Chang, M. et al. (Eds.) (2019). Proceedings of the 27th International Conference on Computers in Education. Taiwan: Asia-Pacific Society for Computers in Education

Discovering the Impact of Student Communities in Educational Recommendations

Yong Zheng
Illinois Institute of Technology
Chicago, Illinois, USA
yzheng66@iit.edu

Abstract: Recommender systems have been applied as one of the technology-enhanced learning techniques for educations. Recently, multi-stakeholder recommender systems were proposed to balance the needs among different stakeholders, especially when there are conflicts of interests. In this paper, we create student communities by using the clustering technique, and seek the impact of these communities in the educational recommendations. Our experimental results identify the best community which can improve the multi-stakeholder recommendations.

Keywords: recommender system, stakeholder, education, clustering, student community

1. Introduction and Related Work

Recommender systems, as one of the technology-enhanced learning methods, has been introduced to the area of educational learning. For example, it is able to suggest books for K-12 users (Pera, 2016), recommend after-school programs (Burke, 2011), or suggest appropriate citations (He, 2010) in paper writings. We successfully have utilized the personality traits to build effective personality-aware recommendation algorithms (Zheng, Subramaniyan, 2019), and develop multi-stakeholder recommendations (Zheng, Ghane, Sabouri, 2019) for the course project recommendations.

In the traditional recommender systems, the receiver of the recommendations is the only stakeholder in the environment. Recently, researchers argue that the perspective of other stakeholders may be also important (Burke, 2016; Zheng, 2017) to be considered in the recommendation process. We made the first attempt to develop multi-stakeholder recommendation methods for educational learning (Zheng, Ghane, Sabouri, 2019; Zheng, 2019a) in order to balance the needs of both instructors and students. More specifically, in the context of course project recommendations, some students may try to select easier projects while the instructors suggest students to work on more challenging projects. Therefore, a multi-stakeholder course project recommendation model is necessary to be built in order to balance the needs of students and instructors.

We have successfully applied the utility-based multi-stakeholder recommendation approaches (Zheng, 2018) to the context of course project recommendations as. In addition, we provide solutions to alleviate the over-/under-expectations (Zheng, Ghane, Sabouri, 2019) and help better capture the different perceptions of the students and instructors (Zheng, 2019a) in our existing work respectively. In this paper, we try to discover the impact of student communities in the multi-stakeholder educational recommendations by creating student communities based on clustering student information.

2. Educational Data and Methodologies

We use the same educational data as the one in our previous work (Zheng, Ghane, Sabouri, 2019; Zheng, 2019a). Each student should select at least three liked and disliked topics of the course, and provide an overall rating to them. In addition, students 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). Both the overall rating and the multi-criteria ratings are in scale 1 to 5. There is a total of 3,306 rating entries given by 269 students on 70 Kaggle data sets. Each rating entry is associated with both overall and multi-criteria ratings. In addition, we have collected the demographic information (i.e., age, gender, nationality) and personality traits represented in five personality dimensions (McCrae, 1992) of the students.

The idea behind the frame work of the utility-based multi-stakeholder recommendations is that the multi-objective learning techniques can be applied to balance the needs of different stakeholders, while we need to build the utility function for each stakeholder in the system. There are different ways to build the utility function. We first claimed that the multi-criteria ratings can be utilized to build these functions (Zheng, 2017). The utility-based multi-stakeholder educational recommendation models have been proposed accordingly (Zheng, Ghane, Sabouri, 2019; Zheng, 2019a).

The utility function in these systems is represented by the similarity between a rating vector on an item, and an expectation vector. These two vectors are mapped to the multiple criteria in the system. In our data set, we have three criteria – App, Data and Ease. An expectation vector by the student (i.e., student expectations) indicates which items they may prefer in terms of the three criteria. By contrast, the expectation vector of the instructor (i.e., instructor expectations) is defined as the instructor's minimal requirements in the system. Both students and instructors will give multi-criteria ratings to the items. Note that the ratings by students can tell student preferences, while the ratings given by instructors indicate how appropriate the item can be selected as projects from the perspective of the instructors. Therefore, the utility of an item *t* from the perspective of a student *s* ($U_{s,t}$) can be represented by the similarity of student expectations and the rating vector on the item by the student. By contrast, the utility of an item *t* from the perspective of an instructor *p* ($U_{p,t}$), can be represented by the dissimilarity between the expectation and evaluation rating vectors based on the "Data" and "Ease" dimensions. We can calculate a *utility score* which is a combination of the item utility from the perspective of students and instructors, i.e., *utility score* = $\alpha \times U_{s,t} + (1-\alpha) \times U_{p,t}$, while α ($0< \alpha<1$) is the weight factor. The optimal value of α may not be 0.5, due to the different distribution of $U_{s,t}$ and $U_{p,t}$. We use the multi-objective learning technique to learn the optimal value of α. We collected the instructor expectations in advance, since there is only one instructor. We can learn the student expectations in advance by the utility-based multi-criteria recommendation models (UBRec) (Zheng, 2019b), or we can learn them in the process of multi-objective learning. In this paper, we decide to learn them in advance. An evaluation metric "score" which is the difference between utility gain and utility loss. For more details, please refer to our previous work (Zheng, 2019a).

In this paper, we would like to discover the impact of student communities in the multi-stakeholder educational recommendations. More specifically, we can update the *utility score* to be $\alpha\times(\beta\times U_{s,t} + (1-\beta)\times U_{c,t}) + (1-\alpha) \times U_{p,t}$, while β is another weight factor in [0, 1]. In comparison with the previous utility score, we use a linear combination of $U_{s,t}$ and $U_{c,t}$, where we use *c* to represent the student community and $U_{c,t}$ therefore as the utility of the item from the perspective of community *c*. We decide to use X-Means (Pelleg, 2000) to cluster students into different communities, while we can utilize the following feature sets for the purpose of clustering -- demographic information (F_d), personality traits (F_p) and ratings on the items (F_r).

3. Experiments and Results

We use the same setting in our previous work (Zheng, 2019a), e.g., we use 5-fold cross validation, and the ε-MOEA as the multi-objective optimizer in the MOEA library[1]. We use three baseline approaches – UBRec (Zheng, 2019b) which only considers the student preferences, SolBasic which is the basic multi-stakeholder solution (Zheng, 2018) and SolCorr which is the improved multi-stakeholder solution which considers preference corrections (Zheng, 2019a).

The experimental results can be observed from Figure 1. The basic multi-stakeholder solution SolBasic failed to beat UBRec which is a traditional recommendation model. Once we applied the preference corrections (Zheng, 2019a), SolCorr is able to perform the best among these baselines. Other solutions are the ones that we incorporate the student communities. We can observe that all of these solutions are able to beat the SolBasic. By using F_d, F_p and F_r respectively, we can observe that F_r is the most effective feature set, while the performance based on clusters using F_d performs the worst.

When we combine the feature sets together to produce clusters, the experimental results vary. The performance was even decreased, when we combined F_d with other feature sets. For example, the

[1] MOEA, http://moeaframework.org

score by using F_d+F_p is lower than the one using F_p only. It infers that F_d is a set of noisy features. A closer look in F_d can reveal that most of the students are from a same country, and most of the students are in the same age group. That's probably the reason why F_d seems to be not useful in our experiments. The combination of F_p and F_r turns out to be the optimal feature set to create the clusters if we combine two feature sets only. However, the score is still lower than the SolCorr. The best solution is the F_p+F_r+Corr in which we utilize F_p and F_r to create student communities, and also apply the preference corrections (Zheng, 2019a). In comparison with SolCorr, the score was improved by around 2%.

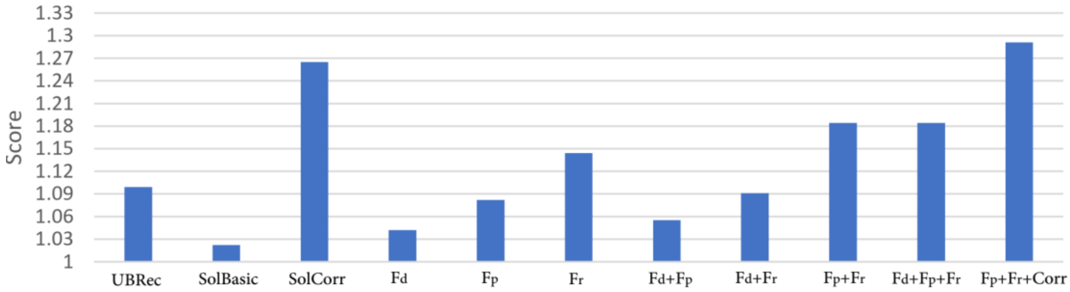

Figure 1. Experimental Results

Recall that β is used to fuse the utility of the item from the perspective of students and the communities. In most of the proposed solutions, the value of β varies from 0.6 to 0.8, while the β is 0.72 in the best solution F_p+F_r+Corr. The contribution by the student communities is 28%, which confirms the effective impact by the student communities.

4. Conclusions and Future Work

In this paper, we found that the student communities could be able to improve the multi-stakeholder educational recommendations, if we use the appropriate feature sets to create the student communities. In our future work, we plan to apply A/B test to further examine the effectiveness of these models.

References

Burke, R. D., Abdollahpouri, H., Mobasher, B., & Gupta, T. (2016). Towards Multi-Stakeholder Utility Evaluation of Recommender Systems. In *ACM UMAP (Extended Proceedings).*

Zheng, Y. (2017). Multi-stakeholder recommendation: Applications and challenges. arXiv preprint arXiv:1707.08913.

Zheng, Y., Ghane, N., & Sabouri, M. (2019). Personalized Educational Learning with Multi-Stakeholder Optimizations. In *Adjunct Proceedings of the ACM UMAP*. ACM.

Zheng, Y. (2019). Multi-Stakeholder Personalized Learning with Preference Corrections. In *Proceedings of the 18th IEEE International Conference on Advanced Learning Technologies (ICALT)*. IEEE.

Zheng, Y., & Pu, A. (2018). Utility-Based Multi-Stakeholder Recommendations By Multi-Objective Optimization. In *IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 128-135). IEEE.

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.

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.

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.

Zheng, Y., & Subramaniyan, A. (2019). Personality-Aware Collaborative Learning: Models and Explanations. In *International Conference on Advanced Information Networking and Applications* (pp. 631-642). Springer.

Zheng, Y. (2019). Utility-based multi-criteria recommender systems. In *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing* (pp. 2529-2531). ACM.

Pelleg, D., & Moore, A. W. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. In *ICML* (Vol. 1, pp. 727-734).

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. Journal of personality, 60(2), 175-215.