Individual Project Update Report

Name: Hongfei Xu
SJSU ID: 011833978

Topic:
Recommendation system in education - Online Courses Recommendation

Topic Background:
With the rapid development of information and network technology, online courses are increasingly favored by people all over the world. This project aims to design a recommendation system that recommends courses to users based on historical data or mines potential users for specific courses.
References:
https://www.sciencedirect.com/science/article/abs/pii/S0950705113001275

Data Collection:
Data has been collected from https://www.coursera.org/ , the online courses provider that is popular all over the world. The website offers dozens categories of courses and certificates, and I gathered the data of the most popular courses for each category by using web-fetching API.
The features I collected for each course are: category (string), level (string), provider (string), skills (context), course summary (context), time to spend (float number), language (string), overall ratings (float), enrolled number (float) and user ratings (int).

Approaches for analysis:
The methodologies I've implemented in progress:

1. Content-based recommendation. The summary of the courses, and the skills provided by the courses are included in the dataset as features. These context features have been used for content-based analysis in my project. I applied TF-IDF method to calculate scores for words, and matched them with user's preferences. The courses with the highest similarity score will be recommended to specific users preferred particular courses.
2. Knowledge-based recommendation. The attributes of category, course level, overall ratings, providers, lecturers, time to spend, and language may compose the constraints for users, since the conditions and demands of users are different, so I've taken these features for knowledge-based recommender system, from which the users can filter the courses based on their own needs.

The extra model I may use if dataset is verified to be applicable:

3. Collaborative-filtering method, user-based or item-based. Reviews of courses on Coursera website has been collected, but the problem I face is that not all the users submitted reviews for the courses, and most of the users only rated for one course. The data might be not enough to process user-based or item-based recommendation.

Progress and Update:
I've applied TF-IDF method for feature "skills" and "about this course". Content-based recommender is nearly completed. Function for knowledge-based recommender is in progress.