An Overview of Automated Machine learning (AutoML)

Author: Arvind Thorat

Abstract

Recent interest in complex, computationally intensive, hype- parameter-rich machine learning models, such as automated machine learning (AutoML) frameworks and deep neural networks, has led to a resurgence in hyper-parameter optimization (HPO) research. This chapter provides an overview of the most important approaches to HPO. First, we describe an optimization method for black box

Functions based on model-free methods and Bayesian optimization. The computationally expensive nature of many modern machine-learning applications makes pure black-box optimization very expensive, so we then use a (much) cheaper variant of the black-box function to optimize the hyper- parameters. We focus on state-of-the-art multi-fidelity techniques to roughly assess quality setting. Finally, we point out open questions and directions for future research.

Introduction

All machine-learning systems have hyper-parameters, and the most fundamental task of automated machine learning (AutoML) is to automatically set these hyper-parameters to optimize performance. In particular, new deep neural networks critically rely on a wide range of hyper-parameter options for neural network architecture, regularization, and optimization.

Automated machine learning (AutoML) has become a trending topic in industrial and academic research in the field of artificial intelligence (AI) in recent years. AutoML shows promise for delivering AI solutions in regulated industries, providing explainable and reproducible results. AutoML makes AI development more accessible to those who do not currently have the theoretical background required for a data science role.

Tasks in AutoML

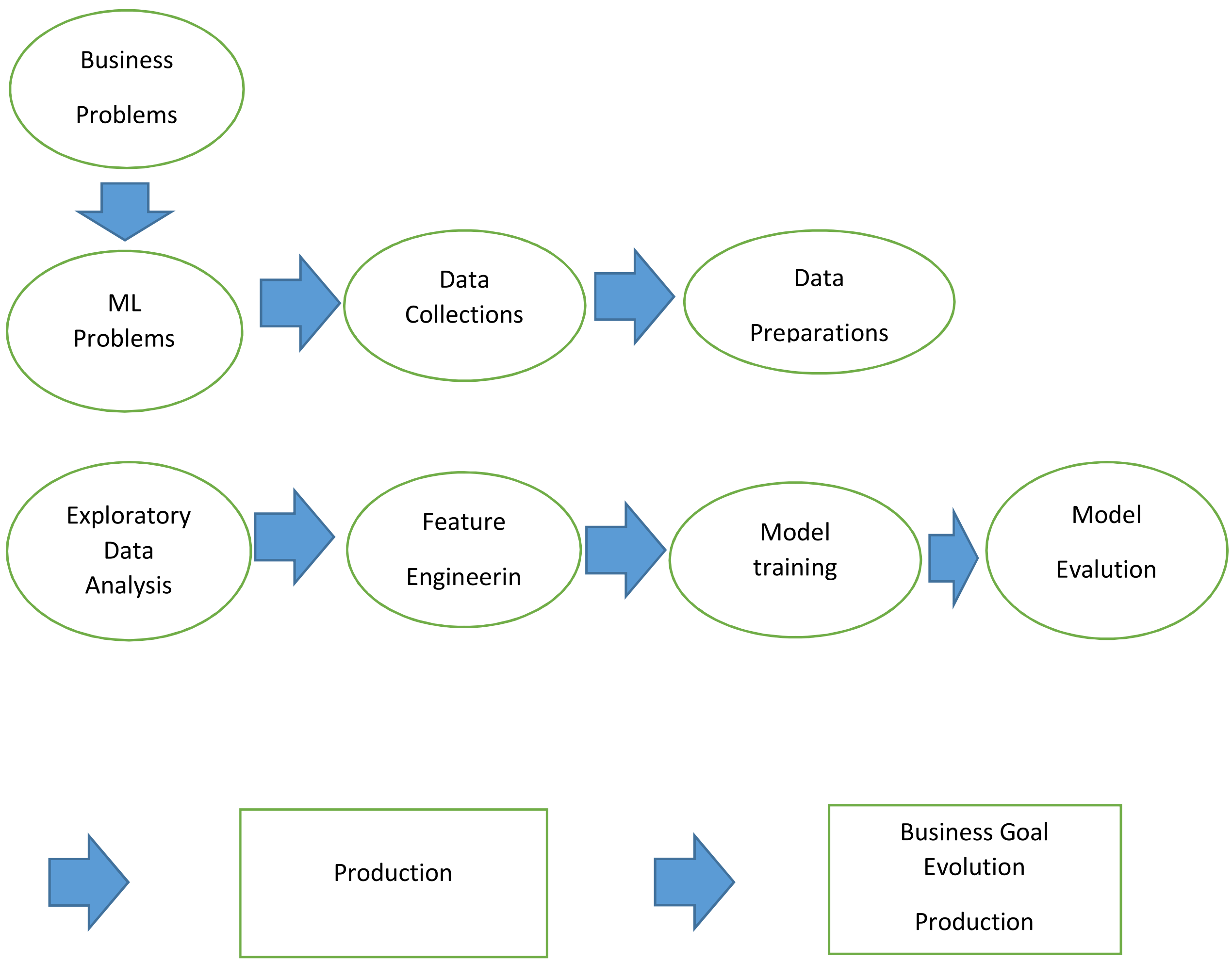
Feature Engineering in AutoML

Data features are part of the input data for machine learning models, and feature engineering refers to the transformation process by which data scientists derive new information from existing data. Feature engineering is one of the most important value-added processes in ML workflows, and good features can make the difference between models with acceptable performance and models with good performance. These mathematical transformations of the raw data are loaded into the model and serve as the heart of the machine learning process. Automated feature engineering (AFE) (link is external to IBM) is the process of exploring the realm of possible combinations of features in a mechanical rather than manual manner.

Tuning Hyper-parameters

Hyper-parameters are part of machine learning algorithms, and are similarly best understood as levers for fine-tuning model performance, but incremental tuning is often very It has a big impact. For small-scale data science modelling, hyper-parameters can be easily set manually and optimized through trial and error.

Following Diagram Represents the simple AutoML structure for Business Problems.



STEPS IN AUTOML

How it works

Data evaluation pre-processing

Data evaluation and pre-processing: Data is prepared , cleaned, and transformed to create datasets useful for training models.

Feature Engineering

Feature Engineering: New data columns are created in existing model training data to better represent predictors of the phenomenon described by the data, or simply to better interact with ML algorithms. There is a possibility.

Feature Selection

Feature Selection: After new features are created, AutoML selects only those that are useful for generating the model.

Algorithm Selection

Algorithm Selection: Competing candidate models are reviewed, and the model that performs best on the desired metric is selected (e.g., precision, recall, balanced precision optimization).

Hyper-parameter Tuning

Hyper-parameter Tuning: The optimal set of hyper-parameters for the learning algorithm is selected.

Different phases of the machine learning process can be the focus of automated machine learning.

The procedures to be automated are:

Data intake and preparation (from various formats and raw data) detection of column type, such as text, discrete numerical, continuous numerical, or Boolean, Target/label, stratification field, numerical feature, categorical text feature, or free text feature are a few examples of column intent detection. Task detection, such as regression, grouping, ranking, or binary classification

Features-based modelling.

1. Choose features

2. Extraction of features

3. Transfer learning and meta-learning

4. Identifying and managing skewed or missing value data

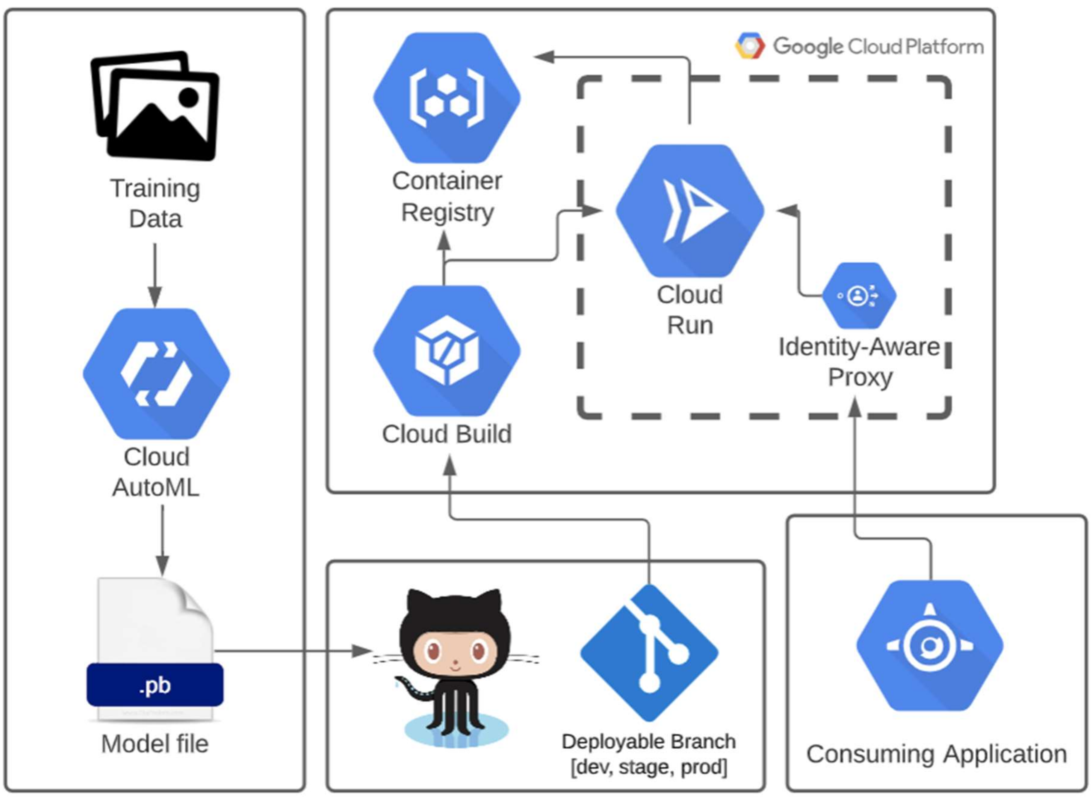
Selecting a model involves deciding the machine-learning algorithm to employ, which frequently involves weighing several rival software implementations.

Ensembling is a consensus-building process whereby employing many models frequently yields superior outcomes than utilizing only one.(5)

Optimization of hyper-parameters for the learning algorithm and feature selection

Pipeline selection in a timely manner

Below Image showing the AutoML in Google cloud Platform.



Importance

Because it marks a turning point in machine learning and artificial intelligence (AI), AutoML is significant. The "black box" argument has been levelled at AI and ML, implying that reverse engineering machine-learning algorithms can be challenging. Even if they increase productivity and processing ability to yield results, it can be challenging to trace the algorithm's exact method of delivery. As a result, selecting the best model for a given situation might be difficult since it can be hard to forecast a model's outcome if it is a "black box."

By making machine learning more approachable, AutoML contributes to its reduction as a mystery. This procedure automates the components of machine learning that use the algorithm in practical situations. A person executing

Applications

AutoML and conventional machine learning have similar use cases. Among these are a few of the following:

Fraud detection in finance, where it enhances the precision and accuracy of models used to detect fraud. Research and development in the medical field, where it is able to make inferences and analyse big data sets.

Facial recognition can benefit from image recognition. Risk evaluation and management in insurance, banking, and finance. Cybersecurity, where it used to testing, monitoring, and risk assessment. In customer service, it used to boost team productivity and do sentiment analysis in chatbots. Spam and malware have the potential to produce adaptive cyber-threats.

In agriculture, it used to speed up the process of quality testing.

USE across several industries

Healthcare:

By helping with disease diagnosis and treatment, AutoML is transforming the healthcare industry. For example, machine learning models that can precisely identify diabetic retinopathy, a major cause of blindness, have been developed using Google's AutoML. By using this technology, medical professionals may better identify patients who are at danger and intervene in a timely manner, leading to better treatment outcomes.

Financial services:

By improving fraud detection and risk assessment, AutoML is revolutionizing the financial services sector. PayPal, for instance, uses AutoML to create fraud detection models that look for fraudulent activity by analysing user behaviour, transaction patterns, and other factors. Users are shield from fraudulent transactions and financial losses by this technology.

Manufacturing:

AutoML is used to optimize manufacturing techniques and improve product fantastic. One amazing instance is Bosch, a multinational engineering and electronics enterprise. Bosch implemented AutoML to optimize its production operations, enhancing efficiency and reducing fees. AutoML fashions examine sensor data, gadget parameters, and historical performance to predict screw-ups and optimize renovation schedules, minimizing downtime and maximizing productiveness.

Retail and ecommerce:

AutoML is supporting shops and ecommerce groups optimize pricing strategies, call for forecasting, and customized guidelines. As an instance, sew restore, a web private styling provider, advantages AutoML to provide personalized style hints to its clients. Via studying purchaser options, previous purchases, and fashion tendencies, AutoML generates tailor-made pointers, enhancing the patron enjoy and using income.

Transportation and logistics:

AutoML is enhancing logistics and supply chain control via call for forecasting and route optimization. For example, UPS, a worldwide package deal delivery employer, makes use of AutoML to optimize delivery routes. AutoML fashions consider variables including traffic patterns, bundle volume, and shipping constraints to generate most effective routes, thereby lowering gas consumption and improving shipping efficiency.

Sample code

```
import sklearn.model_selection

import sklearn.datasets

import sklearn.metrics

import autosklearn.regression


def main():

X, y = sklearn.datasets.load_boston(return_X_y=True)

feature_types = (['numerical'] * 3) + ['categorical'] + (['numerical'] * 9)

X_train, X_test, y_train, y_test = \

sklearn.model_selection.train_test_split(X, y, random_state=1)


automl = autosklearn.regression.AutoSklearnRegressor(

time_left_for_this_task=120,
```

```
per_run_time_limit=30,
tmp_folder='/tmp/autosklearn_regression_example_tmp',
output_folder='/tmp/autosklearn_regression_example_out',
)
automl.fit(X_train, y_train, dataset_name='boston',
feat_type=feature_types)

print(automl.show_models())
predictions = automl.predict(X_test)
print("R2 score:", sklearn.metrics.r2_score(y_test, predictions))
```

References

AutoML.Org

Hutter F, Caruana R, Bardenet R, Bilenko M, Guyon I, Kegl B, and Larochelle H. "AutoML 2014 @ ICML". AutoML 2014 Workshop @ ICML. Retrieved 2018-03-28.[permanent dead link]

De Bie, Tijl; De Raedt, Luc; Hernández-Orallo, José; Hoos, Holger H.; Smyth, Padhraic; Williams, Christopher K. I. (March 2022). "Automating Data Science". Communications of the ACM. 65 (3): 76–87. doi:10.1145/3495256.

Predicting Carpark Prices Indices in Hong Kong Using AutoML, Computer Modeling in Engineering & Sciences 2023, 134(3), 2247-2282. https://doi.org/10.32604/cmes.2022.020930

Erickson, Nick; Mueller, Jonas; Shirkov, Alexander; Zhang, Hang; Larroy, Pedro; Li, Mu; Smola, Alexander (2020-03-13). "AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data". arXiv:2003.06505 [stat.ML].