AutoKeras

Tomasz Kurpiński, Jan Kruszewski, Damian Skowroński

May 2022

1 Introduction

The following article will discuss the results of our group's work during a project which was part of a subject 'research workshops'. The task set before us was to get to know the AutoML framework of our choice, create a function that prepares any data set describing a binary prediction problem, which then creates a model for it. At the end we were to test the created function on sets concerning binary classification from benchmark. In the project we will refer to the research just presented in the article of our benchmark in an attempt to create a kind of benchmark for the AutoKeras package omitted from the study.

2 Descripion of the framework

In this paragraph, we will briefly describe the AutoKeras framework.

2.1 Authors

AutoKeras is an AutoML framework based on the Keras package, developed by a group of people associated with Texas AM University. The main idea of the developers was to create a package that would give a chance to use machine learning to people even without much experience in this field. The package has an extensive and clear documentation.

2.2 What distinguishes AutoKeras?

A couple of benefits set AutoKeras apart from other AutoML frameworks. There is no need to use cloud services, which are not free and not affordable for anyone who wants to use machine learning. AutoKeras also ensures data security and privacy. It is also worth mentioning that AutoKeras focuses on deep learning problems unlike frameworks like SMAC, TPOT, Auto-WEKA or Auto-Sklearn which focus on shallow models.

2.3 API

The inteface of AutoKeras has been created in a similar way to Sklearna. It can be built in 3 lines of code using the constructor and the fit and predict methods. It is worth noting that it has two levels 'task-level' for users with less knowledge of the system and 'search-level' for advanced users who control preprocessing and neural network architecture themselves. In general, AutoKeras offers us the following models: ImageClassifier, ImageRegressor, TextClassifier, TextRegressor, StructuredDataClassifier, Structured DataRegressor and AutoModel. The framework has also many user-friendly features.

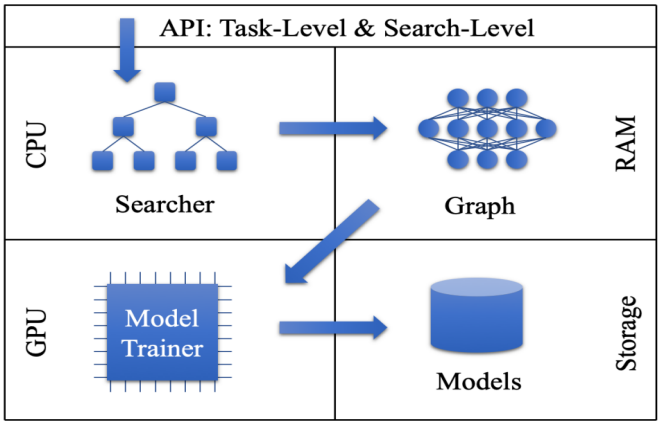
2.4 Use of resources

AutoKeras is designed to take full advantage of the CPU, GPU and RAM, where are stored only currently important information. The rest of the data is stored in other places such as the hard drive.

2.5 Division of AutoKearas into modules

The framework is divided into 4 main modules.

- The Searcher module, is responsible for searching the neural architecture. It uses Bayesian optimisation and a Gaussian process. The search algorithms use the CPU.
- Neural networks are trained in the Trainer module.
- Graph is a module that processes computational graphs that Searcher controls to morph the network.
- The last module is Storage. It is responsible for saving the final models. Due to their size they are not stored in RAM.



3 Preprocessing

Package does the necessary preprocessing for us before training. Hence, our function does not have any additional code snippets devoted to this activity. AutoKeras performs for us the basic vectorisation, data cleaning and normalisation, which is an important step for neural networks.

4 Results of the evaluation

We managed to check the performance of our function for 20 binary sets from the benchmark. The results for accuracy metric oscillated around the value 0.725. The best results were obtained for jasmine, mfeat-factors and shuttle. For each of these sets, the model achieved a score of around 0.97. It seems that we got by far the worst score for the sylvine set, but we are dealing with binary classification, so it would be enough to swap labels to get a very good result.