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

Yong Zheng, Illinois Institute of Technology, Chicago, Illinois, USA

ILLINOIS TECH
College of Computing

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.

Background & Data Collections

The data set was collected by using Qualtrics from 2017 to 2022.

- Subjects: graduate students at the ITM department in Illinois Tech.
- Courses: database (DB), data analytics (DA), and data science (DS).
- Items: topics for final projects, e.g., Kaggle data sets in DA & DS courses.
- Survey 1: data collection of individual preferences
- Survey 2: data collection of group preferences

Survey 1: individual preferences

- Student demographic info, e.g., age, gender, marriage status
- Selected topics/data: at least 3 liked and 3 disliked ones for final projects
- Overall rating: 3 – 5 for liked items, 1 – 3 for disliked items
- Multi-criteria ratings: ratings on App, Data, Ease
  - App: how they like the application domain of the data sets/topics
  - Data: how they like the data processing or storage
  - Ease: how they like the overall ease if using it for final projects
- Example of rating collections: see table below

Survey 2: group preferences

- Students may eventually work on projects on their own, or via a teamwork.
- For students with teamwork, students must fill out survey 2 by a group discussions. Each group filled a single copy of survey 2.
- Questions in survey 2 are similar to the ones in survey 1. There are only questions related to rating collections.
- Eventually, we have group compositions, group preferences and individual preferences of each group member in our data sets.

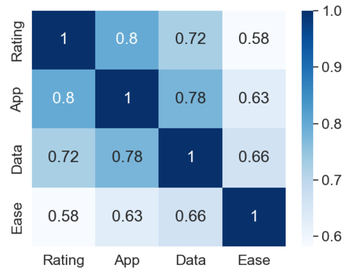
	Overall Rating			App					Data					Ease				
	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
Credit Card Fraud Detection	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Congress Trump Score	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
Homicide Reports	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○

Data & Descriptions

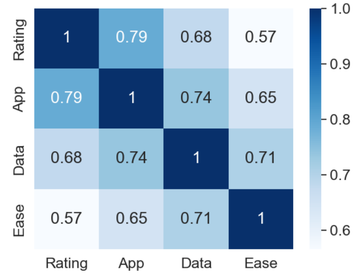
The whole data set is composed of five major files:

- **users.csv**
  columns: UserID, Gender, Age, Married
  description: Meta data about students
- **items.csv**
  columns: Item, Title, URL, Descriptions (texts or texts crawled from URL)
  description: Meta data about the topics of projects
- **ratings.csv**
  columns: UserID, Item, Rating, App, Data, Ease, Class, Semester, Lockdown
  description: Students' individual ratings on items, including the overall rating and multi-criteria ratings, as well as three contextual variables, e.g., class, semester, lockdown periods (pre, dur, pos)
  statistics: 5,230 ratings given by 476 students on 70 items, sparsity: 84.30%
- **group.csv**
  columns: GroupID, UserID
  description: the compositions of groups
- **group_size.csv**
  columns: GroupID, Size
  description: the number of students in each group
  statistics: group size: {2: 88, 3: 42, 4: 9, 5: 4}
- **group_ratings.csv**
  columns: GroupID, Item, Rating, App, Data, Ease, Class, Semester, Lockdown
  description: Ratings on items given by groups, rather than individuals
  statistics: 1,117 ratings given by 143 groups on 70 items , sparsity: 88.84%

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



Correlations on Individual Ratings



Correlations on Group Ratings

Example of Usage

The data set can be utilized for educational analysis and educational recommender systems (EdRec). Below are some examples of EdRec.

- **Traditional EdRec with side information [1]**
  user demographic info: age, gender, marriage status
  item features: textual descriptions
- **Context-Aware EdRec [2]**
  context variables: course, semester
  hint: adapter recommendations to contextual situations
- **Multi-Criteria EdRec [3]**
  criteria: app, data, ease
  hint: enhance recommendations from multi-criteria preferences
- **Group EdRec [4]**
  necessary info: group compositions and group preferences
  hint: evaluate different group aggregation and recommendation strategies
- **Multi-Objective EdRec [5]**
  hint: balance multiple objectives in educational scenarios
  examples: balance accuracy, diversity and novelty in recommendations
- **Integrating of Context and Multi-Criteria Preferences [6]**
  hint: enhance EdRec by fusing context and multi-criteria info together
- **Integrating of Group and Multi-Criteria Preferences [7]**
  hint: enhance Group EdRec by fusing multi-criteria preferences

Open Discussions

- Any other ways to take advantage of this data set?
- Any other suggestions in future data collections?
- Any similar open data sets in education?
- Any potential workshops to host a data or recommendation challenge?

Release & Downloads

kaggle

SCAN ME

GitHub

SCAN ME

References

1. Santos, O. C., & Boticario, J. G. (Ed.). (2011). Educational Recommender Systems and Technologies: Practices and Challenges, IGI Global.
2. Zheng, Y. and Mobasher, B. (2018). Context-Aware Recommendations. In Book: Collaborative Recommendations: Algorithms, Practical Challenges and Applications, World Scientific Publishing, 173-202
3. Zheng, Y., & Wang, D. (2022). Multi-Criteria Ranking: Next Generation of Multi-Criteria Recommendation Framework. IEEE Access, 10, 90715-90725.
4. 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.
5. Zheng, Y., & Wang, D. X. (2022). A survey of recommender systems with multi-objective optimization. Neurocomputing, 474, 141-153.
6. 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.
7. Zheng, Y. (2019). Educational group recommendations by learning group expectations. In IEEE TALE.