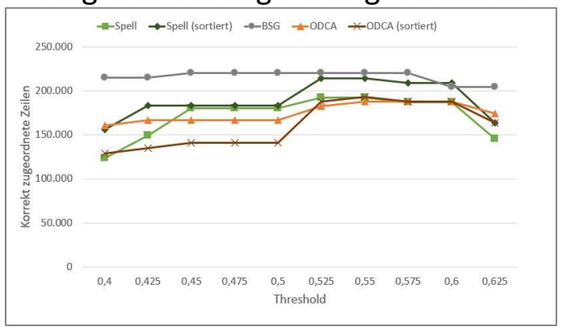
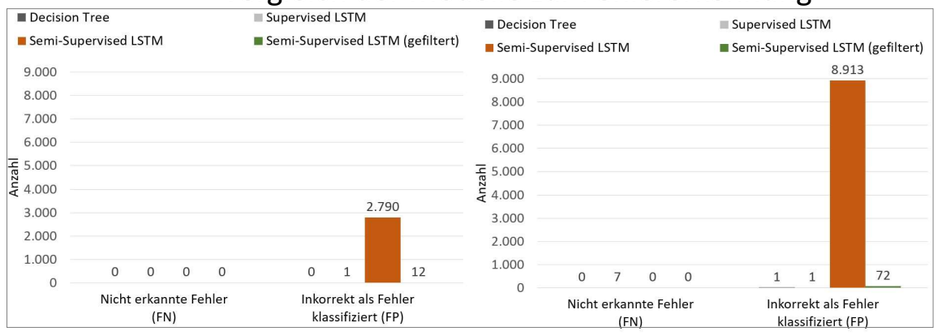
Automatisierte Fehlererkennung einer Microservice-Anwendung basierend auf Log-Dateien

Valerie Restat

Vergleich der Log-Parsing-Verfahren



Vergleich der Modelle zur Fehlererkennung



Motivation

Log-Dateien dokumentieren das Verhalten einer Anwendung oder eines Systems, was deren Analyse zu einem Schlüsselfaktor für die Sicherheit, Stabilität und Nutzbarkeit eines Systems macht. Dabei gibt es in vielen Applikationen in den Log-Dateien Fehler oder Warnungen, bei denen kein Handlungsbedarf besteht. Die hohe Anzahl solcher Meldungen im Vergleich zur geringen Menge an tatsächlichen Fehlern führt dazu, dass sich die Analyse sehr aufwändig gestaltet und viel Zeit kostet. Diese Problemstellung findet sich auch bei der ORDIX AG bei einem Mircoservice für interne Zwecke wieder.

Ziel und Vorgehensweise

Ziel dieser Arbeit war eine automatisierte Fehlererkennung basierend auf Log-Dateien. Zu diesem Zweck musste zunächst eine geeignete Methode zur Vorverarbeitung entwickelt werden, um aus den unstrukturierten Log-Dateien Features zu gewinnen, die für den Einsatz von Machine-Learning-Modellen zur Fehlererkennung verwendet werden können.

Hierfür wurde aus einer Log-Meldung der konstante Teil, der sogenannte Log-Key, extrahiert. Der BSG-Algorithmus (Basic Signature Generation) konnte die meisten Log-Zeilen korrekt zuordnen und wurde deshalb als Verfahren für diese Arbeit gewählt. Anschließend wurde die Fehlererkennung betrachtet und zu diesem Zweck diverse Modelle untersucht, die auf den vorverarbeiteten Daten aufsetzen. Hierfür wurden zunächst die Anforderungen analysiert und in Abhängigkeit derer Methoden der Supervised-, Semi-Supervised- und Unsupervised-Fehlererkennung unter-sucht. Dabei hat sich gezeigt, dass die Unsupervised-Verfahren nicht für den Anwendungsfall dieser Arbeit geeignet sind, sondern in erster Linie für Anwendungen, in denen die Daten mittels numerischer Werte verglichen und so Fehler identifiziert werden können. Dies ist beispielsweise der Fall, wenn die Auslastung des Systems eine Rolle spielt oder ein ungewöhnlich hohes Auftreten einzelner Meldungen. Im Rahmen dieser Arbeit war dies jedoch nicht gegeben, weshalb jene Verfahren in der Umsetzung nicht berücksichtigt wurden.

Ergebnisse

Für die Evaluierung wurden die nicht erkannten Fehler sowie die Meldungen, die inkorrekterweise als Fehler klassifiziert wurden, berücksichtigt. Die Supervised-Verfahren erzielten sehr gute Ergebnisse, insbesondere Mithilfe des Decision Tree konnten fast alle Meldungen korrekt klassifziert werden. Im Bereich der Semi-Supervised-Anomaliedetektion konnten mittels eines LSTM alle Fehler erkannt werden, das Modell wies jedoch, im Vergleich zu den Supervised-Methoden, eine höhere Anzahl an Meldungen auf, die inkorrekt als Fehler klassifiziert wurden. Die Anzahl war jedoch akzeptabel für den Aufwand einer manuellen Analyse.

Fazit

Sowohl der Supervised- als auch der Semi-Supervised-Ansatz kann zur automatisierten Fehlererkennung auf Basis der Log-Dateien des Microservice eingesetzt werden. Bei den Modellen des Supervised Learnings besteht der Aufwand dabei in der manuellen Vergabe der Labels, während er bei der Semi-Supervised-Fehlererkennung in der manuellen Analyse der Modellergebnisse liegt.

h_da
HOCHSCHULE DARMSTADT
UNIVERSITY OF APPLIED SCIENCES

Kontakt
Valerie Restat

Fachbereiche Informatik & Mathematik
und Naturwissenschaften

Referentin
Prof. Dr. Uta Störl

Korreferentin
Prof. Dr. Antje Jahn

ORDIX AG