Chapter 87

Resolving the Paradox of Overconfident Students With Intelligent Methods

Denis Smolin
American University in Bosnia and Herzegovina, Bosnia

Sergey Butakov
Concordia University, Canada

ABSTRACT

The chapter presents a case study of using data mining tools to solve the puzzle of inconsistency between students' in-class performance and the results of the final tests. Classical test theory cannot explain such inconsistency, while the classification tree generated by one of the well-known data mining algorithms has provided reasonable explanation, which was confirmed by course exit interviews. The experimental results could be used as a case study of implementing Artificial Intelligence-based methods to analyze course results. Such analyses equip educators with an additional tool that allows closing the loop between assessment results and course content and arrangements.

INTRODUCTION

Introduction briefs the reader into the problem of the objective course material assessment, flaws in the standard statistical solutions to this problem, and the potential areas where some approaches from Artificial Intelligence (AI) domain could be useful.

Problem of student knowledge and skills assessment is as old as the education itself. All educators need a tool to check efficiency and effectiveness of their teaching materials and approach used in the classroom. Various test styles has been around for centuries but they always raise questions if they are appropriate for checking student comprehension of the subject matter.

The area of student testing is much wider than one can assume. A test, as a measurement tool, and a testing algorithm, as a method to interpret test results, are very generic models of any measurement. It is applicable not only for measurement of student skills and knowledge, but also for almost any other

activity. Testing consists of two major steps - data collection and interpretation. Interpretation implies some decision making, which is a large domain where methods of AI play important role. On one hand, classical testing is based on mathematical statistics and AI uses many statistical methods. On the other hand, modern tests for knowledge and skills are much more complex, as compared with their classical ancestors. AI equips the interpretation phase of testing with appropriate processing methods. These methods give fruitful results that are impossible to get with classical algorithms. This paper demonstrates an example of such an AI method.

Interpretation phase of the testing process is covered by two main theories in the field:

- **Classical Testing Theory (CTT):** Which is closely related to mathematical statistics. For example, Spearman correlation was initially developed for psychological testing (Traub, 1997).
- **Item Response Theory (IRT):** A latest theory that expands CTT and is claimed to be more precise for the case of computer-based testing process (Thissen & Mislevy, 2000).

Currently, classical testing theory, which is over 100 years old, is the most popular one. The question of comparative efficiency of these theories is ambiguous and requires detailed examination. Instead of attempting to invalidate the entire theory, the more productive approach is to investigate possible theory inconsistencies - so-called paradoxes, which consist of the cases the theory cannot explain. While applying CTT to student knowledge and skills one can simply find two paradoxes: a student with bad knowledge getting higher exam score and a student with presumably good knowledge of subject matter failing a major test.

CTT answers these questions with the concept of quality, defined with the statistical evidence. A test is of high quality, when it has high validity and reliability. There are different methods for validity and reliability evaluation. The majority of these methods are based on correlation coefficients, which obviously require a representative sample. In practice, it could be difficult to get enough data to prepare such a sample. A test with proved statistical characteristics is called standardized and considered an unbiased measurement instrument. This test is correct for almost all students. All other tests assumed to be tests with indefinite quality. In some cases, it can also be considered a paradox, because the quality of tests can be proved with some other methods. Additionally the paradox may occur, when a student with presumably good knowledge fails high quality tests. It rarely happens and CTT has eliminated these cases from its scope by arguing that there is not enough information to evaluate them. For example, we may be unaware of all the personal circumstances that happened to the student during the exam time.

Standards provided by the American Educational Research Association (AERA) recommend certain number of implementations to be performed to verify quality of a test or a question (AERA, 1999). In many cases, educators do not have enough data to perform such a comprehensive statistical analysis. The underlying reasons for that could vary from cases where the number of students in a class is limited to very dynamic subject matters where noticeable changes in course elements must be implemented in every teaching cycle. Lack of statistics makes it difficult for educators to setup a proper verification process for course elements. As a solution they often use "hand-made" tests of undefined quality (Miller, 2009). In such cases, teachers cannot improve course elements because they have no adequate tools to perform verification. Without such a verification process, any of participants of the educational process may raise questions about the quality and relevance of the course and its elements.

Related Content

Artificial Intelligence in the Consumer Behavior Process in Business
Albérico Travassos Rosário (2024). *Enhancing and Predicting Digital Consumer Behavior with AI (pp. 39-73).*
www.irma-international.org/chapter/artificial-intelligence-in-the-consumer-behavior-process-in-business/347195

Another Kind as First Person
 (2019). *Human Behavior and Another Kind in Consciousness: Emerging Research and Opportunities (pp. 53-59).*
www.irma-international.org/chapter/another-kind-as-first-person/226488

Fuzzy Joint Gaussian-Impulsive Noise Removal Using Joint Distribution Modelling in Sparse Domain
V. V. Satyanarayana Tallapragada, D. Venkat Reddyand Suresh Varma K. N. V. (2022). *International Journal of Fuzzy System Applications (pp. 1-15).*
www.irma-international.org/article/fuzzy-joint-gaussian-impulsive-noise-removal-using-joint-distribution-modelling-in-sparse-domain/312216

AI-Driven Speech Sign Converter for Business: Improving Communication and Collaboration
Harddik Bafna, Gagan Parashar, Farhan Ansari, A. Krishnamoorthyand N. Prabakaran (2024). *Digital Twin Technology and AI Implementations in Future-Focused Businesses (pp. 49-60).*
www.irma-international.org/chapter/ai-driven-speech-sign-converter-for-business/336449

Towards a New Model of Storage and Access to Data in Big Data and Cloud Computing
Houcine Matallah, Ghalem Belalemand Karim Bouamrane (2017). *International Journal of Ambient Computing and Intelligence (pp. 31-44).*
www.irma-international.org/article/towards-a-new-model-of-storage-and-access-to-data-in-big-data-and-cloud-computing/187066