Can't Get More Satisfaction? Game-Theoretic Group-Recommendation of Educational Resources

Zacharoula Papamitsiou
University of Macedonia
156 Egnatia Avenue, Thessaloniki, 54621
Greece
papamits@uom.edu.gr

Anastasios A. Economides
University of Macedonia, Information Systems Program
156 Egnatia Avenue, Thessaloniki, 54621
Greece
economid@uom.gr

ABSTRACT

Students' satisfaction from educational resources is a subjective perception of how well these resources meet students' expectations for learning. Recommending educational resources to groups of students, targeting at optimizing all students' satisfaction, is a complicated task due to the lack of joint group profiles. Instead of merging individual profiles or fusing individual recommendations, this paper follows a game-theoretic perspective for solving conflict of interest among students and recommending resources to groups in online collaborative learning contexts: the group members are the players, the resources comprise the set of possible actions, and maximizing each individual member's satisfaction from the selected resources is a problem of finding the Nash Equilibrium. In case the Nash Equilibrium is Pareto efficient, none of the players can get more payoff (satisfaction) without decreasing the payoff of any other player, indicating an optimal benefit for the group as a whole. The comparative evaluation of the suggested approach to other state-of-the-art methods provided statistically significant results regarding the error in predicted group satisfaction from the recommendation and the goodness of the ranked list of recommendations.

CCS CONCEPTS

• **Applied computing~Collaborative learning**

KEYWORDS

Collaborative learning, group recommendation, non-cooperative games, satisfaction

ACM Reference Format:

Z. Papamitsiou, and A. A. Economides. 2018. Can't Get More Satisfaction? Game-Theoretic Group-Recommendation of Educational Resources. In *LAK'18: International Conference on Learning Analytics and Knowledge, March 7–9, 2018, Sydney, NSW, Australia. ACM, New York, NY, USA*, 8 pages. https://doi.org/10.1145/3170358.3170371

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
LAK'18, March 7–9, 2018, Sydney, NSW, Australia
© 2018 Association for Computing Machinery.
ACM ISBN 978-1-4503-6400-3/18/03…$15.00
https://doi.org/10.1145/3170358.3170371

1 INTRODUCTION

Student satisfaction is a subjective perception of how well a learning object prompts students' thinking and learning and supports success [20] . It has been acknowledged for reflecting the feeling of sufficiency from the accomplishment of needs, as well as the consistency between expected gain and the actual experience [22]. Relevant studies shown that learner-learner and learner-item interactions are essential for improving students' satisfaction from online learning experiences and are significant predictors of students' learning gain [8, 17–19, 35, 36].

In order to boost students' satisfaction in online activities, two conditions should meet (among others): (a) the students should interact with each other (i.e., in groups), and (b) the most suitable resources should be recommended to the groups. However, recommending educational resources to groups is not a trivial task [23]; students in a group may not be fulfilled by the same items, yet wish to meet their own expectations, making it difficult to reach to a consensus between group members. Inspired from [7], we argue that solving conflicts of interest between group members could be facilitated by Game Theory.

Game theory is "a study of mathematical models of conflict and cooperation between intelligent rational decision-makers (players)" [28]. The present work demonstrates a method for recommending educational resources to groups of students based on non-cooperative games. *Non-cooperative* is a technical term and not an assessment of the degree of cooperation among players in the game, i.e., a non-cooperative game can model cooperation, focusing on predicting individual players' choices (actions) and payoffs, but the players make self-enforced decisions independently [29].

The problem we address is how to optimally recommend educational resources to a group of students with respect to each individual member's satisfaction from the recommendation. More precisely, in our approach, the group members (students) are the players, the educational resources (items) comprise the set of possible actions, and performing a rational, self-enforced selection of items, i.e., that will maximize each group member's satisfaction (payoff), is a problem of finding the Nash Equilibrium (NE). In this state, if the other students will not modify their own actions, the student who has the option of moving away should have no incentive to unilaterally do so (the payoff doesn't improve). In case the state is Pareto optimal, none of the students can improve their payoff without decreasing the payoff of any other student, indicating a fair and optimal benefit for the group as a whole.

The remainder of the paper is organized as follows. Section 2 briefly reviews existing methods for group recommendations, highlighting the need for additional research in the educational domain. Section 3 formalizes the problem of recommendation of educational resources to groups of students as a non-cooperative game and demonstrates an illustrative example. Section 4 presents the results from the evaluation of our approach, and Section 5 elaborates on our findings and concludes the paper.

2 RELATED WORK AND MOTIVATION OF THE RESEARCH

Recommender systems (RSs) are programs that target at suggesting to their users those items that will best satisfy their preferences, by predicting the users' potential interest in the items; the decision is driven by the collected and analyzed information about the items, the users and the user-item interactions [4]. Literature on the topic is rich [4, 21, 31]. Prevalent approaches include collaborative filtering [34], content-based [32] and knowledge-based [6] techniques. Due to drawbacks and limitations of these techniques (e.g., prediction accuracy, data sparsity, cold-start issues), more sophisticated approaches have been proposed, e.g., fuzzy-logic based [38], social network-based [11] and context aware RSs [1].

There are application domains in which the users need to carry out an activity together, as a group. In these cases, the goal is to recommend items that would meet all users' preferences as much as possible [3]. However, recommending to groups is more complex than recommending to individuals [23]; group members usually don't have the same preferences and interests, making it difficult to reach to an agreement between them and satisfy them all. In order to address this issue, commonly used practices in group-recommender systems either (a) generate recommendations for each group member sparsely and next fuse the lists of individual recommendations, or alternatively (b) merge individual preferences in a single group profile (pseudo user) and apply a recommendation technique on it [3,13].

In each case, group aggregation strategies are recruited to guide the selection of those items that will satisfy all group members, and establish an automatic way of how a group of people can reach to a consensus [25]. [23] summarized eleven strategies including average, least misery, most pleasure, average without misery, plurality vote, etc. For example, MusicFX, a group recommender systems for selecting a music station, uses a variant of the average without misery strategy for group profile aggregation [26]. INTRIGUE, a hybrid system for sightseeing destination recommendation to tourists, takes into account characteristics of sub-groups, creates a model for each of them, and then aggregates the sub-group recommendations [2]. PolyLens recommends movies to groups, using the least misery criterion for fusing recommendations [30], whereas HappyMovie uses the individuals' personality and "social trust" in an average profile strategy [33]. For a systematic review, see [24].

The research in educational group recommender systems is rather sparse, focusing on recommending learning resources to groups of students or to groups of instructors. The researchers aggregated learner profiles in pseudo group profiles prior to generating the recommendation [10] or generated single-user recommendations, constructed homogeneous groups, and next recommended resources to these groups [15]. For recommending learning objects to groups of instructors, DELPHOS applies classification algorithms on meta-data, including students' characteristics and the results from the evaluation of five aggregation methods [39].

In these approaches, aggregations of group members' interests, learning styles and personalities are employed to reach to an agreement between the group members. However, not all aggregation strategies work efficiently in all cases, whereas evaluating the aggregation strategies prior to applying one of them is time consuming. Besides, these methods recommend only one item per time, though it is very likely that students would possibly like to access multiple learning resources. In this case, they would be more pleased with a sequence of suggested items. Yet, homogeneous groups is an unwanted restriction.

Towards addressing these issues, we argue that non-cooperative games could efficiently solve conflicts of interest between group members and guide the recommendation of a sequence of learning resources.

3 PROBLEM FORMULATION AS A NON-COOPERATIVE GAME

3.1 Problem definition

Consider a set of students $L=\{l_i\}$ and a set of educational resources $R=\{r_j\}$; we use the indexes *i, j* to refer to an individual student or resource (item), respectively. For each student *i* and each item *j*, the student's satisfaction s_{ij} can be estimated from the student's self-enforced evaluation of how much the particular item corresponds to the student's expectations and how much it motivates the student's thinking, with respect to the student's learning goals. Students' satisfaction from each item is measured with appropriate questionnaire (see section 4); if a student has not yet evaluated an item, then $s_{ij}=0$.

Also, consider $\hat{s}_{ij}$ as the predicted satisfaction for student *i* from item *j*; $\hat{s}_{ij}$ is computed with Matrix Factorization technique [16] (sub-section 3.2), and is a *decision criterion* for selecting an item *j* given the predicted satisfaction it will excite to a student *i*.

We examine the case of having students who collaboratively solve problems in groups (at least two members). Let G be a set of all groups that may be formed by L; then $|G|=2^n-n-1$. If $g\in G$, then $|g|=k$, the number *k* of group members in group g, with $k\geq 2$. The goal is to recommend to each group those items (single or sequence) that will optimally be beneficial to the group as a whole – supporting the group members to efficiently complete the assigned collaborative task – and that are expected to maximize each individual member's satisfaction as well. The items to be recommended to each group should not have been previously seen or evaluated by any of the group members, and are strategically designed to promote the students' interest in a specific learning topic. The group members are students with potentially competing/conflicting learning goals, expectations and preferences (i.e., the groups might be homogeneous, mildly heterogeneous or heterogeneous groups).

We model the group recommendation problem as a non-cooperative game, i.e., a tuple *(k, Q, f)*, where:

- The *k* students (group members) are the players.
- The set of unrated educational resources (items) $Q = \{q_z\} = \bigcup_i \{j \mid \hat{s}_{ij} = 0\}$, $Q \subseteq R$, are the available actions. A vector $x = (q_1, q_2, ..., q_\mu) \in Q$ is a strategy profile.
- The *payoff* function for a student *i* and a strategy profile *x*, $f_i(x) = \frac{\sum_z \hat{s}_{iz}}{|q|}$, where $|q|$ is the total number of items in the strategy, calculates the predicted satisfaction for student *i in the group*, resulting from the actions (items selected) by all group members – including himself – as the average individual predicted satisfaction from all items in the given strategy.

The items that will be recommended to the group of students are those in the Nash Equilibrium (NE) (single item or sequence of items). This state includes those items that, considering that the other students will not modify their own strategy, the student who has the option of deviating should have no benefit by unilaterally changing his own strategy (does not improve his payoff). Let x_i be a strategy profile for student *i* and x_{-i} be a strategy profile of all students except for student *i*; a strategy profile $x^* \in Q$ is a NE if: $\forall i, x_i \in Q : f_i(x_i^*, x^*_{-i}) \geq f_i(x_i, x^*_{-i})$. In case there are more than one strategies that are NE, the recommendation solution for this group is the one that is "*socially optimum*": no other strategy $x^* \in Q$ has both a weakly better payoff for all students and a strictly better payoff for some student: $\forall i \; f_i(x^*) \geq f_i(x)$ and $\exists i \; f_i(x^*) > f_i(x)$. In other words, if the recommendation solution is Pareto efficient, it is impossible to improve the satisfaction of a student without worsening the satisfaction of another student, indicating an optimal solution for the group. However, it is very possible that none of the NE is Pareto efficient. In this case, we calculate the distance between the highest and lowest payoffs in the strategies that are NE, and select the strategy that minimizes this distance, indicating a fair solution for the group.

Finally, for calculating the predicted group satisfaction, a group consensus function *S(g,Q)* computes the average satisfaction from each item in the recommended strategy *Q* for the group *g*: $S(g,Q) = \frac{\sum_{x \in Q} f_i(x)}{|g|}$, where $f_i(x)$ is the payoff for each member *i* and *x* are the items in the *Q* strategy (i.e., the NE).

The overall architecture of the suggested approach for educational group recommendations is illustrated in Figure 1.

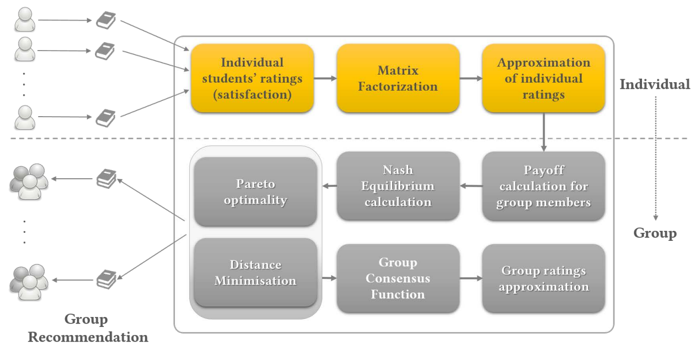


Figure 1: Architecture of non-cooperative game-theoretic group recommender system for educational resources.

3.2 Matrix Factorization

By definition, in non-cooperative games, the students act rationally (i.e., they would select those items that would increase their own satisfaction), and know that the other students act rationally as well. Moreover, in games, it is assumed that the students are aware of their predicted satisfaction from each available strategy, and of the predicted satisfaction of the other group members from their choices. In order to suggest items to the group members, this information should be available to the game-theoretic group recommender, to guide decision support.

As stated in sub-section 3.1, the predicted satisfaction $\hat{s}_{ij}$ for student *i* from item *j* is computed with the Matrix Factorization technique [16]. The basic idea is to view the student-item satisfaction as a sparse matrix, for which we wish to predict the values of its empty cells, such that the values would be consistent with the existing satisfactions in the matrix. This is achieved by computing a low-rank approximation of the satisfaction matrix. As notational convention, bold small letters denote vectors, and bold capital letters denote matrices.

Let **S** be the matrix of size |L| x |R| that contains the satisfaction that the students get from the items. Each student l_i is associated with an *f*-dimensional factor vector $\boldsymbol{l_i}$, and similarly each item r_j with an *f*-dimensional factor vector $\boldsymbol{r_j}$. To get the predicted (approximated) satisfaction from an item r_j for student l_i, the inner product of the corresponding factor vectors is computed: $\hat{s}_{ij} = \boldsymbol{l}_i^T \boldsymbol{r}_j$. The resulting dot product captures the student's l_i overall satisfaction from the item r_j, and models this interaction. The major challenge is then to compute the mapping of each item and each student to the factor vectors, $\boldsymbol{r_j}$, $\boldsymbol{l_i}$, so that they accurately estimate the known satisfactions without over-fitting. The simplest approach to learn the factor vectors is to minimize the regularized squared error on the set of known satisfactions: $min \sum_{(l,r) \in K} (s_{ij} - \hat{s}_{ij})^2 + \lambda(\|l_i\|^2 + \|r_j\|^2)$, where K is the set of (l_i, r_j) pairs for which s_{ij} is known. The constant λ controls the extent of regularization and is usually determined by cross-validation. To minimize this function and determine the factor vectors, Stochastic Gradient Descent [5] can be applied.

3.3 Illustrative example

Consider a group with two members, i.e., the students A and B. Table 1 demonstrates the individual students' predicted satisfaction $\hat{s}_{ij}$ from the educational resources r1, r2, r3, r4 and r5, after matrix factorization. None of the students A and B has previously seen or evaluated any of these five items.

Table 1: The individual students' predicted satisfaction from the educational resources

	r1	r2	r3	r4	r5
A	3.6	4.2	1.8	2.6	3.2
B	1.2	3.4	2.4	4.6	4.6

Table 2 illustrates the payoff (satisfaction) for each student from all the possible actions (strategies) taken by himself and the other group member.

Table 2: The payoff (satisfaction) for each student from all the possible actions (strategies)

		B				
		r1	**r2**	**r3**	**r4**	**r5**
A	**r1**	(3.6, 1.2)	(3.9, 2.3)	(2.7, 1.8)	(3.1, 2.9)	(3.4, 2.9)
	r2	(3.9, 2.3)	(4.2, 3.4)	(3.0, 2.9)	(3.4, 4.0)	(3.7, 4.0)
	r3	(2.7, 1.8)	(3.0, 2.9)	(1.8, 2.4)	(2.2, 3.5)	(2.5, 3.5)
	r4	(3.1, 2.9)	(3.4, 4.0)	(2.2, 3.5)	(2.6, 4.6)	(2.9, 4.6)
	r5	(3.4, 2.9)	(3.7, 4.0)	(2.5, 3.5)	(2.9, 4.6)	(3.2, 4.6)

From this table, it can be seen that there are two NE, the strategies profiles (r2, r4) and (r2, r5). The reason is that if student A chooses action r2, then student B has the same benefit from actions r4 and r5, and does not benefit in changing his action to r1 or r2 or r3. Likewise, considering that student B chooses action r4 or r5, then student A has no benefit to change the action from r2 to r1 or r3 or r4 or r5. Between these two strategies, (r2, r5) is Pareto efficient. This means that these two items (r2 and r5) should be recommended to the group members in order to optimize the satisfaction for each individual member, whereas, no-one of the students can get more payoff (satisfaction) without decreasing the payoff of the other student, indicating an optimal solution for the group as a whole.

4 EXPERIMENTAL EVALUATION

4.1 Participants and experimental setup

The proposed game-theoretic group recommendation method was evaluated on a realistic setting with data from an empirical study with 105 students (59 girls [56.2%] and 46 boys [43.8%], aged 16 years old) from a European High School. The study involved a collaborative activity and was conducted in three phases, in September 2017.

During the first phase, 155 educational resources (i.e., worked examples, solved exercises, self-assessment questions with their answers, etc.), designed to trigger the students' interest on the Python programming language, were randomly assigned to the individuals. Each student had to study and rate at least 3, but not more than 5 items, within 2 days. For the rating of the items, the students had to assess their own perceived usefulness of each item (adopted from [9]) and their own perceived clarity of each item (adopted from [37]), in a 5-point Likert-like scale (Table 3). The average score per student was considered as the student's perceived satisfaction from the corresponding item.

Table 3: Constructs and items for the measurement of satisfaction

Construct	Items	
Perceived Usefulness of item (PU_i)	***PU_i1***	*The item helped me improve my learning*
	PU_i2	*The item enhanced my effectiveness*
	PU_i3	*The item increased my productivity*
	PU_i4	*The item met my expectations*
Perceived Clarity of item's content (PC_i)	***PC_i1***	*The item was clear and understandable*
	PC_i2	*The item was relative with the syllabus*
	PC_i3	*The item helped me greatly to understand the course theme*
	PC_i4	*The item triggered intellectual curiosity*

The resulting dataset from this phase consisted of $|L|$=105 students, $|R|$=155 items and $|S^L|$=605 student-item ratings.

For the needs of the second phase, the students were arranged into four general, equivalent groups: one experimental (*E* – 27 students) and three control groups (*C1* – 26 students, *C2* – 25 students, *C3* – 27 students). Each of these general groups was further partitioned in $|G|$=9 sub-groups (i.e., 36 sub-groups in total), with $|g|$ varying from 2 to 3 students per sub-group. Three types of sub-groups were formed with respect to their members' previous ratings: (a) homogeneous, (b) heterogeneous, and (c) mildly heterogeneous. For the group formation, k-means clustering was applied; the group members were either selected from the same clusters (homogeneous), or they were chosen based on the proximity of the cluster centroids (heterogeneous and mildly heterogeneous). Details on the group formation are beyond the scope of this paper. One (or more) item(s) were delivered to each sub-group regularly (every two days) for two weeks, depending on the recommendation strategy employed. More precisely, the suggested game-theoretic method (*GT*) was used to decide on the recommendations for the sub-groups of *E*, whereas the respective recommendations delivered to the sub-groups of *C1* were generated according to the popular Average aggregation method (*AVG*), the Least Misery method (*LM*) provided the recommendations to the sub-groups of *C2*, and finally, the recommendations to sub-groups of *C3* were decided according to the Most Pleasure method (*MP*). *AVG, LM* and *MP* are briefly demonstrated in section 4.2.2.

After studying the recommended items for two days, all group members had to rate them both individually, and as a team. Every second day the **S** matrix, containing the real individual ratings, was updated. Another matrix, **V**, containing the actual group ratings on the items was also constructed and updated. At the end of the second week, the student-item ratings were $|S^L|$ =1133, and the group-item ratings were $|S^G|$=196. It should be noted that all sub-groups of the control groups received one item per recommendation cycle (every two days), whereas the sub-groups of the experimental group received up-to-three items per cycle. Throughout the experimental process, the items recommended to each sub-group should not have been previously seen and rated by any of the sub-group members.

Finally, the third phase of the activity was about collaboratively writing simple functions in Python, using their knowledge gained during the previous two phases.

4.2 Methods and Evaluation metrics

4.2.1 Group decision strategies. As stated in the previous sub-section, for each one of the control groups (i.e., *C1, C2, C3*), the expected group satisfaction from an item was provided by a different group decision (aggregation) method, formulating how the corresponding sub-groups of students reach to a consensus and come up with a decision about that particular item. More precisely, the group satisfaction ratings were assigned according to the following strategies:

- ***C1 – Average (AVG):*** A consensus-based approach, where all group members jointly and equally make a decision. Let k be the number of students in a group, s_{ij} the satisfaction of student i from item j, then the group satisfaction equals the average satisfaction ratings across the group members: $S(k,j) = \frac{\sum_{i \in g} s_{ij}}{k}$.
- ***C2 – Least-Misery (LM):*** A borderline approach that targets to please the least happy member of the group, resulting the group to behave under a least-misery principle. In this case, the group satisfaction equals the minimum satisfaction among all group members: $S(k,j) = \min_{i \in g} s_{ij}$.
- ***C3 – Most-Pleasure (MP):*** Another borderline group decision strategy satisfying the highest rating within the group. The satisfaction a group of k students gets from an item j equals the maximum satisfaction within the group: $S(k,j) = \max_{i \in g} s_{ij}$.

All solutions were implemented in MATLAB. Furthermore, the Gambit tool [27] was used to verify the correct identification of Nash equilibria.

4.2.2 Evaluation measures. Our proposed method targets at solving conflicts of interest by minimizing the prediction error of group satisfaction from the recommended educational resources (items). In the context of prediction accuracy estimation, the Root Mean Square Error (RMSE) is generally accepted as a good measure of precision, commonly used as an evaluation metric to compare prediction errors of different models for the same data. It measures the sample standard deviation of the difference between values approximated by an estimator and the values actually observed [12]. In our study, we explore the precision of our prediction with respect to satisfaction from the recommended items, as it is actually rated by a given group of students. RMSE is computed as: $RMSE = \sqrt{\frac{\sum_{j=1}^{n}(s_{kj} - \hat{s}_{kj})^2}{n}}$, where n is the number of items rated. Lower values indicate better predictions, and consequently, better decision strategy.

We also used maximum RMSE for capturing the robustness of the recommender system, as it corresponds to the worst-case accuracy *across any group*. Lower mRMSE values indicate that *all groups* will receive good recommendations. This measure is computed as: $mRMSE = max\sqrt{\frac{\sum_{j=1}^{n}(s_{kj} - \hat{s}_{kj})^2}{n}}$.

Furthermore, to measure the quality of the ranked list of recommended items delivered to groups of students, i.e., to evaluate its goodness, we used a measures from Information Retrieval, specifically crafted for ranking: the Normalized Discounted Cumulative Gain (nDCG) which assumes multiple levels of relevance [14].

In simple terms, Discounted Cumulative Gain (DCG) measures the gain of an item (i.e., the relevance score – if rating is missing, zero value is set) based on its position in the resulting list. The gain from the list is accumulated from top to bottom, and more relevant items are preferable to be on the top of the list. Thus, prior to accumulation, the scores are divided by the logarithm of the item's position, leading to a discount. DCG for a group of k students at position N (length of recommendation list), is computed as: $DCG_k@N = s_{kj_1} + \sum_{i=2}^{N} \frac{s_{kj_i}}{log(i+1)}$. However, comparing DCGs between groups of students is not valid. As such, normalized DCG (nDCG) values are computed by arranging all items in an ideal order, and next dividing DCG by the ideal one (IDCG). Accordingly, nDCG is defined as: $nDCG_k@N = \frac{DCG_k@N}{IDCG_k@N}$, where IDCG is the maximum possible DCG, and $nDCG_k@N$ getting values between 0 and 1, with 0 indicating the worst ranking and 1 representing the ideal ranking of items. In our study, due to limitations in available educational resources to be used as the recommendation items set, we only used short lists of up-to five items per group. Thus, we calculated nDCG with N=3 and N=5.

4.3 Results

Tables 4, 5, and 6 demonstrate the results for the evaluation measures for all decision support strategies compared in this study, i.e., the currently proposed game-theoretic method (*GT*) applied on the experimental group, and the Average (*AVG*), Least-Misery (*LM*), and Most-Pleasure (*MP*) methods applied on each one of the control groups, for homogeneous (high inner sub-group similarity), mildly heterogeneous (medium inner sub-group similarity), as well as heterogeneous (low inner sub-group similarity) synthesis of the sub-groups respectively. The sub-groups sizes was firm, varying from two to three students, as explained in section 4.1.

Table 4: Prediction accuracy and goodness of ranked list of recommendations for homogeneous groups

	RMSE	*mRMSE*	*nDCG@3*	*nDCG@5*
GT	**0.348**	0.515	0.967	**0.969**
AVG	0.352	**0.458**	**0.968**	0.968
LM	0.373	0.633	0.889	0.887
MP	0.412	0.839	0.884	0.879

Table 5: Prediction accuracy and goodness of ranked list of recommendations for mildly heterogeneous groups

	RMSE	*mRMSE*	*nDCG@3*	*nDCG@5*
GT	**0.365**	**0.464**	**0.957**	**0.953**
AVG	0.401	0.906	0.925	0.925
LM	0.526	1.212	0.846	0.841
MP	0.815	2.007	0.824	0.798

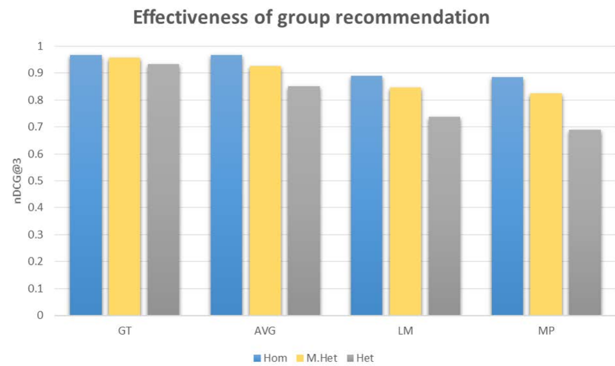
Table 6: Prediction accuracy and goodness of ranked list of recommendations for heterogeneous groups

	RMSE	*mRMSE*	*nDCG@3*	*nDCG@5*
GT	**0.432**	**0.825**	**0.934**	**0.923**
AVG	0.647	1.414	0.852	0.851
LM	0.722	1.883	0.737	0.726
MP	1.206	2.704	0.689	0.624

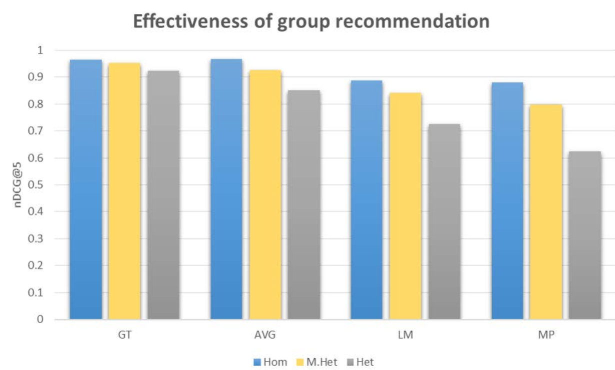
According to these results, all decision support methods achieve low approximation error in prediction of satisfaction ratings for the homogeneous students' sub-groups. On the contrary, for (mildly) heterogeneous sub-groups, accuracy is high for the GT and AVG methods, but the prediction error significantly increases when the aggregation strategy is LM or MP. Furthermore, the group recommendations effectiveness tends to decrease only for the heterogeneous sub-groups. Figure 2 illustrates the goodness of the ranked list of recommended items delivered to the sub-groups of students (a) when the top ranked items are 3 (nDCG@3) and (b) when the top ranked items are 5 (nDCG@5), according to the inner similarity of the sub-groups, and by considering the decision support strategy.

5 DISCUSSION AND CONCLUSIONS

Recommending educational resources (items) to groups of students, targeting at optimizing all students' satisfaction, is a complicated task. The core issue is to determine how a group of students reaches a consensus about the rating for each item in a way that reflects the interests and satisfaction of all group members. This study focuses on solving conflict of interest among students and recommending educational resources to groups in online collaborative learning contexts, and follows a non-cooperative game-theoretic perspective.



(a) 3-top ranked recommendation items



(b) 5-top ranked recommendation items

Figure 2: Effectiveness of group recommendations with respect to the aggregation strategy and the group inner similarity.

Game theory is about social situations, providing solid recommendations to the players regarding their own optimal strategy, as well as administering an external observer that predicts the outcome of interactions (i.e., in our approach, the decision support system). However, the best collective result does not always come from each individuals following their own interest, but rather from reaching the group's consensus; whereas a Nash Equilibrium does not correspond to a socially optimal outcome, a Pareto optimal equilibrium describes a social optimum in the sense that no individual player can improve their payoff without making at least one other player worse off. Pareto efficiency is not a solution concept, but is used to evaluate the overall gain.

An empirical study with a realistic dataset was conducted for the evaluation of the suggested approach. The goal was to compare the performance accuracy and the effectiveness of ranked lists of recommended items delivered to groups of students by the suggested method to other state-of-the-art decision support methods. The following novel facts and important observations have risen.

Firstly, from tables 4, 5 and 6, it becomes apparent that the proposed game-theoretic strategy minimizes the prediction error of the sub-group satisfaction ratings, as, by far, it scores the lowest RMSE values for all categories of inner sub-group similarity. Especially for the highly heterogeneous sub-groups, the other aggregation methods combine potentially conflicting rankings that could create a group recommendation which might not be satisfactory for the group members. In this case, the GT decision strategy resolves sufficiently the conflict of interest and delivers the most appropriate items to the students. However, mRMSE demonstrates some variance in the prediction error *across sub-groups*. In particular, for the homogeneous sub-groups, the GT method did not have the lowest prediction error; in this case, it turns out that the AVG strategy was a better approach, although only slightly. Yet, this was an expected finding, since the average method works well in most cases of homogeneous groups.

Moreover, we also observe that our method has a good overall performance (i.e., the nDCG values reflecting the effectiveness of ranked list or recommendations), although not always the best. However, it is important to notice that, compared to the other methods, the performance of the proposed GT seems to be stable and robust, regardless of the inner sub-group similarity, targeting ranking quality and demonstrating only small variations. From the evaluation results it was found that nDCG for the GT method is close to 1.0 (higher than 0.9) in all cases of sub-group homogeneity, whereas the respective values for the other methods decrease as the inner group similarity decreases.

However, there are some limitations. Firstly, the samples of the 155 educational resources and 105 students considered in the evaluation process are small; bigger datasets should be analyzed. Secondly, we investigated only groups of two to three students; the behavior of GT with larger groups of students (e.g., 4 to 5 members) should be explored as well. Thirdly, for all the involved decision support methods, we left out the comparison of the gain of group recommendations with respect to individual recommendations, i.e., comparing the effectiveness of group recommendations to the gain for individuals from the recommendations; additional evaluation is required in this case. Lastly, we assumed that the group formation method used in this study would not raise issues of uncertainty; more accurate (unbiased) methods for group formation should be applied.

Furthermore, a number of challenges for future work has emerged. For example, more sophisticated measures of satisfaction could be applied (e.g., incorporating the students’ affective states, perceived enjoyment, challenge). The learning analytics research could contribute towards this direction. Yet, another challenging issue is focusing on the transparency of the group recommendation: showing each individual’s payoff and eventually, how satisfied the other group members are, could improve the particular student’s understanding of the recommendation process, and perhaps make it easier to accept the educational resources that initially he/she did not like.

To conclude, the contribution of this paper is the launching of a non-cooperative game-theoretic method for recommending balanced sequences of educational resources (items) to groups of students. The proposed solution demonstrates a socially optimum group recommendation method, beyond aggregation of individual profiles or merging of individual recommendation approaches, and yields statistically significant results even for highly heterogeneous groups of students.

REFERENCES

[1] G Adomavicius and A Tuzhilin. 2008. Context-aware Recommender Systems. *In Proceedings of the 2008 ACM Conference on Recommender Systems*, 335–336. https://doi.org/10.1145/1454008.1454068

[2] L Ardissono, A Goy, G Petrone, M Segnan, and P Torasso. 2003. INTRIGUE: Personalized recommendation of tourist attractions for desktop and hand held devices. *Applied Artificial Intelligence* 17, 8: 687–714. https://doi.org/10.1080/713827254

[3] S Berkovsky and J Freyne. 2010. Group-based recipe recommendations: Analysis of Data Aggregation Strategies. *In Proceedings of the Fourth ACM Conference on Recommender Systems (RecSys ’10)*, 111–118. https://doi.org/10.1145/1864708.1864732

[4] J Bobadilla, F Ortega, A Hernando, and A GutiéRrez. 2013. Recommender Systems Survey. *Know.-Based Syst.* 46: 109–132. https://doi.org/10.1016/j.knosys.2013.03.012

[5] L Bottou. 2010. Large-scale machine learning with stochastic gradient descent. *In Proceedings of COMPSTAT’2010: 19th International Conference on Computational Statistics*, 2010 Keynote, Invited and Contributed Papers, Yves Lechevallier and Gilbert Saporta (eds.). Physica-Verlag HD, Heidelberg, 177–186. https://doi.org/10.1007/978-3-7908-2604-3_16

[6] R Burke. 2002. Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction* 12, 4: 331–370. https://doi.org/10.1023/A:1021240730564

[7] L A M C Carvalho and H T Macedo. 2013. Users’ satisfaction in recommendation systems for groups: An approach based on noncooperative games. *In Proceedings of the 22Nd International Conference on World Wide Web (WWW ’13 Companion)*, 951–958. https://doi.org/10.1145/2487788.2488090

[8] S-H H Chang and R A Smith. 2008. Effectiveness of personal interaction in a learner-centered paradigm distance education class based on student satisfaction. *Journal of Research on Technology in Education* 40, 4: 407–426. https://doi.org/10.1080/15391523.2008.10782514

[9] F D Davis. 1989. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Q.* 13, 3: 319–340. https://doi.org/10.2307/249008

[10] P Dwivedi and K K Bharadwaj. 2015. e-Learning recommender system for a group of learners based on the unified learner profile approach. *Expert Systems* 32, 2: 264–276. https://doi.org/10.1111/exsy.12061

[11] J He and W W Chu. 2010. A social network-based recommender system (SNRS). *In Data Mining for Social Network Data*, Nasrullah Memon, Jennifer Jie Xu, David L Hicks and Hsinchun Chen (eds.). Springer US, Boston, MA, 47–74. https://doi.org/10.1007/978-1-4419-6287-4_4

[12] R J Hyndman and A B Koehler. 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22, 4: 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

[13] A Jameson and B Smyth. 2007. *The Adaptive Web*. In Peter Brusilovsky, Alfred Kobsa and Wolfgang Nejdl (eds.). Springer-Verlag, Berlin, Heidelberg, 596–627.

[14] K Järvelin and J Kekäläinen. 2002. Cumulated Gain-based Evaluation of IR Techniques. *ACM Trans. Inf. Syst.* 20, 4: 422–446. https://doi.org/10.1145/582415.582418

[15] M Kompan and M Bielikova. 2016. Enhancing existing e-learning systems by single and group recommendations. *International Journal of Continuing Engineering Education and Life Long Learning* 26, 4: 386–404. https://doi.org/10.1504/IJCEELL.2016.080980

[16] Y Koren, R Bell, and C Volinsky. 2009. Matrix Factorization Techniques for Recommender Systems. *Computer* 42, 8: 30–37. https://doi.org/10.1109/MC.2009.263

[17] Y-C Kuo, A E Walker, K E E Schroder, and B R Belland. 2014. Interaction, Internet self-efficacy, and self-regulated learning as predictors of student satisfaction in online education courses. *The Internet and Higher Education* 20, Supplement C: 35–50. https://doi.org/10.1016/j.iheduc.2013.10.001

[18] M Kurucay and F A Inan. 2017. Examining the effects of learner-learner interactions on satisfaction and learning in an online undergraduate course. *Computers & Education* 115, Supplement C: 20–37. https://doi.org/10.1016/j.compedu.2017.06.010

[19] H-J Lee and I Rha. 2009. Influence of structure and interaction onstudent achievement and satisfaction in web-based distance learning. *Journal of Educational Technology & Society* 12, 4: 372–382.

[20] C C Lo. 2010. How student satisfaction factors affect perceived learning. *Journal of Scholarship of Teaching and Learning* 10, 1: 47–54.

[21] J Lu, D Wu, M Mao, W Wang, and G Zhang. 2015. Recommender System Application Developments. *Decis. Support Syst.* 74, C: 12–32. https://doi.org/10.1016/j.dss.2015.03.008

[22] C Martin. 1988. Enhancing children's satisfaction and participation using a predictive regression model of bowling performance norms. *The Physical Educator* 45, 4.

[23] J Masthoff. 2011. Group recommender systems: Combining individual models. *In Recommender Systems Handbook*, Francesco Ricci, Lior Rokach, Bracha Shapira and Paul B Kantor (eds.). Springer US, Boston, MA, 677–702. https://doi.org/10.1007/978-0-387-85820-3_21

[24] J Masthoff. 2015. Group recommender systems: Aggregation, satisfaction and group attributes. *In Recommender Systems Handbook*, Francesco Ricci, Lior Rokach and Bracha Shapira (eds.). Springer US, Boston, MA, 743–776. https://doi.org/10.1007/978-1-4899-7637-6_22

[25] J Masthoff and A Gatt. 2006. In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems. *User Modeling and User-Adapted Interaction* 16, 3: 281–319. https://doi.org/10.1007/s11257-006-9008-3

[26] J F McCarthy and T D Anagnost. 1998. MusicFX: An arbiter of group preferences for computer supported collaborative workouts. *In Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98)*, 363–372. https://doi.org/10.1145/289444.289511

[27] R D Mckelvey, A M Mclennan, and T L Turocy. 2006. Gambit: Software Tools for Game Theory. Retrieved from http://econweb.tamu.edu/gambit/

[28] R B Myerson. 1986. An Introduction to Game Theory, In S. Reiter (ed.) *Studies in Mathematical Economics*, The Mathematical Association of America, 1-61.

[29] J Nash. 1951. Non-Cooperative Games. *Annals of Mathematics* 54, 2: 286–295.

[30] M O'Connor, D Cosley, J A Konstan, and J Riedl. 2001. PolyLens: A recommender system for groups of users. *In ECSCW 2001: Proceedings of the Seventh European Conference on Computer Supported Cooperative Work*, Germany, Wolfgang Prinz, Matthias Jarke, Yvonne Rogers, Kjeld Schmidt and Volker Wulf (eds.). Springer Netherlands, Dordrecht, 199–218. https://doi.org/10.1007/0-306-48019-0_11

[31] D H Park, H K Kim, Il Y Choi, and J K Kim. 2012. A literature review and classification of recommender systems Research. *Expert Syst. Appl.* 39, 11: 10059–10072. https://doi.org/10.1016/j.eswa.2012.02.038

[32] M J Pazzani and D Billsus. 2007. *The Adaptive Web.* In Peter Brusilovsky, Alfred Kobsa and Wolfgang Nejdl (eds.). Springer-Verlag, Berlin, Heidelberg, 325–341.

[33] L Quijano-Sanchez, J A Recio-Garcia, B Diaz-Agudo, and G Jimenez-Diaz. 2013. Social factors in group recommender systems. *ACM Trans. Intell. Syst. Technol.* 4, 1: 8:1--8:30. https://doi.org/10.1145/2414425.2414433

[34] B Sarwar, G Karypis, J Konstan, and J Riedl. 2001. Item-based collaborative filtering recommendation algorithms. *In Proceedings of the 10th International Conference on World Wide Web (WWW '01)*, 285–295. https://doi.org/10.1145/371920.372071

[35] A Sher. 2009. Assessing the relationship of student-instructor and student-student interaction to student learning and satisfaction in Web-based Online Learning Environment. *Journal of Online Interactive Learning* 8, 2: 102–120.

[36] H-J So and T A Brush. 2008. Student perceptions of collaborative learning, social presence and satisfaction in a blended learning environment: Relationships and critical factors. *Computers & Education* 51, 1: 318–336. https://doi.org/10.1016/j.compedu.2007.05.009

[37] V Terzis and A A Economides. 2011. The acceptance and use of computer based assessment. *Computers & Education* 56, 4: 1032–1044. https://doi.org/10.1016/j.compedu.2010.11.017

[38] R R Yager. 2003. Fuzzy Logic Methods in Recommender Systems. *Fuzzy Sets Syst.* 136, 2: 133–149. https://doi.org/10.1016/S0165-0114(02)00223-3

[39] A Zapata, V H Menéndez, M E Prieto, and C Romero. 2015. Evaluation and selection of group recommendation strategies for collaborative searching of learning objects. *International Journal of Human-Computer Studies* 76, Supplement C: 22–39. https://doi.org/10.1016/j.ijhcs.2014.12.002