

Big Data in Education and Standardized Testing

Huiyi Chen
Indiana University
huiychen@indiana.edu

ABSTRACT

This paper provides an insight of how big data is related to standardized testing. It will also describe some applications of big data in standardized testing, and how they function.

KEYWORDS

Big Data, standardized testing, education

1 INTRODUCTION

Big data has been in our daily life for quite a long time. It is not surprising when we hear the word "big data" any more. We have also heard stories, for example, Target used their data to predict the personal sale items. However, it seems like big data is still a new thing in education, or at least, people have not been making use of big data in education to improve the standardized testing environment and the students' learning experience. In this report, we will look into some questions and concerns that have been raised in the past of the similar topic and some applications that have been in place for the development of big data in the area of standardized testing and education administrative at a high level.

2 CURRENT ENVIRONMENT

The current environment of big data in education is controversial. Even without the concept of big data, standardized testing has already been a controversial topic in the fields due to the pressure from result efficiency and usability. Some people says that standardized tests are useful and effective predictors of academic measurement in universities. [2]At least until now, there is no other better measurement to fairly compare students among themselves. Ferreira also mentioned in the article that he believes standardized test scores are more accountable on the top end scale than on the lower. The reason behind that is that higher scores normally indicates one's capability. However, lower scores do not always mean one's incapability. [2] Other people argue that the problems with these standardized testing are not used properly. While 1410 and 1390 could mean absolutely nothing in difference, some admissions offices would do the cutoff at 1400, which make the result unfair. Additionally, a score starts with 14 looks just better than that starts with 13.

3 BIG DATA IN STANDARDIZED TESTING

With the conflicts that exist in the society, there is a strong need for big data to get involved in order for the system to generate useful data to balance with standardized tests, easing the opponent for that and to make more fair decisions. While big data has been incorporated in many different fields to better decision making process such as business and pharmaceutical firms, research in education learning analytic has been very limited to only examination indicators.

4 APPLICATIONS IN TESTING

There are increasing number of big data applications in the education and standardized testing fields for learning these tests and predict performance and measure proficiency. One area that has been focusing on is knowledge discovery (KDD). It is an interdisciplinary area that focus on methodologies of identifying and retrieving meaningful patterns from large data set, for example, the College Board testing data. KDD draws research in statistics, pattern recognition, machine learning and etc. to analyze testing data to make sure the tests are focusing on the important content but not uncommon knowledge that will potentially lower students' scores without proficiently indicating their abilities[1]. There is a variety of method we can use to conduct data mining research in education. Within these different approaches, different tasks can be done.

4.1 Clustering algorithms

Clustering algorithms is a useful approach to access group data items which based on defined logical relationships[1]. The goal of the approach is to maximize the inter-cluster similarity and minimize intra-cluster similarity. With this approach, we will be able to group testing questions based on their commonality and relationships with other questions to determine whether a question is proficient on testing or not.

4.2 Classification

Classification is another approach in data mining research that helps to classify test data into predetermined categories. This approach is often referred to machine learning. These algorithms can help us learn form a large set of pre-classified data, for example, reading, vocabulary, and math are three different categories that we can analyze in SAT [1]. To dig even deeper into each categories, we can even categorized different types of math questions such as algebra, calculus, and etc.

4.3 Association

Association approach is also a kind of algorithm to extract the characteristics of some data within a set. These are driven by rule-based algorithms, which usually examine correlation between variables. This approach is to study the frequency associated with a data set [1].

4.4 Factor analysis

One of the most common approach in education is Factor analysis. It is to find the different variables that could be grouped together, or split the set of variables into a set of factors [1]. "Within the education field, factor analysis includes algorithms such as principal component analysis and exponential-family principal components analysis and used for dimensional reduction, including

in pre-processing to reduce the potential for over-fitting and to determine meta-features. Factor analysis has also been used in learning environments to develop predictive models.

5 APPLICATIONS IN EDUCATION ADMINISTRATION

Not only big data can be used in analyzing standardized testing data, it can also be applied to many higher education administrations including admission processing, student performance monitoring in colleges and etc[3]. With the similar data mining methods, big data is going to make managing students' and staff's information more convenient with more significant and useful data. While in the past, we need to go to the registration office to register for classes, or the staff need to take a long time to schedule around classrooms, professors, times, and sections, we can now use something called course management or learning management system (CMS) [3]. This system gives us immediate report of student activities and course arrangements. It would also record and analyze students' homework and grades to determine what would need to be done to bring up students' grades or generate a class report on testing to compare with previous classes for the professor to make decisions on potential class materials. Nowadays, a CMS becomes critical for institutions to collect data to process important learning analytic application.

6 CONCLUSIONS

Competition in education will continue to be increasingly strong due to the growing population of students all over the world[4]. Both standardized tests and institutions are under pressure of responding to better the education systems with quality of learning programs and testing outcomes. More and more organizations and testing center are utilizing data to make better decisions about learning programs and questions distribution [4]. While higher education is largely rely on a political and economic situation of a country, big data will help these institutions and testing centers to maximize their ability to utilized existing data to generate the most information. Data-driven decision making is already in use to help institutions and testing centers to build strategies. Big data will continue to prove its work in education and standardized testing by making them more efficient and useful in the future.

REFERENCES

- [1] Ben Daniel. 2015. Big Data and Analytics in Higher Education: Opportunities and Challenges. *British Educational Research Association* (2015).
- [2] J. Ferreira. 2014. Itfis Not fiBig Data vs. Teachersfi fi!? Itfis fiBig Data vs. Standardized Testsfi. (2014). <https://www.knewton.com/resources/blog/ed-tech/big-data-vs-standardized-tests/>
- [3] Anthony G. Picciano. 2012. The Evolution of Big Data and Learning in Analytics in American Higher Education. *Journal of Asynchronous Learning Networks, Volume 13: Issue 3* (2012).
- [4] Jeffrey Selingo (Ed.). 2017. How Colleges Use Big Data to Target the Students They Want. (2017).

7 BIBTEX ISSUES

Warning-no number and no volume in Daniel2015

Warning-page numbers missing in both pages and numpages fields in Daniel2015

Warning-no journal in Ferreira2014

Warning-no number and no volume in Ferreira2014

Warning-page numbers missing in both pages and numpages fields in Ferreira2014

Warning-no number and no volume in Picciano2012

Warning-page numbers missing in both pages and numpages fields in Picciano2012

Warning-no journal in Selingo2017

Warning-no number and no volume in Selingo2017

(There were 9 warnings)

8 ISSUES

DONE:

Well done!