A Recommendation System for Delivering Items Customers Like

Tego Chang

Master in Interdisciplinary Data Science (MIDS)

Duke

Outline

- Exploratory data analysis (EDA)
- Methodology & design flow
- Evaluation

EDA

	user_id	quiz_type	quiz_number	question_number	item_id	rating
0	0	random	0	0	45	1
1	0	random	0	1	734	1
2	0	random	0	2	4145	1
3	0	random	0	3	3231	0
4	0	random	0	4	361	1

1. Variables of Interest, **item_id** and **rating**, are both expected delivery of the recommendation system.
2. No **characteristic** variables for item_id, but there is **user_id** to know others' preference.
3. **quiz_type** states whether the question comes from randomized or personalized by another recommendation system.

We proposed a recommendation system through
Collaborative Filtering – User-based Filtering

Duke

Methodology – User-based Filtering

Input User Selection

- Select user_id with the most observations as the input.
- Treat the personalized questions of this user as a ground truth.

Find Similar Users

- Based on the input to locate all users in the dataset who have rated the same items.

Similarity Calculation

- Apply **Pearson Correlation Coefficient** to decide the similarity between the input user and the users in the previous step.

Recommend

- Extract **all the items these most similar users have rated** and as a delivery for recommendation.

Duke

Evaluation

- Take user_id = 13613 as our input user as it has the most observation.
- Our recommendation system provides **565 items** it might like.
- Among the items recommended, 36 of them exactly match the items in the personalized dataset for user_id = 13613.
- The recommendation system achieve a **74% accuracy** on these matched items.

Duke