
Extended Abstract of Task 4 For NeurIPS Education Challenge 2020

Yuren Zhang, Yufei Wu, Jie Huang and Songtao Fang

Anhui Province Key Laboratory of Big Data Analysis and Application, School of Computer Science and Technology & School of Data Science, University of Science and Technology of China, {yr160698, wyf1997, jiehuang, songtao}@mail.ustc.edu.cn

Abstract

In the task4, we mainly integrate a variety of diagnosis models cooperating many targeted preprocessing strategies to model students' knowledge level and further predict. The major limitation is the lack of records (only 10) for each student in inference while the average records in training data more than 280. Meanwhile, we find some redundant or low-importance in provided data on the basis of data analysis. To address above limitations, we propose the following solutions: (1) We utilize the User-based collaborative filtering (UserCF) method to initialize the students vectors while performing inference, which alleviates the problem of cold start and lack of students interaction records, (2) For the purpose of robustness, we compress the total number of subjects the questions belong to by combining or removing the unimportant ones. (3) In order to improve the prediction accuracy, we integrate FM, NeuralCD and many other models which have proved their efficiency in recommender system and cognitive diagnosis areas. (4) Finally, we try various Adaptive Testing strategies such as MFI, KLI, and MAAT in order to accelerate the convergence. In the end, our solutions achieve scores of 0.729 and 0.735 on the public and private leaderboards respectively, which shows excellent prediction performance and generalization.