

ABSTRACT

Diagnostic Trouble Codes (DTCs) in sequential vehicle data indicate technical problems. However, diagnostic systems in the vehicle do not log all DTCs. In addition, not every DTC is time-stamped. In this case, it is not possible to trace when a particular problem occurred. Analysing such sequential data is time- and resource-consuming and requires expertise in this area. This effort can be reduced using Automated Machine Learning (AutoML). This thesis addresses the question of the contribution of a selected AutoML solution in the analysis of sequential vehicle data.

This work aims to first provide an overview of the existing AutoML solutions for time series data through the literature review. Next, a selected solution for time series classification will be integrated into an existing Data Science Platform – Ontology-based Meta AutoML. Finally, the Platform will be applied to the classification of sequential vehicle data.

Based on the literature review, the deep learning-based Mcfly AutoML for the time series classification was selected and integrated into the existing Data Science Platform. Finally, the Platform was applied to the classification of sequential vehicle data. In the experiment on a partial dataset, an F1 Score of 0.70 was obtained. Although 94% of DTCs were detected using AutoML (Recall of 0.94), the false positive rate was high in the experiment (Precision of 0.57).

By using the Data Science Platform, the effort to create machine learning models for time series classification and hyperparameter optimization can be reduced. The AutoML solution integrated into the Platform is based on deep learning network architectures. Therefore, a longer runtime for machine learning pipeline optimization can be expected.

Keywords: Automated Machine Learning, CAN bus, Deep Learning, Diagnostic Trouble Code, Ontology-based Meta AutoML, Time Series Classification

KURZFASSUNG

Diagnose-Fehlercodes (DTCs) in sequenziellen Fahrzeugdaten weisen auf technische Probleme hin. Diagnosesysteme im Fahrzeug protokollieren jedoch nicht alle Fehlercodes. Außerdem wird nicht jeder Fehlercode mit einem Zeitstempel versehen. In diesem Fall ist es nicht möglich nachzuvollziehen, wann ein bestimmtes Problem aufgetreten ist. Die Analyse von solchen sequenziellen Daten ist zeit- und ressourcenaufwendig und erfordert Expertise in diesem Bereich. Dieser Aufwand kann mit Hilfe von Automated Machine Learning (AutoM) verringert werden. Die vorliegende Arbeit befasst sich mit der Frage, welchen Beitrag eine ausgewählte AutoML-Lösung bei der Auswertung von sequenziellen Fahrzeugdaten leistet.

Ziel der Arbeit ist es, zunächst durch Literaturrecherche einen Überblick über die existierenden AutoML-Lösungen für Zeitreihen zu schaffen. Als nächstes soll eine ausgewählte Lösung für die Zeitreihen-Klassifikation in die Data-Science-Plattform – Ontologie-basiertes Meta AutoML – integriert werden. Abschließend soll die Plattform für die Klassifikation von sequenziellen Fahrzeugdaten angewandt werden.

Basierend auf der Literaturrecherche wurde das Deep-Learning-basierte Mcfly-AutoML für die Zeitreihen-Klassifikation ausgewählt und in die bestehende Data-Science-Plattform integriert. Abschließend wurde die Plattform für die Klassifikation von sequenziellen Fahrzeugdaten angewandt. Bei dem Experiment auf einem Teildatensatz wurde ein F1-Score von 0,70 erzielt. Obwohl die 94% der Diagnose-Fehlercodes mit Hilfe von AutoML detektiert wurden (Recall von 0,94), war die Falsch-Positiv-Rate in dem Experiment hoch (Precision von 0,57).

Durch die Anwendung der Data-Science-Plattform kann der Aufwand bezüglich der Erstellung von Machine-Learning-Modellen für die Zeitreihen-Klassifikation sowie der Hyperparameteroptimierung verringert werden. Die in die Plattform integrierte AutoML-Lösung basiert auf Deep-Learning Netz-Architekturen. Daher ist eine längere Laufzeit für die Optimierung von Machine Learning Pipelines zu erwarten.

Schlagwörter: Automated Machine Learning, CAN-Bussystem, Deep Learning, Diagnose-Fehlercode, Ontologie-basiertes Meta AutoML, Zeitreihen-Klassifikation