%	44%
30	57%	59%	53%	55%	47%	48%
35	60%	61%	56%	58%	50%	50%
40	61%	63%	58%	60%	52%	53%
45	63%	65%	59%	62%	54%	55%
50	64%	67%	61%	64%	56%	57%

5.5 Discussion

Clustering was chosen as a pre-processing step to form neighborhood in a collaborative filtering process. The conventional clustering techniques, for instance k-means, try to minimize the average squared distance between each point and its closest cluster center. To group the most similar users, these

Table 4 Precision and recall for $\sigma = 0.87$

Top	CF-GT		CF		K_means	
	Precision	Recall	Precision	Recall	Precision	Recall
5	26%	26%	20%	20%	19%	19%
10	38%	39%	33%	33%	29%	29%
15	46%	47%	40%	41%	37%	38%
20	53%	52%	46%	48%	43%	43%
25	56%	56%	50%	52%	47%	47%
30	58%	60%	53%	55%	50%	51%
35	61%	62%	56%	58%	53%	54%
40	62%	64%	58%	60%	55%	56%
45	64%	66%	59%	62%	57%	58%
50	65%	68%	61%	64%	59%	60%

algorithms use the distance from the user to different centroids. The distance user-user is not taken into account.

Our study presents a new collaborative filtering method based on clustering. The clustering technique is based on the solution concept “Shapley value”. This later gives an optimal solution in keeping both the point-to-center and point-to-point distances, within a cluster, to a minimum.

The superiority of our method can be explained by the fact that the preselection of neighborhood take into account the importance of every user from its closest representative and others users in the same cluster.

6 Conclusion

In this study, we presented a game theory based recommendation system. We have integrated a solution concept in cooperative game theory in order to form preselect similar communities user, all taking into account the intrinsic relationship among users. In order to validate our proposed algorithm, several experiments have been carried out. CF-GT was evaluated using *MAE*, precision and recall metrics. Also, a comparative study is presented that compares our proposed algorithm with collaborative filtering algorithm and *k-means* based collaborative filtering algorithm. Experimental results show that the proposed algorithm improve extensively the predictive accuracy.

The strength of our approach is that the preselecting of neighborhood take into account the importance of every user from its closest representative

and others users in the same cluster. Unlike to classical methods which take into account only the importance of every user from its closest representative. That's what makes our approach better than the conventional CF and clustering-based CF. Moreover, as a key feature of our proposed algorithm is that it can be applied in many domains such as e-commerce, e-tourism, e-learning, etc. The main limitation of our approach is that , during the test we only considered the quality of the recommendations provided we did not verify whether our model mitigates the known problems of FC such as sparsity and scalability. As a perspective, improvements and tests will be made to tackle other challenges of CF. Also we intend to validate our approach in educational field. For this, in our future work we envisage to take into account the characteristics of learners by using ontologies.

References

- [1] L.R. Divyaa and N. Pervin. Towards Generating Scalable Personalized Recommendations: Integrating Social Trust, Social Bias, and Geo-spatial Clustering. *Decision Support Systems*, Elsevier ,2019.
- [2] H. Liu, Z. Hu, A. Mian, H. Tian and X. Zhu. A new user similarity model to improve the accuracy of collaborative filtering. *Knowledge-Based Systems*. Elsevier, 56:156–166, 2014.
- [3] M. Sridevi, R.R. Rao and M.V. Rao. A survey on recommender system. *International Journal of Computer Science and Information Security*. LJS Publishing, 14(5):265, 2016.
- [4] R. Pagare, S.A. Patil. Study of collaborative filtering recommendation algorithm-scalability issue. *International Journal of Computer Applications*. Foundation of Computer Science, 67(25):265, 2013.
- [5] A. Jeyasekar, K. Akshay and others. Collaborative Filtering using Euclidean Distance in Recommendation Engine. *Indian Journal of Science and Technology*. Foundation of Computer Science, 9(37):265, 2016.
- [6] R.S. Michalski, R.E. Stepp and E. Diday. A recent advance in data analysis: Clustering objects into classes characterized by conjunctive concepts. *Progress in pattern recognition*. Elsevier, 33–56, 1981.
- [7] V.K. Garg, Y. Narahari and N.N. Narasimha. Novel biobjective clustering (BiGC) based on cooperative game theory. *IEEE Transactions on Knowledge and Data Engineering*. IEEE, 25(5):1070–1082, 2012.
- [8] J. Beel, B. Gipp, S. Langer and M. Genzmehr. Docear: An academic literature suite for searching, organizing and creating academic

- literature. *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*. ACM, 465–466, 2011.
- [9] J.P. Lucas, M. Luz, M. Moreno, R. Anacleto, A.A. Figueiredo and C. Martins. A hybrid recommendation approach for a tourism system. *Expert Systems with Applications*. Elsevier, 40(9):3532–3550, 2013.
- [10] M. Salehi, M. Pourzaferani and S.A. Razavi. Hybrid attribute-based recommender system for learning material using genetic algorithm and a multidimensional information model. *Egyptian Informatics Journal*. Elsevier, 14(1):67–78, 2013.
- [11] G.R. Xue, C. Lin, Q. Yang, W. Xi, H.J. Zeng, Y. Yu and Z. Chen. Scalable collaborative filtering using cluster-based smoothing. *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 14(1):114–121, 2005.
- [12] L. Song, C. Tekin and V.D.S. Mihaela. Clustering based online learning in recommender systems: a bandit approach. *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 4528–4532, 2014.
- [13] N.T. Son, D.H. Dat, N.Q. Trung and B.N. Anh. Combination of Dimensionality Reduction and User Clustering for Collaborative-Filtering. *Proceedings of the 2017 International Conference on Computer Science and Artificial Intelligence*. ACM, 125–130, 2017.
- [14] H. Zorzour, Z. Al-Sharif, M. Al-Ayyoub, Y. Jararweh. A new collaborative filtering recommendation algorithm based on dimensionality reduction and clustering techniques. *2018 9th International Conference on Information and Communication Systems (ICICS)*. IEEE, 102–106, 2018.
- [15] U. Gupta and N. Patil. Recommender system based on hierarchical clustering algorithm chameleon. *2015 IEEE International Advance Computing Conference (IACC)*. IEEE, 1006–1010, 2015.
- [16] G. Chalkiadakis, E. Elkind and M. Wooldridge. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool Publishers, 5(6):1–168, 2011.
- [17] J.L. Herlocker, J.A. Konstan, L.G. Terveen and J.T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems (TOIS)*. ACM, 22(1):5–53, 2004.
- [18] F. Cacheda, V. Carneiro, D. Fernández and V. Formoso. Comparison of collaborative filtering algorithms: Limitations of current techniques and

- proposals for scalable, high-performance recommender systems. *ACM Transactions on the Web (TWEB)*. ACM, 5(1):2, 2011.
- [19] N. Mehrbakhsh, O. Ibrahim, K. Bagherifard. A recommender system based on collaborative filtering using ontology and dimensionality reduction techniques. *Expert Systems with Applications*. Elsevier, 92:507–520, 2018.
- [20] S. Benkessirat, N. Boustia and N. Rezoug. Overview of Recommendation Systems. *Smart Education and e-Learning 2019*. Springer, 357–372, 2019.
- [21] J. Aligon, E. Gallinucci, M. Golfarelli, P. Marcel and S. Rizzi. A collaborative filtering approach for recommending OLAP sessions. *Decision Support Systems*. Elsevier, 69:20–30, 2015.
- [22] R.B. Myerson. *Game Theory: Analysis of Conflict*, Harvard U. Press, Cambridge, MA. 1991.
- [23] M. Tahmasebi, F.F. Ghazvini and M. Esmaili. Implementation and evaluation of a resource-based learning recommender based on learning style and web page features. *Journal of Web Engineering*. River Publishers Alsbjergvej 10, Gistrup, 9260, Denmark, 17(3–4):284–304, 2018.
- [24] B. Kitchenham. Procedures for performing systematic reviews. *Keele, UK, Keele University*.33: 1–26, 2004.

Biographies



Selma Benkessirat is a Ph.D. student at the Computer Sciences Department, Saad Dahlab University, Blida, Algeria. She is a member of the Laboratory LRDSI (Laboratory of Research for the Development of Computerized Systems). Her research interests include machine learning, deep learning, recommender systems and game theory.



Narhimene Boustia received the PhD degree in computer science from from USTHB Algiers in 2011. She is Professor and searcher at Blida 1 University. Her research focuses on security of information system, access control and knowledge management.



Rezoug Nachida is a senior Lecturer in the Computer science department at Saad Dahlab University in Algeria. Her research interests are primarily related to data warehousing solutions, data mining and OLAP system, decision support system, decision making and context-aware recommender system. Dr. Rezoug has co-authored in numerous papers in proceedings of international conferences and journals. She graduated for his ph.D in Computer Sciences from Superior school of Computer science at Algeria in 2016.