**Project Title: “Students Dropouts Prediction in Massive Open Online Courses”**

# Abstract

Massive Open Online Courses (MOOCs) are attracting unprecedented interest and have been growing rapidly around the world during the coronavirus lockdown. In 2020, enrolments in Coursera were reported to be 640% higher, growing from 1.6 to 10.3 million. However, MOOCs courses have been criticized for the low completion ratio and high number of dropouts. To counter this, researchers have built machine learning algorithms to analyze student data and predict the level of dropout. However, to date, these algorithms have only achieved an accuracy rate of 80%. Our aim is to investigate the dropout prediction problem. Additionally, we will also propose that the learning sophisticated feature interactions behind the user learning activity are critical to maximizing Click-through rate (CTR), and thus improving user retention in MOOCs. A dataset from XuetangX, one of the largest MOOCs in China, was acquired, and a systematic study for dropouts' problems in MOOCs was conducted. We investigated two variants of the CFIN Model with a personalized attention mechanism to model and predict user dropout behaviors. The core of the approach is a course representation model and a user representation model. The CFIN model utilizes the context smoothing technique to smooth feature values with different contexts and use an attention mechanism to combine user and course information into the modeling framework. In addition, the same course can have different dropout rates for different user groups, it was proposed to build a personalized attention network that exploits the embedding of a user ID and embedding of course ID to generate the query vector for the course and learning-activity level attentions. Furthermore, I will also investigate the DeepFM model, which combines the power of factorization machines for recommendation and that of deep learning for feature learning in a new neural network architecture but has not yet been explored for the dropout prediction problem. Experiments on the dataset show that the two proposed variants have similar performance with the CFIN model, predicting dropouts with a 93% accuracy, and it achieves better performance than SVM, RF, and ensemble modeling techniques. Furthermore, then conducted comprehensive experiments to demonstrate the effectiveness and efficiency of the DeepFM model in user CTR for dropout prediction problems.

