

Difficulty score for detecting erroneous data points in intensive care data based on Item Response Theory

(Master Thesis)



MUHAMMAD ATIQUE
NAZAR

Motivation

The incorporation of Machine Learning (ML) techniques in the health sector is undergoing rapid transformation. ML algorithms help in analysing vast amounts of complex data, extracting insights and predicting outcomes. The success and efficiency of ML techniques heavily rely on the quality of datasets. Thus, enhancing the quality of the datasets becomes a crucial aspect. Error detection in datasets forms a necessary foundation for improving the quality of datasets, as it can significantly reduce faulty data and outliers. This not only improves the accuracy of the result but also allows for unbiased evaluation of different models and techniques. The effectiveness of error detection can differ depending on the complexity and nature of the data. To address this, a difficulty score can be assigned for error detection which aids in ML models and techniques to reach their full potential and provide exceptional results.

State of the Art

Under a thesis titled “Entwickeln eines Schwierigkeitsscores für Datensätze bei der Fehlererkennung in intensivmedizinischen Daten” by Tobias Klinkhart at Informatik 11, a method for calculating difficulty score for detecting erroneous data points has already been implemented. Exploring a different engaging approach, such as Item Response Theory (IRT) in calculating the difficulty score, is worth considering. IRT is a framework being used for analysing and scoring the assessments in education, psychology and medicine. IRT serves as a statistical approach, that allows to measure the relationship between an individual’s cognitive abilities and their responses to test items. Recently, IRT has been used for a variety of applications e.g. to examine the difficulty of estimating mortality from intensive care data.

Objective

The main objective of this master’s thesis is to develop an algorithm by using Item Response Theory (IRT), which assigns a difficulty score for error detection, on a scale of 0 to 1 (0% to 100%), based on the annotated intensive care data. In addition, an analysis is to be made to identify the data characteristics that particularly influence the score. The developed difficulty score is to be evaluated against the difficulty score from Klinkhart’s thesis.

Planned Procedure

The first step will be a comprehensive literature review to identify and discover possible solutions that implement IRT in Python and can be applied to the annotated datasets to develop an algorithm. Subsequently, a difficulty score will be implemented using the identified solution on the datasets annotated by two physicians. Finally, the developed difficulty score will be compared with the existing difficulty score from Klinkhart’s thesis.