Project Report

Big Data Project

Name : Liuyang Zheng

Instructor : Fan Zhang

Date : Sep 15, 2019

Introduction

1. Abstract Objective

This project is aimed at building an application to identify handwritten digits. The final application is expected to identify the handwritten digits on pictures by using the trained neural network model and store the results into Cassandra database.

2. Basic Ideas

- Core Functionalities
  - a. Users should be able to use this application to identify the handwritten digits on their chosen picture.
  - b. Users should be able to check Cassandra database to get the results of predictions.
- Main Components
  - a. Neural network model based on tensorflow
  - b. Cassandra database

Background and Objectives

1. Background

- **Background Knowledge**

a. TensorFlow : TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries, and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

b. The Apache Cassandra database is the right choice when you need scalability and high availability without compromising performance. Linear scalability and proven fault-tolerance on commodity hardware or cloud infrastructure make it the perfect platform for mission-critical data. Cassandra's support for replicating across multiple datacenters is best-in-class, providing lower latency for your users and the peace of mind of knowing that you can survive regional outages.

- Background of Design

First, I want to identify the handwritten digits on the pictures. I create a simple neural network and train it 2000 times with a mnist dataset b. Then I want to add the accuracy of prediction, then I change the

network into a 2-layer convolutional network and train it for 20000 times with the same sample dataset. After that I get a more accurate model.

Finally, I restore the parameters and configuration from the saved model and the application is able to predict a picture with handwritten digits.

2. Objectives

● Functional Objectives

This application should be able to predict handwritten digits with an output result, and meanwhile store the result into Cassandra database.

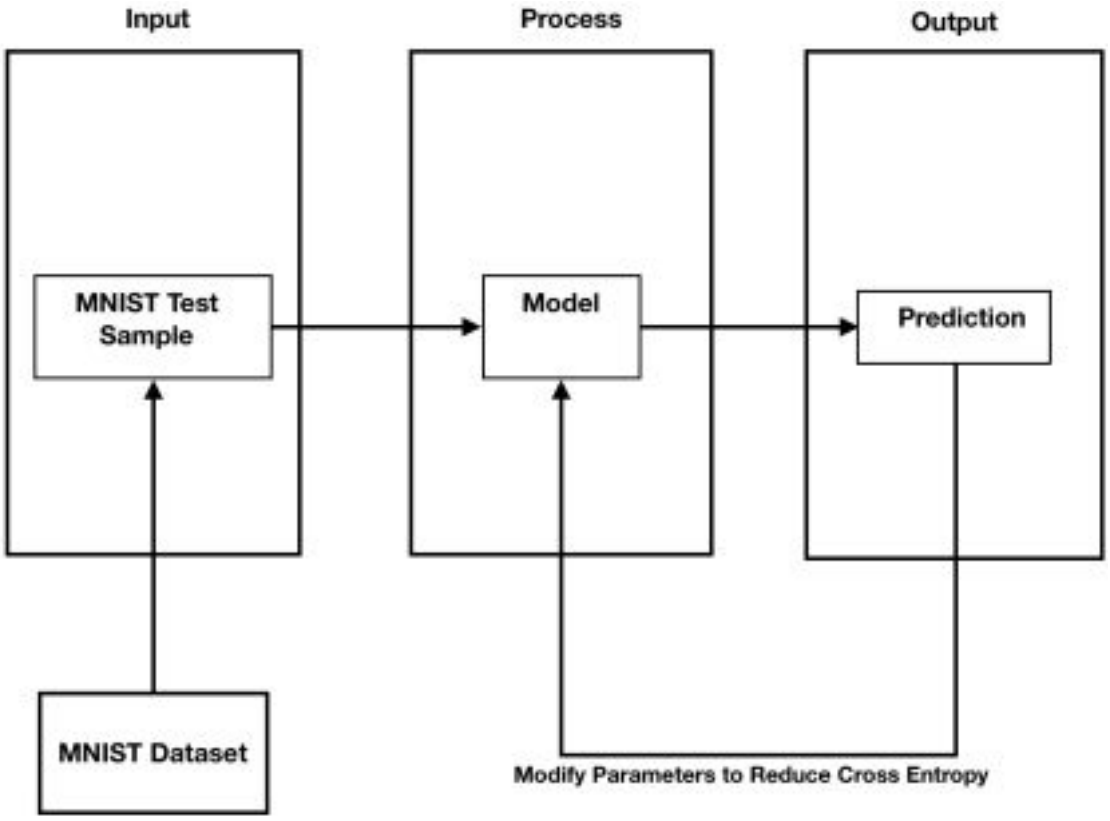
● Non-Functional Objectives

a. The product shall be delivered on schedule

b. The final product shall be concise, clear and user friendly

Approach/Methodology/Algorithm

1. Application Architecture



2. Component Approach

- Develop Environment : Jupiter Notebook
- Develop Language : Python
- Database : Cassandra

3. Algorithm

- Convolutional Neural Network

Results, Finding and Analysis

1. Results

- The expected prediction accuracy is at least above 90%
- The actual prediction accuracy that I test with my own pictures are pretty low from 90%

2. Problem Found

- The pre-preparation of the input mnist picture is too simple and not appropriate.
- The layers of the convolutional neural network are not enough. ● The pictures are not pixel pictures

3. Analysis

- For problem 2, more hidden layers could be added

Conclusion and Reflection

1.Accomplishment

This application is accomplished for about 60%, it is able to use the trained model to predict the handwritten digits and write results back to the cassandra database.

2. Defects

The prediction accuracy of this application is not satisfying and needs further enhancement.