

머신러닝 파이프라인

쿠베플로우 파이프라인 Part 3

송호연





목차



쿠베플로우 파이프라인 Part 3



1-1. 실습 1 - Tensorflow MNIST, CatBoost





학습목표



쿠베플로우 파이프라인 Part 3



01. 쿠베플로우 실습을 통해 작동 방식을 이해한다.

실습을 진행하면서 쿠베플로우 파이프라인의 기본 사용법에 대해 공부한다.



실습

Tensorflow MNIST

CatBoost

01



실습 1


Tensorflow MNIST

https://github.com/chris-chris/kubeflow-tutorial/tree/master/lesson10_tf_mnist

Experiments > Train TF MNIST

← ✓ tf_mnist_pipeline 2021-04-27 06-46-33 Retry Clone run Terminate Archive

Graph Run output Config



tf-mnist-pipeline-bd69v-52004730

	Artifacts	Input/Output	Volumes	Manifest	Logs
1	2021-04-26 21:50:21.880918:	W	tensorflow/stream_executor/platform/default/dso_loader.cc:60]	Could	
2	2021-04-26 21:50:21.880980:	I	tensorflow/stream_executor/cuda/cudart_stub.cc:29]	Ignore above cuda	
3			(60000, 28, 28)		
4			(60000,)		
5	2021-04-26 21:50:24.223335:	I	tensorflow/compiler/jit/xla_cpu_device.cc:41]	Not creating XLA device	
6	2021-04-26 21:50:24.223640:	W	tensorflow/stream_executor/platform/default/dso_loader.cc:60]	Could	
7	2021-04-26 21:50:24.223672:	W	tensorflow/stream_executor/cuda/cuda_driver.cc:326]	failed call to cu	
8	2021-04-26 21:50:24.223702:	I	tensorflow/stream_executor/cuda/cuda_diagnostics.cc:156]	kernel driver	
9	2021-04-26 21:50:24.223971:	I	tensorflow/core/platform/cpu_feature_guard.cc:142]	This TensorFlow b	
10			To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.		
11	2021-04-26 21:50:24.225210:	I	tensorflow/compiler/jit/xla_gpu_device.cc:99]	Not creating XLA device	
12	2021-04-26 21:50:24.290580:	W	tensorflow/core/framework/cpu_allocator_impl.cc:80]	Allocation of 47	
13	2021-04-26 21:50:24.366670:	I	tensorflow/compiler/mlir/mlir_graph_optimization_pass.cc:116]	None of	
14	2021-04-26 21:50:24.367467:	I	tensorflow/core/platform/profile_utils/cpu_utils.cc:112]	CPU Frequen	
15	1/1875 [...]			ETA: 13:45 - loss: 191.8114 - sparse_categorical_accu	
16	1/313 [...]			ETA: 1:15 - loss: 0.2129 - sparse_categorical_accuracy:	
17	2021-04-26 21:50:33.190414:	W	tensorflow/python/util/util.cc:348]	Sets are not currently considere	
18					



실습 1



Tensorflow MNIST

```
import kfp
from kfp.components import func_to_container_op, OutputPath, InputPath

EXPERIMENT_NAME = 'Train TF MNIST'      # Name of the experiment in the UI
KUBEFLOW_HOST = "http://127.0.0.1:8080/pipeline"

def download_mnist(output_dir_path: OutputPath()):
    import tensorflow as tf

    tf.keras.datasets.mnist.load_data(output_dir_path)
```



실습 1

Tensorflow MNIST

```
def train_mnist(data_path: InputPath(), model_output: OutputPath()):  
    import tensorflow as tf  
    import numpy as np  
    with np.load(data_path, allow_pickle=True) as f:  
        x_train, y_train = f['x_train'], f['y_train']  
        x_test, y_test = f['x_test'], f['y_test']  
  
    model = tf.keras.models.Sequential([  
        tf.keras.layers.Flatten(input_shape=(28, 28)),  
        tf.keras.layers.Dense(128, activation='relu'),  
        tf.keras.layers.Dense(10)  
    ])  
    model.compile(  
        optimizer=tf.keras.optimizers.Adam(0.001),  
        loss=tf.keras.losses.SparseCategoricalCrossentropy(from_logits=True),  
        metrics=[tf.keras.metrics.SparseCategoricalAccuracy()],  
    )  
    model.fit(x_train, y_train)  
    model.evaluate(x_test, y_test)  
    model.save(model_output)
```



실습 1



Tensorflow MNIST

```
def tf_mnist_pipeline():
    download_op = func_to_container_op(download_mnist, base_image="tensorflow/tensorflow")
    train_mnist_op = func_to_container_op(train_mnist, base_image="tensorflow/tensorflow")
    train_mnist_op(download_op().output)

if __name__ == '__main__':
    import kfp.compiler as compiler
    compiler.Compiler().compile(tf_mnist_pipeline, __file__ + '.zip')
    kfp.Client(host=KUBEFLOW_HOST).create_run_from_pipeline_func(
        tf_mnist_pipeline,
        arguments={},
        experiment_name=EXPERIMENT_NAME)
```




실습 1

CatBoost

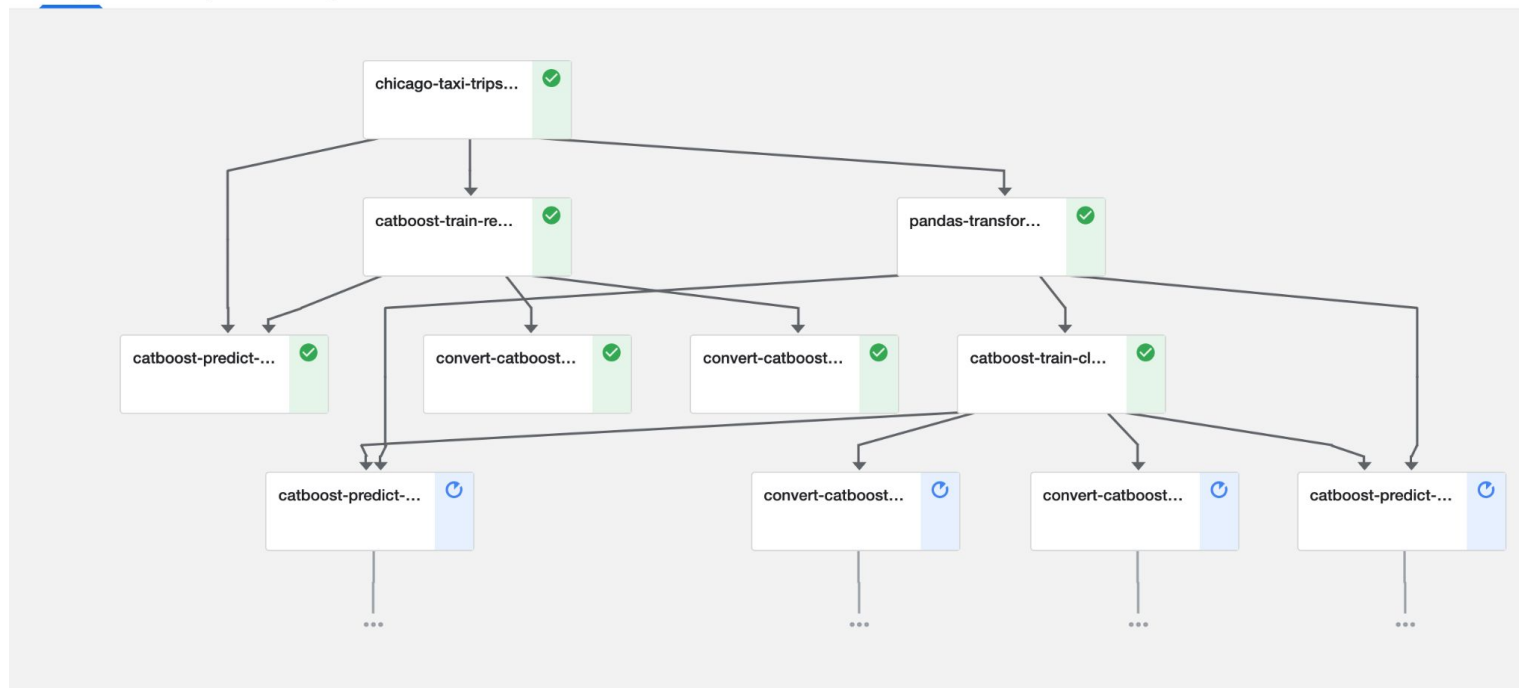
https://github.com/chris-chris/kubeflow-tutorial/tree/master/lesson11_catboost

Experiments > CatBoost pipeline

Retry Clone run Terminate Archive

← catboost_pipeline 2021-04-27 06:47:16

Graph Run output Config





실습 1

CatBoost

```
import kfp
from kfp import components

EXPERIMENT_NAME = 'CatBoost pipeline'    # Name of the experiment in the UI
KUBEFLOW_HOST = "http://127.0.0.1:8080/pipeline"

chicago_taxi_dataset_op =
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/e3337b8bdcd63636934954e592d4b3
2c95b49129/components/datasets/Chicago%20Taxi/component.yaml')
pandas_transform_csv_op =
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/e69a6694/components/pandas/Trans
form_DataFrame/in_CSV_format/component.yaml')
```



실습 1



CatBoost

```
catboost_train_classifier_op =  
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/f97ad2/components/CatBoost/Train_classifier/from_CSV/component.yaml')  
catboost_train_regression_op =  
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/f97ad2/components/CatBoost/Train_regression/from_CSV/component.yaml')  
catboost_predict_classes_op =  
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/f97ad2/components/CatBoost/Predict_classes/from_CSV/component.yaml')  
catboost_predict_values_op =  
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/f97ad2/components/CatBoost/Predict_values/from_CSV/component.yaml')  
catboost_predict_class_probabilities_op =  
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/f97ad2/components/CatBoost/Predict_class_probabilities/from_CSV/component.yaml')  
catboost_to_apple_op =  
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/f97ad2/components/CatBoost/convert_CatBoostModel_to_AppleCoreMLModel/component.yaml')  
catboost_to_onnx_op =  
components.load_component_from_url('https://raw.githubusercontent.com/kubeflow/pipelines/f97ad2/components/CatBoost/convert_CatBoostModel_to_ONNX/component.yaml')
```



실습 1

CatBoost

```
def catboost_pipeline():
    training_data_in_csv = chicago_taxi_dataset_op(
        where='trip_start_timestamp >= "2019-01-01" AND trip_start_timestamp < "2019-02-01"',
        select='tips,trip_seconds,trip_miles,pickup_community_area,dropoff_community_area,fare,tolls,extras,trip_total',
        limit=10000,
    ).output

    training_data_for_classification_in_csv = pandas_transform_csv_op(
        table=training_data_in_csv,
        transform_code='df.insert(0, "was_tipped", df["tips"] > 0); del df["tips"]',
    ).output

    catboost_train_regression_task = catboost_train_regression_op(
        training_data=training_data_in_csv,
        loss_function='RMSE',
        label_column=0,
        num_iterations=200,
    )
```



실습 1



CatBoost

```
regression_model = catboost_train_regression_task.outputs['model']

catboost_train_classifier_task = catboost_train_classifier_op(
    training_data=training_data_for_classification_in_csv,
    label_column=0,
    num_iterations=200,
)

classification_model = catboost_train_classifier_task.outputs['model']

evaluation_data_for_regression_in_csv = training_data_in_csv
evaluation_data_for_classification_in_csv = training_data_for_classification_in_csv

catboost_predict_values_op(
    data=evaluation_data_for_regression_in_csv,
    model=regression_model,
    label_column=0,
)
```



실습 1



CatBoost

```
catboost_predict_classes_op(  
    data=evaluation_data_for_classification_in_csv,  
    model=classification_model,  
    label_column=0,  
)  
  
catboost_predict_class_probabilities_op(  
    data=evaluation_data_for_classification_in_csv,  
    model=classification_model,  
    label_column=0,  
)  
  
catboost_to_apple_op(regression_model)  
catboost_to_apple_op(classification_model)  
  
catboost_to_onnx_op(regression_model)  
catboost_to_onnx_op(classification_model)
```



실습 1



CatBoost

```
if __name__ == '__main__':  
    kfp.compiler.Compiler().compile(catboost_pipeline, __file__ + '.zip')  
    kfp.Client(host=KUBEFLOW_HOST).create_run_from_pipeline_func(  
        catboost_pipeline,  
        arguments={},  
        experiment_name=EXPERIMENT_NAME)
```

❶ 짚어보기

❶ 쿠베플로우 파이프라인 Part 3

01. 쿠베플로우 실습을 통해 작동 방식을 이해한다.

실습을 진행하면서 쿠베플로우 파이프라인의 기본 사용법에 대해 공부한다.

머신러닝 파이프라인

쿠베플로우 파이프라인 Part 3

송호연



감사합니다.

THANKS FOR WATCHING

