# Preference corrections: capturing student and instructor perceptions in educational recommendations

Check for updates

Yong Zheng

Correspondence: yzheng66@iit.edu
Department of Information
Technology and Management,
School of Applied Technology,
Illinois Institute of Technology,
Chicago, IL, USA

## Abstract

Recommender systems (RS) have been applied in the area of educations to recommend formal and informal learning materials, after-school programs or online courses. In the traditional RS, the receiver of the recommendations is the only stakeholder, but other stakeholders may be involved in the environment. Take educations for example, not only the preference of the student, but also the perspective of other stakeholders (e.g., instructors, parents, publishers, etc) may be important in the process of recommendations. Multi-stakeholder recommender systems (MSRS) were recently proposed to balance the needs of multiple stakeholders in the recommender systems. We use course project recommendations as a case study, and the perspectives of both students and instructors will be considered in our work. However, students and instructors may have different perceptions on the technical difficulty of the projects. In this paper, we particularly focus on the solution of preference corrections which can be used to capture different perceptions of students and instructors in the multi-stakeholder educational recommendations.

**Keywords:** Recommender systems, Educational recommendation, Multi-stakeholder, Preference correction

## Introductions

Recommender systems (RS) produce item recommendations tailored to user preferences. The end user is the only stakeholder in the traditional RS. Recently, researchers claimed that it's necessary to balance the needs of multiple stakeholders in the process of recommendations (Burke, Abdollahpouri, Mobasher, & Gupta, 2016). Take recommending books to the students for example, not only the student's preferences, but also the perspectives of instructors, parents and even publishers may also be important. Maximizing the preference of the end users may hurt the utility of the item from the perspective of other stakeholders. Multi-stakeholder recommender systems (MSRS), therefore, suggest to produce recommendations by considering the perspective of multiple stakeholders, and balance their needs in the environment.

We believe that MSRS is necessary in the area of educations for at least two major reasons. On one hand, the suggestions by other stakeholders are useful or helpful for learners. For example, researchers (Ekstrand, Azpiazu, Wright, & Pera, 2018) pointed

out that the views of parents or instructors on the educational learning materials (e.g., books, textbooks) could be useful to help students select the appropriate learning resources. On the other hand, there could be conflicting interests in the process of teaching and learning. For example, students may prefer to work on easier projects, while the instructors may encourage them to try more challenging ones (Zheng, Ghane, & Sabouri, 2019). In this case, a balance between these stakeholders may be required. The situation could be more complicated, since students and instructors may have different perceptions on the technical difficulty of the projects (Zheng, 2019a). For example, a project seems to be easy from the perspective of instructors, while students may still believe it is difficult for them.

In this paper, we build our MSRS to adapt to the case of course project recommendations. In addition, we propose preference correction as a solution to capture these difference perceptions in the multi-stakeholder educational recommendations.

The major contributions in this paper can be summarized as follows:

- We examine two types of the multi-stakeholder recommendations.
- We additionally evaluate the performance of different multi-objective learning algorithms for multi-stakeholder educational recommendations.
- We deliver more insights about the process of preference corrections by using the extended experimental results.

## Related work and problem statements

### Educational recommender systems

Educational recommender systems emerged as one of the technology-enhanced learning (Drachsler, Verbert, Santos, & Manouselis, 2015) methods. They have been successfully applied to suggest books for K12 users (Pera & Ng, 2013), recommend after-school programs in informal learning (Burke, Zheng, & Riley, 2011), or suggest appropriate citations (He, Pei, Kifer, Mitra, & Giles, 2010) in paper writings.

### Multi-stakeholder recommender systems

Multi-stakeholder recommender systems (MSRS) (Burke et al., 2016) were proposed in 2016, in order to balance the needs of multiple stakeholders. The idea of "multi-stakeholder" is not new, while we can find the earliest research in the category of reciprocal recommendations, such as the applications in dating (Pizzato, Rej, Chung, Koprinska, & Kay, 2010) and recruitment (Yu, Liu, & Zhang, 2011). The idea behind MSRS is that the perspectives of other stakeholders may also be important in the process of recommendations. Take the car advertising as shown by Fig. 1 for example, the *advertising agency* would like to present the car advertisement to any *viewers* who may click it. The receiver of the Ads may just want to view any Ads they are interested in. However, the Ads should be delivered to potential customers from the perspective of *car producers* or *sellers*. Teenagers may like cars but they may not have the capability to make purchases, which decreases the utility of the Ads in view of the producers or sellers.

Researchers have extended the notion of MSRS to several domains, including movies (Burke et al., 2016), music (Abdollahpouri & Essinger, 2017), marketplace (Nguyen,

**Fig. 1** Example in Car Advertising
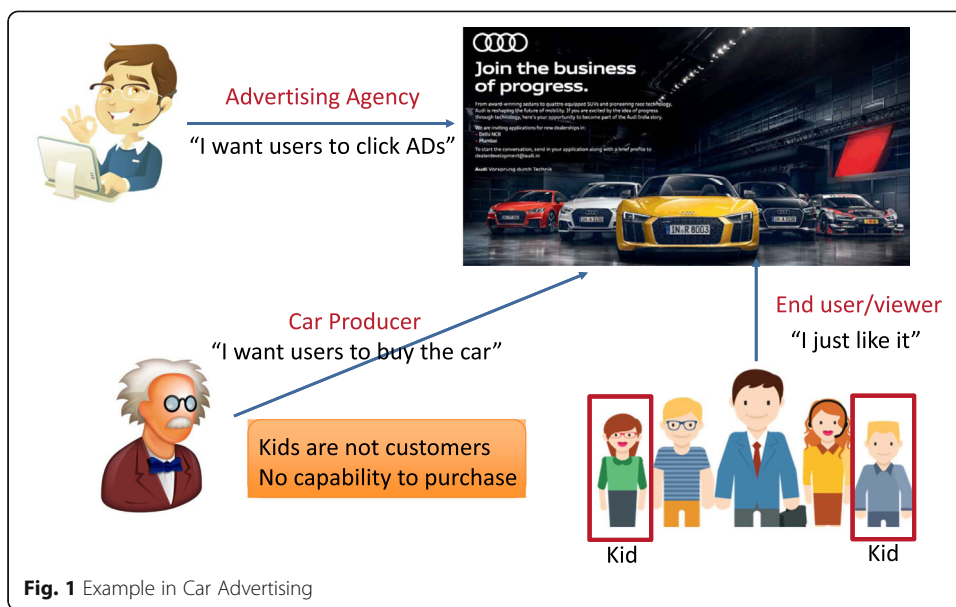
Dines, & Krasnodebski, 2017), dating (Zheng & Pu, 2018a), educations (Burke & Abdollahpouri, 2016; Ekstrand et al., 2018), and so forth. There are two existing work which point out the potential usefulness of MSRS in the area of educations. Ekstrand et al. (2018) believe that the suggestions by parents or instructors are useful to help students select the appropriate learning materials. Burke and Abdollahpouri (2016) discuss the possibility of applying MSRS in after-school programs. For example, an organizer may propose an educational event (e.g., robotics tutorial) and they may predefine some constraints (e.g., it is targeted to 9th grade student only, and they hope to achieve a gender equity in the event). In this case, the recommendations of the after-school programs should be produced by not only matching the student preferences, but also considering the constraints by the organizers. Unfortunately, these two work just discussed the potential applications of MSRS in educations, but no technical solutions were proposed or built. We made the first attempt to build the solution of multi-stakeholder recommendations for the area of educations [Zheng, 2019a; Zheng, Ghane, & Sabouri, 2019].

The major characteristics of MSRS can be summarized as follows:

- There are at least two stakeholders in the system, and these stakeholders must have underlying interactions or connections. Take the dating application for example, the relationship between the target user and the partners to be recommended is reciprocal or bilateral. The target user can select any partners, while he or she can be selected by others as well.
- There may be conflicts of interests among different stakeholders. As a result, maximizing the preference of one stakeholder may hurt the utility of the item from the perspective of other stakeholders.

#### Problem statements

We make our contributions in this paper to address the following problems:

- How can we find a real-world education context in which we need MSRS? and how can we obtain the appropriate data?
- How can we develop the technical MSRS solutions which balance the needs of students and instructors.
- Are there any particular issues or concerns we need to address in an educational MSRS? In our case, we identify the issue of different perceptions of the students and instructors, and propose our solutions.

## Educational setting and data

In this section, we introduce our educational case study and the data set we have. We will discuss how we build the models and produce multi-stakeholder recommendations in the next sections.

We use the data collected by ourselves in the process of academic teaching and learning (Zheng, 2018b). We collected the data from a Web-based learning portal that improves the process of teaching and learning for instructors and students. One of the components in our portal is the project recommendations. Students are required to complete a project for the data analytics courses. They need to find a data set from Kaggle.com, define research problems (e.g., hypothesis testing, regressions, classifications, etc), and utilize their skills to solve the proposed problems. We design a questionnaire in which we provide a list of potential Kaggle data sets and collect student preferences on them. In the questionnaire, each student should select at least three liked and disliked data sets or projects, and give an overall rating to them. In addition, they were asked to rate each selected project on three criteria: how interesting the application area is (*App*), how convenient the data processing will be (*Data*), how easy the whole project is (*Ease*) by using this data set. The rating scale for all ratings is 1 to 5. The dimension "Data" and "Ease" indicate the degree of ease of the project from different perspectives. Table 1 presents an example of our data.

We assign this questionnaire to the students in the data analytics class every semester, and we have collected the data for 2 years. There is a total of 3306 rating entries given by 269 students on 70 Kaggle data sets. Each rating entry is composed of an overall rating and multi-criteria ratings by a student on a selected item.

The course project is used to let students have hands-on practice, and it is better for them to have practical experience in multiple aspects of data analytics. However, some Kaggle entries may provide a preprocessed data, which decreases the difficulty in data processing. Or, they may provide examples of research ideas or problems, which reduces the burdens of brainstorming or critical thinking. Therefore, we also ask instructors to give ratings in the two criteria "Data" and "Ease" for all of the 70 items. These ratings basically reflect the degree of difficulty of the projects from the perspective of

**Table 1** Example of The Educational Data

| Student | Item | Overall Rating | App | Data | Ease |
|---------|------|----------------|-----|------|------|
| 10 | 41 | 4 | 4 | 4 | 4 |
| 10 | 60 | 2 | 2 | 2 | 2 |
| 12 | 21 | 4 | 4 | 5 | 4 |

the instructors. And these ratings can be used to estimate the utility of the projects from the perspective of the instructors. We currently only have one instructor who teaches the data analytics class. We are planning to extend the questionnaire to other courses and instructors in the future. Note the "App" refers to how students like the application or the domain of the data, while instructors have no limitations on it – that's the reason why we did not collected the instructor's rating on "App".

Apparently, students and instructors are the two stakeholders in this case study. On one hand, the instructors respect students' choices and encourage them to look for any data sets they are interested in. They have no limitations on the criterion "App". However, the instructors encourage students to take advantage of this chance and select more challenging projects. The ease of the project is also taken into account in the final grading. This is similar to the diving competition in Olympics games. More specifically, an athlete can select a diving style or action with higher or lower degree of difficulty (Burke et al., 2016). And the final score depends on the degree of difficulty and their performance in the final projects. On the other hand, from the perspective of the students, some of them may prefer to select easier topics since they would like to save time and efforts so that can complete the projects easily and quickly. Some others may prefer to choose more challenging topics so that they can learn more by these hands-on practice. Therefore, a multi-stakeholder recommender system is necessary to recommend appropriate projects to the students by balancing the needs of both students and instructors.

## Utility-based multi-stakeholder recommendations

In this section, we introduce the utility-based multi-stakeholder recommendation models.

### Utility-based multi-stakeholder framework

Utility-based recommendation is one of the recommendation models, according to the classification of recommender systems by Burke (2002). A utility function is necessary to be built to capture the value of the item from the perspective of the end users. The utility score associated with an item and a user can be used to rank the candidate items and produce the top-N recommendations to the end user.

The utility-based multi-stakeholder framework was first proposed in (Zheng, Dave, Mishra, & Kumar, 2018) in 2018, and extended to the area of educations [Zheng, Ghane, & Sabouri, 2019]. It is general enough for the multi-stakeholder recommendations. The workflow can be described as follows:

- First of all, we need to figure out the stakeholders in a system. The utility function should be defined in order to capture the value of the item from the perspective of each stakeholder. The utility function may be the same or different for several stakeholders. Each utility function can produce a utility score associated with an item in view of a stakeholder.
- The ranking score which will be used to produce the top-N recommendations could be simply a function of the utility scores from multiple stakeholders. The

most straightforward approach is a linear aggregation of these utility scores, where the weights are the parameters to be learned in the recommendation process.

- Meantime, we need to define the objectives to balance the needs of multiple stakeholders, and usually there are multiple objectives involved. The weights in the linear aggregations finally can be learned by the multi-objective optimization process.

### Utility functions and educational recommendations

The utility-based multi-stakeholder recommendation models were first proposed for speed dating (Zheng & Pu, 2018a). In this section, we introduce our practice for the educational case study [Zheng, Ghane, & Sabouri, 2019]. More specifically, the key components in the utility-based multi-stakeholder educational recommendations can be described as follows.

#### Utility of the items from the perspective of students and instructors

Given a student $s$ and a candidate item $t$, we first predict how $s$ will rate $t$ in the three criteria, "App", "Data" and "Ease", respectively by using the biased matrix factorization (Koren, Bell, & Volinsky, 2009) which is a standard benchmark in traditional recommender systems. These predicted multi-criteria ratings are used to create a rating vector $R_{s,t}$. For each student, we assume there are student expectations in the same three criteria. These expectations are the latent standard to select the appropriate Kaggle data from the perspective of students. And we represent them as the student expectation vector $E_s$. Namely, in $E_s$, we have a student's expectations on "App", "Data" and "Ease" respectively for the projects he or she likes.

Note that the expectations are not always the "full-stack". The rating in each criterion is not the full or optimal one. Take the hotel reviews on the TripAdvisor.com for example, there may be several criteria, such as location, room size, cleanliness, and so forth. The ideal situation is that we would like to book a hotel with all fivestar in these criteria. However, there are always some factors which may persuade us to lower our expectations, such as the budget in the example of hotel bookings. Similarly, there are also factors in the case of educations, such as the capability of the students which result in the situation that not everyone would like to select more challenging projects in the class.

The instructors encourage students to select more challenging course projects. However, there are always underperformed and outperformed students in the class, and instructors cannot require every student to select more challenging data sets. In other words, the problem in this case study cannot be simply solved by a filter-based or constraint-based recommendation models. To simplify the problem, instructors or professors set up a minimal expectation or requirement which can be described by the vector $E_p$ associated with the criteria "Data" and "Ease" only, since students can select the projects or data sets in any domains or applications (i.e., the criterion "App"). Recall that we have already collected the instructor's rating on each item in our data, and we use the $R_{p,t}$ to represent the rating vector. Therefore, the dissimilarity between $E_p$ and $R_{p,t}$ can be used to denote the utility of an instructor, $U_{p,t}$. The reason why we use dissimilarity is because the $E_p$

represents the minimal requirements, instead of the maximal expectations. The instructors do not have a limit on the more challenging projects, but the instructors would like the students to avoid much easier ones at least.

In terms of the similarity measures, we figured out that Pearson correlations and cosine similarity may not be reliable when the number of multiple criteria in the data is limited. As a result, we calculate the Euclidean distance between the vector of expectations and ratings, normalize it to the scale [0, 1], and use 1 minus the normalized Euclidean distance as the similarity measure in our experiments.

### Multi-stakeholder recommendations by multi-objective learning

For each item to be recommended, we can calculate the utility of the item from the perspective of the student and instructor respectively, as denoted by $U_{s,t}$ and $U_{p,t}$. The utility score is a linear aggregation, $\phi \times U_{s,t} + (1 - \phi) \times U_{p,t}$, while $\phi$ is the weight factor in scale [0, 1]. This utility score is used to rank items to produce the top-$N$ recommendations. Note that 0.5 may not be the best choice for $\phi$, since $U_{s,t}$ and $U_{p,t}$ may be in different distributions. The optimal value of $\phi$ can be learned through a process of multi-objective learning by using the open-source library MOEA (http://moeaframework.org). The multiple objectives can be set up as follows:

- $U_{s,L}$ refers to the utility of student by given the top-$N$ recommendation list $L$. It is the average of the $U_{s,t}$ over all items in $L$, while $t$ is an item in the list $L$.
- $U_{p,L}$ refers to the utility of instructor by given the top-$N$ recommendation list $L$. It is the average of the $U_{p,t}$ in $L$.
- The difference between $U_{s,L}$ and $U_{p,L}$, and we want to minimize this difference for the purpose of balance.
- The recommendation performance, such as precision, recall, NDCG. These metrics can be viewed as another representative of the student utilities. They may be decreased when we additionally consider the utility of instructors.

The multi-objective learning will minimize the difference between $U_{s,L}$ and $U_{p,L}$, and maximize other objectives. The optimal solution is expected to balance the needs of students and professors. It may decrease the recommendation performance since these performance can be viewed as a representation of matching the preferences of the end users, but we expect it is still acceptable. It could be better if the recommendation performance can be improved.

### Student and instructor expectations

Due to the fact that there is only one instructor in our data, we acquire the $E_p$ from the instructor. It is 4 and 4 for the criteria "Data" and" Ease" respectively. In other words, a project with rating 5 in "Data" and" Ease" may not be suggested as the project in the class from the perspective of the instructor. In terms of the student expectations, we can learn these expectations in advance or learn them later with other parameters in the process of multi-objective optimizations. Namely, there are two possible workflows:

- Two-Stage Learning. We can learn student expectations in advance, and finally learn the parameter $\phi$ in the multi-objective learning process. To learn these student expectations, we used the utility-based multi-criteria recommender (UBRec) (Zheng, 2019b). More specifically, the similarity between the expectation vector $E_s$ and rating vector on an item $R_{s,t}$ can be used as the score to rank items for the top-$N$ recommendations. We can learn these student expectations by listwise ranking which maximizes the ranking metric normalized discounted cumulative gain (NDCG) (Valizadegan, Jin, Zhang, & Mao, 2009).
- One-Stage Learning. Alternatively, we can learn both the student expectations and the parameter $\phi$ in the multi-objective learning process.

**Preference corrections**

In this case study, the instructors encourage students to select more challenging projects, while some students may prefer to choose easier ones. We propose the multi-stakeholder recommendation approaches above to balance the needs of these two stakeholders. The degree of difficulty of the projects is the key concern in this.

topic. However, students and instructors may have difference perceptions on the technical difficulty (i.e., the criteria "Data" and "Ease") of the items. For example, from the perspective of the instructor, he or she may find out that students overestimate or underestimate the difficulty of the Kaggle data in terms of the ratings in the "Data" and "Ease". Assume a student's expectation on "App", "Data" and "Ease" is < 4, 4, 3>, we may recommend wrong items to the student if he or she overestimate or underestimate the ratings associated with "Data" and "Ease" on the items. Therefore, the correction of student ratings may be required. From the perspective of students, a similar thing may happen. Students may find out that instructors are too critical on the degree of difficulty, and instructors may overestimate or underestimate the difficulty in the ratings on "Data" and "Ease". In this case, the correction of instructor ratings may be required. Our previous work [Zheng, 2019a]pointed out preference corrections as one of the possible solutions, and we examine these solutions in both one-stage and two-stage learning in this paper. Morespecifically, we can derive three solutions as a process of preference corrections:

- *Student Corrections.* We adjust students' predicted ratings on the items $R_{s,t}$ by aggregating known ratings by instructors $R_{p,t}$. As a result,

$$R_{s,t} = \beta_1 \times R_{s,t} + (1-\beta_1) \times R_{p,t}$$

- *Instructor Corrections.* It is a similar process applied to instructor's ratings.

$$R_{p,t} = \beta_2 \times R_{p,t} + (1-\beta_2) \times R_{s,t}$$

- *Combined Corrections.* We apply both of these corrections.

$\beta_1$ and $\beta_2$ are two weight factors that lie in [0, 1]. Note that $R_{s,t}$ is composed of three dimensions, while there are only two dimensions ("Data" and "Ease") in $R_{p,t}$. Therefore, the corrections above only adjust ratings in the criteria "Data" and "Ease". Students' rating on "App" will not be affected. After corrections, the adjusted ratings will be used to further calculate student and professor utilities. We expect the process of preference corrections can work as a communication between students and instructors, and we may produce better solutions for the multistakeholder recommendations.

## Experiments and results

### Setting and evaluations

We use a 5-fold cross validation for evaluation purpose since the data is relatively small. We evaluate the recommendations by using top-$N$ recommendations. We examine the results by using $N$ as 5 and 10. We only present the results in top-5 recommendations since they present same patterns. We define the relevant items as the items which were given a rating no less than 3 in the test sets. We use F1-Measure (FM) and NDCG as the evaluation metrics for the top-$N$ recommendations. Precision is defined as the ratio of relevant items selected to number of items recommended, and recall presents the probability that a relevant item will be selected. FM is a metric which combines precision and recall, as shown by Eq. 1.

$$FM = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{1}$$

NDCG is a ranking measure from information retrieval, where positions are discounted logarithmically. It is used to evaluate the quality of the ranks in the list of top-$N$ recommendations. Assuming each user $u$ has a "gain" $G_{ui}$ from being recommended an item $i$, the average Discounted Cumulative Gain (DCG) for a list of $J$ items is defined as shown in Eq. 2.

$$DCG = \frac{1}{N} \sum_{u=1}^{N} \sum_{j=1}^{J} \frac{G_{uij}}{\max\left(1, \ \log_b j\right))} \tag{2}$$

NDCG is the normalized version of DCG given by Eq. 3, where $DCG^*$ is the maximum possible DCG.

$$NDCG = \frac{DCG}{DCG*} \tag{3}$$

We use the following approaches as the baselines which are the recommendation methods without considering multiple stakeholders.

- *MF* refers to the biased matrix factorization technique (Koren et al., 2009) which produces the recommendations by using user, item, and overall ratings in the data.
- *SVR* is a linear-aggregation based multi-criteria recommendation method (Adomavicius & Kwon, 2007). We first predict a user's multi-criteria ratings on the items, and build a linear regression model by using support vector regression to estimate the overall rating from these predicted multi-criteria ratings.

- *UBRec* is the utility-based multi-criteria recommendation approach (Zheng, 2019b) we used to learn user expectations.
- $Rank_p$ is one recommendation method that only considers the utility of instructors or professors. We calculate the utility of instructors as the dissimilarity between instructor's expectations and rating vectors. Afterwards, we rank and produce the top-$N$ recommendations based on instructor utilities only.

In terms of the multi-objective learning, we use the MOEA library which is a Java-based open source framework for multi-objective optimization. It defines the whole learning framework, implements the state-of-the-art multi-objective optimization algorithms, and suggests empirical settings for quick experiments. We adopt six mainstream multi-objective learning techniques in the MOEA library:

- NSGA-II (Deb, Pratap, Agarwal, & Meyarivan, 2002) is one of the most popular multi-objective learning techniques. It is composed of two principal parts: a fast non-dominated sorting solution part and the preservation of the solution's diversity.
- NSGA-III (Deb & Sundar, 2006) is an improved version of NSGA-II, in which adopts many new selection mechanisms and it can handle more than two objectives at the same time.
- MSOPS (Hughes, 2003) is a multiple single objective which optimize single objectives respectively and aggregate them together to produce the final solution.
- e-MOEA (Deb, 2003) is a steady-state algorithm, meaning only one individual in the population is evolved per step, and uses an -dominance archive to maintain a well-spread set of Pareto-optimal solutions.
- SMPSO (Nebro et al., 2009) is a multi-objective learning approach based on the particle swarm optimizer (PSO).
- OMOPSO (Sierra & Coello, 2005) is an improved multi-objective PSO by using crowding, mutation and -dominance, and it was demonstrated as one of the top PSO methods to address the multi-objective issues.

We use the suggested empirical settings in the MOEA framework for these multi-objective learning approaches. MOEA setups these quick-run environments to avoid complicated parameter tuning in the experiments. We set the maximal number of function evaluations as 5000, so that it is able to find the best solution within an acceptable running time.

In addition, the multi-objective optimizers may produce multiple solutions. We need to select the best solution by using a pre-defined metric. We introduce the utility loss as the metric for this purpose, as shown below. The "max" values, such as $maxU_{s,L}$, $maxU_{p,L}$, $maxFM$ and $maxNDCG$, are the best values for each metric from the baseline approaches. The loss is composed of the loss in three components – the utility of the recommendation list from the perspective of students and instructors, as well as the recommendation performance. Recall that the multi-stakeholder recommendations will produce the recommendations by considering the perspective of multiple stakeholders. Maximizing the preferences of one

**Table 2** Experimental Results

|  |  | $U_{s,L}$ | $U_{p,L}$ | FM | NDCG | UtilityLoss | $\varphi$ | $\beta_1$ | $\beta_2$ |
|---|---|---|---|---|---|---|---|---|---|
| Baseline Methods | MF | 0.1590 | 0.1143 | 0.0787 | 0.1094 | 0.2804 | N/A | N/A | N/A |
|  | SVR | 0.1794 | 0.1272 | *0.0845* | 0.1156 | 0.2088 | N/A | N/A | N/A |
|  | UBRec | *0.1812* | 0.1338 | *0.0845* | *0.1264* | **0.1838** | N/A | N/A | N/A |
|  | Rank$_p$ | 0.0720 | *0.2982* | 0.0274 | 0.0390 | 0.4288 | N/A | N/A | N/A |
| Two-Stage Learning | Without Correction | 0.1611 | 0.2387 | 0.0622 | 0.0921 | 0.1925 | 0.8400 | N/A | N/A |
|  | Student Correction | 0.2028 | 0.1348 | 0.0810 | 0.1232 | 0.1541 | 0.6100 | 0.9800 | N/A |
|  | Instructor Correction | 0.2021 | 0.1392 | 0.0826 | 0.1257 | 0.1439 | 0.5800 | N/A | 0.3000 |
|  | Combined Correction | 0.2029 | 0.1418 | 0.0838 | 0.1262 | **0.1366** | 0.5400 | 0.9800 | 0.4200 |
| One-Stage Learning | Without Correction | 0.1994 | 0.2506 | 0.0738 | 0.1069 | **0.0665** | 0.8000 | N/A | N/A |
|  | Student Correction | 0.2079 | 0.1339 | 0.0832 | 0.1264 | 0.1371 | 0.9800 | 0.98 | N/A |
|  | Instructor Correction | 0.2079 | 0.1339 | 0.0832 | 0.1264 | 0.1371 | 0.9800 | N/A | 0.9800 |
|  | Combined Correction | 0.2079 | 0.1343 | 0.0832 | 0.1263 | **0.1368** | 0.98 | 0.98 | 0.73 |

Bold entries are significant values

stakeholder may hurt other stakeholders. Therefore, we expect a utility loss in comparison of the baseline methods.

$$UtilityLoss = \frac{1}{3}\left( \frac{\max U_{s,L} - U_{s,L}}{\max U_{s,L}} + \frac{\max U_{p,L} - U_{p,L}}{\max U_{p,L}} + \frac{1}{2}\left( \frac{\max FM - FM}{\max FM} + \frac{\max NDCG - NDCG}{\max NDCG} \right) \right)$$

$$(4)$$

## Results and findings

First of all, we present the experimental results in Table 2 by using e-MOEA which was demonstrated as the best multi-objective optimizer in our experiments. The numbers in italic are the "max" values denoted in Eq. 4. The numbers in bold are the best performing results in the multi-stakeholder recommendations.

In terms of the baseline methods, UBRec is the best performing approach, since it obtains the minimal utility loss in comparison with other baselines. Rank$_p$ obtains the largest $U_{p,L}$ since it is baseline approach which considers the perspective of instructors only.

We build the multi-stakeholder recommendation models by using both two-stage and one-stage learning. In the process of two-stage learning, we learn the student expectations by using UBRec first, and learn other parameters (e.g., $\varphi$, $\beta_1$, $\beta_2$) in the multi-objective learning. By contrast, we will learn all the parameters including the student expectations in the process of one-stage learning.

In two-stage learning, we can observe that by using preference corrections, we are able to reduce the utility loss in comparison with the method without corrections. Particularly, we can obtain the minimal loss by using the combined corrections. In terms of the $\beta$ values, we can observe that $\beta_1$ is large and close to 1, while $\beta_2$ is smaller than 0.5. It tells that the degree of instructor corrections is much larger than the student corrections. Namely, the instructor may underestimate the technical difficulty of the projects from the perspective of the students. Another interesting finding is that the optimal value of $\varphi$ will be closer to 0.5 with preference corrections. As mentioned before, the optimal value of $\varphi$ is not always 0.5 since the distribution of the item utilities from the perspective of students and instructors may be very different. With

appropriate preference corrections, these distributions will also be adjusted, and the optimal value of $\phi$ will be closer to 0.5.

The results in the one-stage learning are different from the ones in two-stage learning, and more difficult to be interpreted. First of all, the approach without preference corrections obtains the lowest utility loss. However, a small loss may indicate that the adjustment is small and the needs of the multiple stakeholders are not well balanced. A further look at this solution reveals that the utility of the recommendation list from the view of the instructors dropped a little bit (0.2506 from 0.2982), which is the major reason why leads to a small utility loss. However, the recommendation performance in terms of FM and NDCG are close to the ones by MF. We believe this solution is not good enough, since the loss is too small and the balance is not well achieved. By using preference corrections, we can obtain similar utility loss as the ones by using the multi-stakeholder solutions in the two-stage learning process. It is still a question that how large or small the utility loss is acceptable. We believe it depends on the tolerance in view of the students and instructors, which requires user studies to learn more insights. In addition, the optimal value of both $\phi$ and $\beta$ parameters are increased in comparison with the optimal ones in the two-stage learning process. Recall that we need to additionally learn the student expectations in the process of one-stage learning. The joint effort of the student expectations and the optimal parameters (e.g., $\phi$, $\beta_1$, $\beta_2$) determines the best performing solutions. In this case, the learned student expectations may affect the optimal values of the $\phi$ and $\beta$ parameters, which makes it more difficult to explain the patterns. More specifically, we believe that learning student expectations together with these parameters may offset the issues that were raised by the different perceptions of the students and instructors. We can still observe that the optimal value of $\beta_2$ was reduced in the combined corrections, which infers that the perspective of instructors needs more corrections.

Finally, we compare different multi-objective optimizers (MOOs) by using the solution with combined corrections in the one-stage learning as an example. MOEA is selected as the best optimizer for two reasons. On one hand, it is able to produce the solutions with less utility loss as shown in Fig. 2. On the other hand, each MOO can



**Fig. 2** Comparison of Multiple MOOs
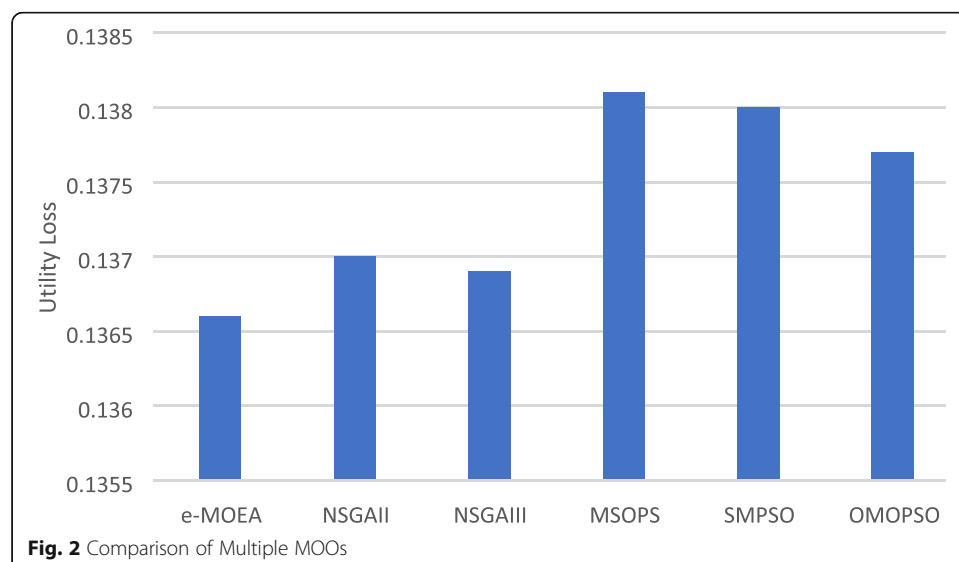
produce multiple solutions, and the average quality of these solutions by using e-MOEA is the best.

## Discussions

First of all, we discuss the validity threats (Wohlin et al., 2012), especially for the reproduction of this study. There are potentially two major threats. One threat to external validity for this study involves the representativeness of our subjects and the educational setting. Students and instructors may exhibit different behaviors, preferences or constraints, if the researchers want to reproduce the results by using a similar setting (i.e., course project recommendations). Another threat to external validity for the experimental results come from the parameter tuning and evaluations by using the library MOEA. As mentioned previously, we use the default parameters in MOEA, and just change the number of function evaluations. However, there are several random initializations in the optimizers, which may leads to different running results. It is suggested to run the optimizations for multiple times in order to identify the best solutions.

In this paper, we propose the MSRS for the course project recommendations. Note that our technical approaches rely on the multi-criteria ratings. The proposed methodologies can also be generalized to other educational settings, as long as the multi-criteria ratings are available. In addition, we can extract multi-criteria ratings or preferences based on the technology of review mining (Chelliah, Zheng, Sarkar, & Kakkar, 2019), if they are not available.

Take the book or learning material recommendations for example, Ekstrand et al. (2018) believed that the suggestions by parents or instructors are useful to help students select the appropriate learning materials. In this case, they can develop some criteria which are associated with the perspectives from students, parents and instructors, such as the content of the books, the storyline, the appropriateness in terms of age and gender (Pera & Ng, 2012), the degree of compatibility with students' major, and so forth. It is worth mentioning that it is not necessary to let all stakeholders to rate the learning materials on all these criteria. Some criteria, such as the appropriateness in terms of age and gender, may be rated by parents, while other criteria may be rated by other stakeholders. In terms of the after-school program recommendations proposed by Burke and Abdollahpouri (2016), the constraints from the perspective of the organizers, such as age and gender, can be developed as the multiple criteria for the organizers only.

## Conclusions and future work

In this paper, we focus on the effect of different perceptions of students and instructors in the multi-stakeholder recommendations, evaluate our solutions by using the one-stage and two-stage multi-stakeholder recommendation processes. Our experimental results discover that both the student and instructor corrections are useful to capture the different perceptions of students and instructors and help find better solutions, especially in the two-stage learning process. In the one-stage process, learning student expectations together with other parameters may offset the issues that were raised by the different perceptions of the students and instructors.

There are two major work we plan to do in the future. On one hand, we use utility loss as one of the metrics to select the optimal solution from the Pareto optimal set. However, it is still a challenge to select the optimal solution based on the loss without the information about the tolerance of the utility loss from the perspective of stakeholders. We plan to assign user studies to learn the tolerance of the utility loss in our future work. On the other hand, we believe user studies and user-centric evaluations are much more important than the offline evaluations. The user experience of each stakeholder, such as the tolerance of the utility loss or the preference of pairwise A/B tests, is the major factor to make a decision on the optimal solution. We will design appropriate user studies to perform user-centric evaluations, in order to better evaluate the proposed algorithms or solutions in our future work.

**References**
Abdollahpouri, H., Essinger, S.: Multiple stakeholders in music recommender systems. arXiv preprint arXiv:1708.00120 (2017).
Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *IEEE Intell Syst, 22.3*, 48–55.
Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Model. User-Adap. Inter., 12*(4), 331–370.
Burke, R., & Abdollahpouri, H. (2016). Educational recommendation with multiple stakeholders. In *2016 IEEE/WIC/ACM international conference on web intelligence workshops (WIW)* (pp. 62–63). IEEE proceedings, Piscataway.
Burke, R., Zheng, Y., & Riley, S. (2011). Experience discovery: Hybrid recommendation of student activities using social network data. In *Proceedings of the 2nd international workshop on information heterogeneity and fusion in recommender systems* (pp. 49–52). ACM Proceedings, New York.
Burke, R. D., Abdollahpouri, H., Mobasher, B., & Gupta, T. (2016). Towards multi-stakeholder utility evaluation of recommender systems. In *UMAP (Extended Proceedings)*.
Chelliah, M., Zheng, Y., Sarkar, S., & Kakkar, V. (2019). Recommendation for multi-stakeholders and through neural review mining. In *Proceedings of the 28th ACM international conference on information and knowledge management*. ACM Proceedings, New York.
Deb, K: A fast multi-objective evolutionary algorithm for finding well-spread pareto-optimal solutions. KanGAL Report No 2003002 (2003).
Deb, K, Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: Nsga-ii. *IEEE Trans. Evol. Comput., 6*(2), 182–197.
Deb, K, & Sundar, J. (2006). Reference point based multi-objective optimization using evolutionary algorithms. In *Proceedings of the 8th Conference on Genetic and Evolutionary Computation* (pp. 635–642).
Drachsler, H., Verbert, K., Santos, O. C., & Manouselis, N. (2015). Panorama of recommender systems to support learning. In *Recommender Systems Handbook* (pp. 421–451). Springer, US.
Ekstrand, M. D., Azpiazu, I. M., Wright, K. L., & Pera, M. S. (2018). Retrieving and recommending for the classroom. *ComplexRec, 6*(2018), 14.
He, Q., Pei, J., Kifer, D., Mitra, P., & Giles, L. (2010). Context-aware citation recommendation. In *Proceedings of the 19th international conference on world wide web* (pp. 421–430). ACM.
Hughes, E. J. (2003). Multiple single objective pareto sampling. In *Evolutionary Computation, 2003*. CEC'03. The 2003 Congress on (Vol. 4, pp. 2678–2684). IEEE proceedings, Piscataway.
Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer, 8*, 30–37.
Nebro, A. J., Durillo, J. J., Garcia-Nieto, J., Coello, C. C., Luna, F., & Alba, E. (2009). Smpso: A new pso-based metaheuristic for multi-objective optimization. In *Computational intelligence in Miulti-criteria decision-making, 2009*. Mcdm'09. Ieee symposium on (pp. 66–73). IEEE proceedings, Piscataway.
Nguyen, P., Dines, J., Krasnodebski, J.: A multi-objective learning to re-rank approach to optimize online marketplaces for multiple stakeholders. arXiv preprint arXiv:1708.00651 (2017).

Pera, M. S., & Ng, Y.-K. (2012). Personalized recommendations on books for k-12 readers. In *Proceedings of the fifth ACM workshop on research advances in large digital book repositories and complementary media* (pp. 11–12). ACM Proceedings, New York.

Pera, M. S., & Ng, Y.-K. (2013). What to read next?: Making personalized book recommendations for k-12 users. In *Proceedings of the 7th ACM conference on recommender systems* (pp. 113–120). ACM Proceedings, New York.

Pizzato, L., Rej, T., Chung, T., Koprinska, I., & Kay, J. (2010). Recon: A reciprocal recommender for online dating. In *Proceedings of the fourth ACM conference on recommender systems* (pp. 207–214). ACM Proceedings, New York.

Sierra, M. R., & Coello, C. A. C. (2005). Improving pso-based multi-objective optimization using crowding, mutation and -dominance. In *International conference on evolutionary multi-criterion optimization* (pp. 505–519). Springer, Berlin, Heidelberg.

Valizadegan, H., Jin, R., Zhang, R., & Mao, J. (2009). Learning to rank by optimizing ndcg measure. In *Advances in Neural Information Processing Systems* (pp. 1883–1891).

Wohlin, C., Runeson, P., H¨ost, M., Ohlsson, M. C., Regnell, B., & Wessl'en, A. (2012). *Experimentation in Software Engineering*. Springer, Berlin, Heidelberg.

Yu, H., Liu, C., & ZHANG, F. (2011). Reciprocal recommendation algorithm for the field of recruitment. *J Inf Comput Sci, 8*(16), 4061–4068.

Zheng, Y., Pu, A. (2018a). Utility-based multi-stakeholder recommendations by multi-objective optimization. In *Proceedings of the 2018 IEEE/WIC/ACM international conference on web intelligence*. IEEE proceedings, Piscataway.

Zheng, Y. (2018b). Personality-aware decision making in educational learning. In *Proceedings of the 23rd international conference on intelligent user interfaces companion* (p. 58). ACM Proceedings, New York.

Zheng, Y.(2019a). Multi-stakeholder personalized learning with preference corrections. In Proceedings of the 18th IEEE International Conference on Advanced Learning Technologies (ICALT). IEEE proceedings, Piscataway.

Zheng, Y. (2019b). Utility-based multi-criteria recommender systems. In *Proceedings of the ACM symposium on applied computing*. ACM Proceedings, New York.

Zheng, Y., Dave, T., Mishra, N., & Kumar, H. (2018). Fairness in reciprocal recommendations: A speed-dating study. In *Adjunct publication of the 26th conference on user modeling, adaptation and personalization* (pp. 29–34). ACM Proceedings, New York.

Zheng, Y., Ghane, N., & Sabouri, M. (2019). Personalized educational learning with multi-stakeholder optimizations. In *Adjunct Proceedings of the ACM Conference on User Modelling, Adaptation and Personalization*. ACM.

## Publisher's Note