BIG DATA PROJECT

Kaiyue Wu

August 17,2019

Contents

ABSTRACT

With the popularity of Docker application container engine, I was deeply attracted by its features of easy installation and deployment, excellent version control, portability and isolation. In this project, I built a CNN cloud platform based on Docker container, which makes it very convenient for users to install, use it to predict image recognition results, and store the obtained data into Cassandra database, so that users can use the saved data to make further analysis.

In order to achieve the above steps, firstly, develop the training model of convolutional neural network (CNN) based on MXNet store the training results and import them into the Flask based restful API server, then implement the related functions of Web front-end based on Vue, and finally integrate the related functions of front-end and server into the Docker container.

1 INTRODUCTION

1.1 Python

Python is an interpreted, high-level, general-purpose programming language. Its design emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library. This project was built on Python 3.7.

1.2 Machine Learning

Machine learning is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead. Machine learning algorithms build a mathematical model based on sample data, known as "training data", in order to make predictions or decisions without being explicitly programmed to perform the task.

1.3 MXNet

Apache MXNet is an open-source deep learning software framework,

used to train, and deploy deep neural networks. It is scalable, allowing for fast model training, and supports a flexible programming model and multiple programming languages (including C++, Python, Julia, Matlab, JavaScript, Go, R, Scala, Perl, and Wolfram Language.)

2 IMPLEMENTATION

See in file Descripiom.md

3 CONCLUSION

Big data is a brand new field to me, before this project, I was totally innocent to relevant knowledge in big data, however, through the course and this project, not only have I got a clear sketch about this field, but also I had a good command of basic skills in big data. First of all, the containerization technology, especially the use of Docker, really astonishes me. The first thing is its magic structure which does not necessarily need a guest OS; The second thing is the application, I’ve never thought there exists such a convenient and easy way to run a service without even having one on the machine. I’ve also learned the principal of NoSQL database and mastered the basic use of Cassandra database. What I think the coolest part of this research is to integrate

deep learning with the Docker and Cassandra, I got a sense of accomplishment to work out a project using many tools and knowledge of different fields.