

Entwickeln eines Schwierigkeitsscores für die Fehlererkennung in intensivmedizinischen Daten

(Bachelorarbeit)



TOBIAS KLINKHART

Motivation

Es gibt gute Gründe, einem Datensatz für ein Problem einen Schwierigkeitsgrad zuzuordnen. Nicht nur kann so eine unausgewogene Auswahl von Datensätzen beim Training von Modellen vermieden werden, sondern auch die Performance von bestehenden Modellen genauer evaluiert werden. So kann das Verhalten eines Modells in Bezug auf die Schwierigkeit des Datensatzes Aufschluss über mögliches Fehlverhalten oder Defizite dieses Modells aufzeigen. Dies wiederum kann zu einer höheren Transparenz und Verständnis für die Ergebnisse des Modells führen. Gerade bei der Fehlererkennung in intensivmedizinischen Daten ist es wichtig zu wissen, wie verlässlich etwaige Ergebnisse sind. Hierbei kann ein Schwierigkeitsscore für die Fehlererkennung in solchen Daten helfen.

Stand der Technik

In der Literatur werden hierzu verschiedene Ansätze vertreten. So kann die Schwierigkeit eingeschätzt werden, indem die Problemlösung durch Menschen ausgewertet wird (z.B. Zeit, die für die Problembewältigung benötigt wird, oder die Annotation eines Datensatzes durch Experten). Weiterhin existieren Ansätze, die aus der Performance von AI und Modellen auf dem Datensatz auf die Schwierigkeit schließen. Es gilt: Je schwieriger der Datensatz, desto schlechter ist die erwartete Performance. Oft wird dabei auf IRT (Item-Response-Theory) verwiesen, bei der die Schwierigkeit des Datensatzes von der Wahrscheinlichkeit abhängt, dass mehrere Modelle (Respondents) einen Datensatz (Item) richtig klassifizieren.

Zielsetzung

Es soll mittels annotierter intensivmedizinischer Daten ein Algorithmus entwickelt werden, der einen Schwierigkeitsscore (0%-100%) für solche Daten für die Fehlererkennung zuweist. Darauf aufbauend soll es möglich sein, auch nicht-annotierten intensivmedizinischen Datensätzen einen Schwierigkeitsscore zuzuweisen, dessen Aussagekraft sich auf die Fehlererkennung beschränkt. Dieser Schwierigkeitsscore wird mit Hilfe von zwei bereits vorhandenen Fehlererkennungsmethoden evaluiert. Insbesondere die Performance in Abhängigkeit zum Schwierigkeitsscore steht im Vordergrund. Abschließend folgt eine Analyse, ob/welche Aspekte der Datensätze diesen Schwierigkeitsscore signifikant beeinflussen.

Geplante Vorgehensweise

Für die Konzeptionierung des Algorithmus wird eine ausgiebige Analyse der annotierten Datensätze vorgenommen. Bei gleichen (von verschiedenen Personen annotierten) Daten werden mögliche Abweichungen der Annotationen als Indikator für Unsicherheit bzw. Schwierigkeit in Fehlererkennung in den Algorithmus einbezogen. Nach der Implementierung des Algorithmus in Python folgt die Evaluation aus der Zielsetzung (insbesondere der signifikanten Zusammenhänge innerhalb der Datensätze).