# Evaluation Suite For Algorithms For The Detection Of Erroneous Data Points

(Masterarbeit)

JIALEI ZHAO

## Motivation

Anomaly-detection-analysis-system (AnDAS) is a data analysis software which can be used to detect data anomalies in ICU datasets. A variety of detectors are included that can analyze the imported patient data and give predictions of erroneous data points. Evaluating classifier performances is a crucial step in this classification problem. Although many evaluation methods already exist, such as accuracy, ROC and AUC curves, they may turn to be inefficient in some classification cases. Because some AI Agents, despite having identical average performance metrics, can demonstrate diverse performance trends when examined at the individual level. In this paper, more methods capable of evaluating the performance of erroneous points detectors in more detail and from more perspectives will be applied.

## State of The Art

For making up for the inefficiency of aggregate metrics such as accuracy, many new evaluation methods have been proposed. For example, identifying relevant and irrelevant features in AI systems is crucial for grasping performance patterns. This distinction enables us to gauge the system's capabilities across different levels of difficulty or circumstances. Beyond theoretical analysis, several open source projects provide methods for automatically evaluating machine learning agents. „Behavior suite" package for reinforcement learning could isolate core capabilities of agents with targeted 'unit tests' and measure properties that can be observed in the environment. It provides not only a snapshot of agent behavior, but also contains a complete description of every experiment, summary scoring and in-depth analysis of each experiment. These methods give us a deeper understanding of the behavior of the algorithms and the opportunity to improve them.

## Target Setting

Different evaluation methods will be applied to the detectors in the AnDAS software. These new methods should help to uncover the performance patterns of the detectors compared to the traditional evaluation methods. The evaluation results can be used to improve the performance of the detectors in different aspects. And when facing diverse datasets, the comprehensive assessment would be useful in choosing the appropriate detector to greatly enhance the capability of error-point prediction.

## Planned Working Steps

First, a literature search on evaluation criteria and forms of presentation of evaluation results will be performed. After organizing and summarizing the literature and open sources searched, a variety of suitable evaluation techniques are applied to the performance of the error point detector in NDAS software. The next step is to develop a survey to ask AnDAS users whether the implemented procedures each improve the understanding of the performance of the classification algorithm. Conducting and assessing the survey marks the final phase of the process.

Betreuer: Alexander Kruschewsky
04.02.2024

Informatik 11
Embedded Software
RWTH AACHEN UNIVERSITY