UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

Bachelor/Master Project

Machine Learning Model Exchange Formats: Comparative Analysis and Evaluation

Term: Summer Term 2022

Language: English/German

Motivation

Machine learning (ML) marks a topic of recent interest for various strings of research and can be applied for a broad range of use cases within organizations. ML development is largely centered around the training and use of so-called "ML models" to make predictions on the basis of novel inputs. This includes several intermediate steps such as data wrangling, model selection, and model serving, which are mostly aggregated under the designator of an "ML pipeline". For this purpose, a multitude of tools and libraries exist that are tailored towards specific parts of the ML pipeline, e.g., NumPy, Sci-kit learn, Keras, TensorFlow, or DeepCognition. The developed ML model must thereby follow some internal representation. Most ML tools, like Google's TensorFlow utilize a proprietary representation not shared by other tools. Using various tools to edit and manage the same ML model is thereby challenging since this would require specifically tailored adapters between both tools. Proprietary representations of ML models thus lead to a lack of interoperability.

In recent years, several representation formats for ML models have been proposed to address this challenge. For instance, Microsoft and Facebook have released the *Open Neural Network Exchange Format* (ONNX; Boyd 2017, Gaskill 2018). Other solutions include the *Predictive Model Markup Language* (PMML; Grossman et al. 1999), the *Neural Network Exchange Format* (NNEF; Krohnos Group n.d.), or the *MMdnn* (Liu et al. 2020). Except for the PMML, the listed exchange formats focus exclusively on so-called "deep learning (DL) models", which mark a special case of ML models in general. These are not be confused with exchange formats used in computational neuroscience such as the NeuroML or NineML.

Description

This project aims at an identification, analysis, and assessment of existing exchange formats for ML models. According to preferences, it is also possible to focus on DL models exclusively. It is therefore required to elucidate the technological foundation of ML models and the respective exchange formats. The comparative assessment must follow an evaluation framework and yield use cases and limitations of the different formats. This might also include a prototypical implementation using some of the analyzed exchange formats based on the number of group members.

Institut für Informatik und Wirtschaftsinformatik (ICB)

Lehrstuhl für Wirtschaftsinformatik und Unternehmensmodellierung

Pierre Maier
Tel.: 0201 / 183 - 4563
pierre.maier@uni-due.de

R09 R04 H41
Universitätsstraße 9
45127 Essen

www.umo.wiwi.uni-due.de

The range of tasks can include the following:

- elucidation of foundational concepts in ML and/or DL
- identification and selection of ML model exchange formats
- elaboration of evaluation framework for comparative assessment
- prototypical implementation and exchange of ML models using the selected exchange formats

Expected Outcomes

The scope of the expected outcome varies according to the number of students. In any case, the students are expected to produce a project report that discusses the results as communicated in the description. If some prototypical implementation is realized, the respective source code should be made available.

Introductory Literature

- Boyd E (2017) Microsoft and Facebook create open Ecosystem for AI Model Interoperability. *Microsoft Blog*. https://azure.microsoft.com/en-us/blog/microsoft-and-facebook-create-open-ecosystem-for-ai-model-interoperability/
- Gaskill B (2018) ONNX: the Open Neural Network Exchange Format. *Linux Journal*. https://www.linuxjournal.com/content/onnx-open-neural-network-exchange-format
- Grossman R, Bailey S, Ramu A, Malhi B, Hallstrom P, Pelleyn I, Qin X (1999) The Management and Mining of Multiple Predictive Models Using the Predictive Modeling Markup Language. *Information and Software Technology* 41(9):589-595
- Guazelli A, Zeller M, Lin W-C, Williams G (2009). PMML: An Open Standard for Sharing Models. *The R Journal* 1(1):60-65
- Khronos Group (n.d.) Neural Network Exchange Format (NNEF). https://www.khronos.org/api/nnef
- Liu Y, Chen C, Zhang R, Qin T, Ji X, Lin H, Yang M (2020) Enhancing the Interoperability Between Deep Learning Frameworks by Model Conversion. *Proceedings of the 28th ACM Joint Meeting in European Software Engineering Conference and Symposium on the Foundations of Software Engineering*, pp. 1320-1330.

Application Procedure:

Please apply via email to the supervisor. Please attach a short letter of motivation (approximately 1/2 A4 page) and a recent transcript of records ('Leistungsnachweis'). You can apply individually or in a group of **3-5 participants** (in this case each person should still send a separate e-mail, however point to the other members of the group).

Application deadline: 20 April 2022, 23:59 h