

Big Data and Massive Online Open Education

Weipeng Yang

School of Education, Indiana University Bloomington

201 N Rose Ave

Bloomington, Indiana 47405

yang306@uemail.iu.edu

ABSTRACT

Massive Open Online Education(MOOC) often refers to a kind of online course that emphasizes free and unrestricted access via the internet. This paper will focus on how to collect such data as length of study, learning time of the day, preferred courses and other learners' behaviors to analyze and make predictions for future course designers and online teaching platform administrators.

KEYWORDS

big data, MOOC

1 INTRODUCTION

Nowadays people in the field from K-12 to institutions of higher education have witnessed a great shift from traditional classroom teaching to distance learning. These online courses are based on Learning Management Systems (LMSs). The more advanced a learning management system could be, the better they could collect a variety of data from students' activities and performances. With this amount of data, educators and researchers could utilize data mining and data visualization techniques to generate synchronous feedback and to keep-track of students' progress with higher efficiency. The source of the data could be the average time a student is learning a certain course, test score of a quiz, response rate in an online discussion session, etc.

2 BIG DATA

Big data are sometimes considered as the amount of data that could be available to an organization, in which contains a gigantic amount of information that would render itself too complex for any household data-processing software to manage. Not only big data emphasizes the amount of data, it also all calls for the need of real-time data and a wide source of data. As for the field of education, big data provide more direct evidence-based approach to learning and could allow researchers to see the difference of students nationwide or even worldwide. Analyzing these data could help LMS providers to make strategical improvements or let online teachers provide personalized tutoring to those student in need.

3 EDUCATIONAL DATA MINING IN MOOC

Educational Data Mining (EDM) refers to the process of analyzing various kinds of data on diverse levels of education (for instance, data could be collected from online students to decision-makers of the MOOC corporation) via a variety of techniques and tools. What is collected includes time, sequence and context of the course being taught. It could be easily inferred that EDM requires cooperation of various subjects such as statistics, artificial intelligence, machine learning, etc. With these interdisciplinary means data could be

processed. EDM aims to predict the academic performance of a student, evaluate student learning in the context of LMS, improve instructional sequences and evaluate add-on software that could provide additional help with the LMS. Pioneers of this field are researching how to improve the modeling of student performance, teaching domain and LMS properties and characteristics. They are also interested in providing students with diverse needs different track of learning courses accordingly. It is concluded that EDM usually requires five means to analyze educational data: prediction, clustering, relationship mining, distillation of data and discovery via models. Prediction stresses that students' academic performance will be analyzed via their behavior in online learning. Clustering means that by sensing such specific characteristics as preference of learning materials or performance styles, the students will be grouped according to the elements mentioned above. Moreover, these resources could be recommended to learners with similar needs. Relationship mining is the most mentioned method in EDM. It focuses on figuring out hidden relationships with such variables as teaching and learning strategies, students' performance in online environment and students' interactions. Distillation of data for human judgement concentrate on ways to filter out most important data in a cluster so that researcher could figure out structures in the data quickly. Discovery via models, the last method, focuses on utilizing existing model to analyze newly collected data. To better utilize EDM, the MOOC cooperation shall establish a data structure first by determining the need of its users and their learning goals as well as the source of the data. Then they shall start defining certain variables and start creating a model or choose from an existing one. In the end they could start using this model to predict students' preferences and make modification accordingly. It is stated in some research that after using EDM to collect more information, MOOC's learning outcomes are improved and course tutors could cater to students' needs more efficiently.

4 LEARNING ANALYTICS

Other than computer science and statistics, Learning Analytics (LAs) are rooted on a wider spectrum of subjects such as sociology and psychology. Those who apply LAs wish to create a learning environment for teachers in which each student's learning need will be satisfied to the greatest extent and they could also choose their own learning tracks via their own learning habits. LAs could also enable facilitator of MOOC to distribute educational resources with better decision-making mechanism, providing feedback to students and help at-risk students (refer to those who haven't participated in learning activities for a long time or those who do not have ideal performance in quizzes). To reach these goals mentioned above, LAs also involve a vast variety of data, from students' learning habits, assignments collected, social interaction online, threads

on discussion forums to generate students' progress and identify those who might be at-risk. With enough data collected, LAs could also be utilized to determine the overall structure of the course, students' learning objectives and the sequence of learning contents. Decisions are often based on models with multiple dimensions such as students' experience, knowledge, their preferred sequence of learning. LAs consist of three steps: Data collection and processing, analyzing data and action and data post processing. It can be concluded that LAs could be efficient for all level of users, tutors and decision-makers in MOOC however it also face lots of challenges as decision makers will need to determine what kind of data to collect and they need to connect separate collected data together via specific algorithm to gain a holistic view of inner connections between.

5 CONCLUSIONS

With the rapid iteration of software and LMSs, learners nowadays could easily gain access to massive amount of learning materials at almost every corner of the world. MOOC brings great flexibility to learners so that they could choose their online instructors, their sequence of learning and their learning materials. It also brought instructors and students closer than ever so that they could interact more frequently, thus enabling a more dynamic atmosphere in learning. Challenges also arises as the numbers of students grow exponentially, it could be more difficult for teachers and managers to keep track of each student's status and provide help accordingly. Thanks to the introduction of big data in this field, decision maker of MOOC could evaluate and investigate students' status more easily and could develop more learning strategies. It can be envisioned that soon big data will relieve more burden on teachers' shoulders.

NO CITATIONS [1], all lables were missing so i put some placeholders in that need to be fixed

REFERENCES

- [1] Terry Anderson. 2008. *Theory and practice of online learning*. Edmonton, AB, Canada: AU Press., Chapter Teaching in an online learning context, 343–366.

6 BIBTEX ISSUES

Warning--empty address in 1

(There was 1 warning)

7 ISSUES

7.1 Uncaught Bibliography Errors

Two few references.

7.2 Formatting

Incorrect number of keywords or HID and i523 not included in the keywords

7.3 Writing Errors

Spelling errors