Deep Learning on Kubeflow

Big Data Praktikum

Iurii Mozzhorin, Rebekka Pech

15.07.2019

What is Kubeflow?

- an open-source platform
- based on Kubernetes
- simple, portable and scalable deployments of ML
- integration of existing tools and libraries

Use Case

- Train a CNN model to distinguish between works of 10 painters
- Dataset from kaggle
- Size of Dataset: (after Resizing:)
- Objective: A model that gives a (good) predicition who painted the incoming image

Milestones

- Create a model (transfer learning)
- Create a Docker image
- Train the model on GKE using TFJobs on CPUs
- Train the model on GKE using TFJobs on GPUs
- Distributed training on multiple pods
- Save the trained model into a GCS bucket
- Serve the model with TFServing

Technologies

- TF for the model $\rightarrow$ keras
- Docker + Kubernetes
- Kubectl + ksonnet $\rightarrow$ kubectl + kustomize $\rightarrow$ kubectl
- GCP dashboard/console & gcutil
- TF Serving: Tensorboard

Create a model

Create a Docker image

Train the model (CPU)

Train the model (GPU)

Problems

- Keras: cannot save models/logs directly to a bucket
- Keras: no official support for distributed learning
- Kustomize/ksonnet

Conclusion

- Requirement of understanding the used technologies (Dockers, Kubernetes, Keras...)
- still a raw version (ksonnet,...)
- Kubeflow is not for data science - it's for model training