Requirements Document

Big Data Boys - Tyler Cole, Arnav Bhutani

Machine learning and data science has been one of the fastest growing fields in computer science and computing for years. The aim of this project is to create a platform in which researchers can develop algorithms without having a lot of machine learning experience. By using advanced algorithms, our project uses the power of the cloud to parallelize the generation and training of a machine learning model that is tailored specifically to an individual application.

Abstract

Index Terms

Cloud computing, Machine Learning, Genetic Algorithms, Big Data.

CONTENTS

1 INTRODUCTION

1.1 System Purpose

The purpose of this is to create a multifaceted platform for running distributed machine learning tasks. This system will be optimized to have very low latency for a variety of algorithms such as you only look once (YOLO) and RESNET-18. The system will also be able to develop algorithms based on input sources. The platform distributes the training and running of these algorithms into various containers using Kubernetes. Due to current circumstances, the algorithm will be restricted to only working with cloud service providers, and can later be modified to work with multiple systems.

1.2 Scope

This document will cover the requirements that the system must meet. It will not describe the implementation of the system in detail, and will not discuss any technology involved in the product. The requirements outlined by this document must be met in order for the project to be deemed completed. Extra time may be used to complete the requirements laid out in the Stretch Goals section, however these goals are optional and will only be completed once the base requirements of the system have been accomplished.

1.3 System Overview

The system will be built on a cloud server. It will have two primary functionalities. The first involves a pipeline ranging from data ingestion with an input file to the output classification. The system will take input data, pre-process and queue it, analyze the processed data, and then output the analysis. The second system is an algorithm development platform, where the platform will develop the 'most' suitable algorithm possible for a given dataset. This section will be composed of docker containers running in a 'master-slave' configuration, with the trainer being the master, while the various worker pods which parallelize the training will be the 'slaves'. The output algorithm of the development process can then be used for analysis by the classification process. The entire program will be handled by a single orchestrator.

2 SYSTEM REQUIREMENTS

2.1 Functional Requirements

1) The system must ingest a form of data.
2) The system must split the input into individual frames.
3) The system must recognize a singular individual using facial recognition technology.
4) The system must be managed by an orchestrator.
5) The system must be built using containerization software.

2.2 Usability/Performance Requirements

1) The system must ingest data at 10 frames per a second.
2) The system must output data at 1 gigabyte per a second.

3) The system must be optimized for performance per a watt per dollar.
4) The system must be fault tolerant.

2.3 Stretch Goals

1) The system must integrate spark to process 'big data' components at the cost of latency.
2) The system must ingest data multiple camera inputs, given each input has a connected edge server.
3) The system must have a maximum latency of 10ms when deployed on the edge.
4) The system must have a visualization that tracks the person of interest across a digital approximation of the sample space(s).
5) The system must interface with a fifth generation wireless network.
6) Write and publish an academic paper on the edge system.

2.4 System Interface

The system should be extensible to potentially work with any number of file inputs. It should be simple to install, and should be able to work with on the cloud. It should be ready to interface with the edge with minimal changes.

2.5 Elements modified due to COVID-19

The project has been modified such that the project can now be completed on the cloud. Due to this the latency requirements have been removed, as cloud applications will be unreliable for timing. Additionally, setting up camera feeds that stream directly to the cloud are infeasible, so the requirements have been changed to reflect that the input source will be a stream of input images rather than a live feed. In addition to these requirement change, an additional system has been added to the requirements. This system which develops algorithms is described in the system overview.

3 GANTT CHART