Improvement of Recommendation Accuracy by Integrating User Demographic Information

Abstract

Shuqi Wu, Master of Science in Computer Science

University of Dublin, Trinity College, 2019

Supervisor: Dr. Bahman Honari

Recommender systems has taken more and more places in our lives, recommending goods or services that satisfy the original expectations of users.Since the first research paper came up in the mid 1990s, attempts of implementing different algorithms to provide more personalized items in order to improve the prediction accuracy can be found in the literature. However, as far as I known, most papers about recommender systems so far are only based on users' previous preferences - ratings, neglecting the user demographic information, such as age, gender,occupation. Theoretically, recommendations made with respect to different users characteristics would be more accurate and personalized. That leads to my research question:

to build a recommender system that is using the user demographic information and test if this recommender systems performance would be improved by using the user demographic information.

The baseline algorithm is based on Collaborative Filtering algorithm combining with matrix factorization technique, which was used to solve the data sparsity problem.

Movielens one million dataset was chosen as the experiment dataset as it contains the complete user demographic information dataset (predictors:x) and the traget value

(predicted value:Y), ratings, that can used for the comparison between different experiments. A recommender system's performance is measured by mean square error(mse), the comparison between the predicted ratings and the true ratings. The overall experiment was processed as follows:

- Experiment 1: Train the ratings dataset without user demographic information and predict users ratings towards different movies. Use the predicted ratings for the test set to get the overall mse as the evaluation result.

- Experiment 2: Train the datset with user demographic information and predict user ratings towards different movies. Use the predicted ratings for the test set to get overall mse as the evaluation result. However, with the complexity of user demographic information, Convolution Neurall Network was added to train the textual data.

- Experiment 3: Train the datset with two dataset in Experiment 1 and Experiment 2 using User-based Collaborative Filtering algorithm, one added experiment to prove that the improvement in Experiment 2 was resulted from the user demographic dataset. The precision was used as the evaluation matrix.

The mse results of experiment 1 and experiment 2 show that the recommender systems performance by using user demographic information(0.661) implemented with Convolution Neural Network has improved around 15% compared to the experiment without user demographic information(0.770). The Experiment 3 was designed to prove that the improvement of experiment 2 was made by user demographic instead of the Convolution Neural Network algorithm. The precision of the test with user demographic information(46.4%) is 17% higher than that without user demographic information(38.5%). Thus, the above experiments show that the use of user demographic information can improve the overall recommender systems performance.