

ITM-Rec: An Open Data Set for Educational Recommender Systems

Yong Zheng

Illinois Institute of Technology, USA
yzheng66@iit.edu

ABSTRACT: With the development of recommender systems (RS), several promising systems have emerged, such as context-aware RS, multi-criteria RS, and group RS. However, the education domain may not benefit from these developments due to missing information, such as contexts and multiple criteria, in educational data sets. In this paper, we announce and release an open data set for educational recommender systems. This data set includes not only traditional rating entries, but also enriched information, e.g., contexts, user preferences in multiple criteria, group compositions and preferences, etc. It provides a testbed and enables more opportunities to develop and examine various educational recommender systems.

Keywords: recommender system, data set, context-aware, group, multi-criteria

1 INTRODUCTION

Nowadays, recommender systems (RS) have become one of the major tools in technology-enhanced learning. RS have been introduced to the education domain (Santos et al., 2011) to deliver personalized learning, recommend textbooks or informal learning materials, assist decision-making in group studies or teamwork, and provide better adaptive learning in mobile environments, etc.

Various RS were well-developed with enriched information, such as contextual variables and multi-criteria user preferences. However, the education domain may not benefit from these RS due to a lack of corresponding data sets with necessary information. In this paper, we fill this gap by announcing and releasing the ITM-Rec data set, which can be utilized as a testbed to develop and examine different types of RS (Zheng & Mobasher, 2018; Zheng, 2018; Zheng et al., 2022) in educations.

2 ITM-REC: DATA COLLECTIONS, STATISTICS AND USAGE

2.1 Data Collections

The data set was collected from a questionnaire using Qualtrics from 2017 (Zheng, 2018). The subjects were graduate students enrolled in the specialization of data management and analytics at the ITM department in Illinois Tech. The questionnaire was designed to collect student preferences on the topics of the final projects in three courses: database (DB), data analytics (DA), and data science (DS).

More specifically, students in the DA and DS classes were given a list of Kaggle data sets as candidates for analysis in their final projects. Each student was asked to select at least three data sets they liked and three that they did not like. They were asked to rate their selections by giving an overall rating and additional ratings for three criteria: App (their liking of the data's application domain), Data (their liking of the data's processing or storage), and Ease (their liking of the degree of ease in using the data for the final project). In the DB class, students were asked to build web information systems with

connections to relational databases, and were asked to rate different potential project topics (e.g., hotel booking systems, hospital appointment systems) by giving overall ratings and the same multi-criteria ratings. The questionnaire was first assigned to each student to collect individual preferences. In addition to user and group preferences, we also collected student demographic information (e.g., age, gender, marriage status) and item content features (e.g., URL of Kaggle data, title of the data, textual descriptions of the data).

Regarding the final projects, students had the option to work independently or as part of a team. If a team was formed for the final project, each group of students were asked to complete the questionnaire again as a second survey. It is worth noting that this second survey was assigned to each group and they were required to submit a single copy of the questionnaire after group discussions. Their input can be considered as group preferences on the items, rather than individual preferences.

2.2 Data Description and Statistics

We have collected the data from 2017 to 2022. The user information has been well anonymized, and data set will be released on Github¹ and Kaggle². The whole data set is composed of five major files:

- Users.csv, where we provide meta data (ID, gender, age, etc.) about 476 unique students.
- Items.csv, where we provide meta data (ID, title and descriptions) about 70 unique items.
- Ratings.csv refers to individual preferences and it contains 5,230 ratings given by 476 users over 70 items. An example is shown in Table 1. In addition to overall and multi-criteria ratings, context information such as the course (DB, DA or DS), semester (Spring, Fall) and COVID-19 lockdown periods (PRE, DUR, POS) is provided. PRE refers to the timeframe from 2017 Fall to 2019 Fall, DUR refers to the timeframe from 2020 Spring to 2021 Spring, and the POS period refers to the timeframe from 2021 Fall to 2022 Fall.
- Group_ratings.csv refers to group preferences and contains 1,117 ratings given by 143 groups.
- Group.csv describes the composition of groups. There are 143 groups, where 88, 42, 9, 4 groups have a group size of 2, 3, 4, 5, respectively.

Table 1: Example of Individual Ratings by Students

<i>UserID</i>	<i>ItemID</i>	<i>Rating</i>	<i>App</i>	<i>Data</i>	<i>Ease</i>	<i>Course</i>	<i>Semester</i>	<i>Lockdown</i>
1173	28	5	4	4	4	DA	Fall	PRE
1175	41	5	4	4	4	DS	Spring	POS
...

2.3 Data Usage

Due to the enriched information, this data can be utilized as a testbed to develop and examine various recommender systems. Below are existing or possible examples by using this educational data set:

¹ <https://github.com/irecsys/RecData>

² <https://www.kaggle.com/datasets/irecsys/itmrec>

- *Traditional RS with side information* (e.g., user demographic information and item features), or identification of *grey sheep users* in RS (Zheng et al., 2017).
- *Context-aware RS* (Zheng & Mobasher, 2018; Zheng, 2014) which adapt the recommendations to different contexts (e.g., course, semester, lockdown periods), or context suggestions (Zheng et al., 2016) which recommend contexts, rather than items.
- *Multi-criteria RS* (Zheng & Wang, 2022) which enhance recommendations by taking advantage of user preferences in multiple criteria (e.g., App, Data, Ease).
- *Group RS* (Zheng, 2018) which recommend items to each group, rather than individuals.
- *Multi-objective RS* (Zheng & Wang, 2022) which optimize multiple objectives in RS.
- *Integrated RS with multi-factors*, where researchers can build RS by integrating various factors together, e.g., integrating the context information in multi-criteria RS (Zheng et al., 2019), utilizing multi-criteria preferences towards group RS (Zheng, 2019), etc.

3 CONCLUSIONS & FUTURE WORK

In this paper, we announce and release an open data set for educational recommender systems, where enriched information (e.g., contexts, multi-criteria preferences, group compositions and preferences) are available. As a result, this data set can be utilized to develop and examine different types of educational recommender systems. We may continue the process of data collections, and release another version of this data in the next few years.

REFERENCES

- Santos, O. C., & Boticario, J. G. (Ed.). (2011). Educational Recommender Systems and Technologies: Practices and Challenges, IGI Global.
- Zheng, Y. (2018). Personality-aware decision making in educational learning. In ACM IUI companion.
- Zheng, Y. and Mobasher, B. (2018). Context-Aware Recommendations. In Book: Collaborative Recommendations: Algorithms, Practical Challenges and Applications, World Scientific Publishing, 173-202
- Zheng, Y. (2014). Deviation-based and similarity-based contextual SLIM recommendation algorithms. In Proceedings of the 8th ACM Conference on Recommender Systems.
- Zheng, Y., Mobasher, B., & Burke, R. (2016). User-oriented context suggestion. In ACM UMAP.
- Zheng, Y., & Wang, D. (2022). Multi-Criteria Ranking: Next Generation of Multi-Criteria Recommendation Framework. IEEE Access, 10, 90715-90725.
- Zheng, Y. (2018). Exploring user roles in group recommendations: A learning approach. In Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization.
- Zheng, Y., & Wang, D. X. (2022). A survey of recommender systems with multi-objective optimization. Neurocomputing, 474, 141-153.
- Zheng, Y., Agnani, M., & Singh, M. (2017). Identification of grey sheep users by histogram intersection in recommender systems. In 13th Conference on Advanced Data Mining and Applications.
- Zheng, Y., Shekhar, S., Jose, A. A., & Rai, S. K. (2019). Integrating context-awareness and multi-criteria decision making in educational learning. In ACM/SIGAPP Symposium on Applied Computing.
- Zheng, Y. (2019). Educational group recommendations by learning group expectations. In IEEE TALE.