Machine Learning for Investing Behaviours on Cryptocurrencies

Chengkai Lu

Submitted for the Degree of Master of Science in

Data Science and Analytics with a Year in Industry

Department of Computer Science
Royal Holloway University of London
Egham, Surrey TW20 0EX, UK

April 15, 2018

Declaration

This report has been prepared on the basis of my own work. Where other published and unpublished source materials have been used, these have been acknowledged.

Word Count:

Student Name:

Date of Submission:

Signature:

Abstract

Your abstract goes here.

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Aims And Objectives . . . 1
- 1.3 Project Structure . . . 1
  - 1.3.1 Technologies . . . 1
  - 1.3.2 Programme Structure . . . 1
  - 1.3.3 Report Structure . . . 1

2 Clustering **1**
- 2.1 Background . . . 1
- 2.2 K-means Clustering . . . 1

3 Recurrent Neural Network **2**
- 3.1 Background . . . 2
- 3.2 Long Short-Term Memory . . . 2

4 K-means Clustering on Price Fluctuations **2**
- 4.1 Data Prepocessing . . . 2
  - 4.1.1 Data Formats . . . 2
  - 4.1.2 Normalisation . . . 2
- 4.2 Data Analysis . . . 2
  - 4.2.1 Shifting Along Timeline . . . 2
  - 4.2.2 Model Training . . . 2
- 4.3 Result . . . 2

5 LSTM on Highest/Lowest Nextday Growth Prediction **3**
- 5.1 Data Prepocessing . . . 3
  - 5.1.1 Data Formats . . . 3
  - 5.1.2 Normalisation . . . 3
  - 5.1.3 Training Set and Validation Set . . . 3
- 5.2 Data Analysis . . . 3
  - 5.2.1 Activation Functions . . . 3
  - 5.2.2 Parameters . . . 3
  - 5.2.3 Regularisation . . . 3
  - 5.2.4 Loss Function . . . 3
  - 5.2.5 Model Training . . . 3
- 5.3 Result . . . 3
  - 5.3.1 Back Test . . . 3

6 Performance **3**
6.1 K-means Clustering 3
6.2 Long Short-Term Memory 3

7 Conclusion and Evaluation **3**
7.1 Visualisation 3
7.2 Further Work 3

References **3**

1 Introduction

1.1 Motivation

1.2 Aims And Objectives

1.3 Project Structure

1.3.1 Technologies

1.3.2 Programme Structure

1.3.3 Report Structure

2 Clustering

2.1 Background

Clustering, cluster analysis or data segmentation is a non-parametric algorithm in the subtree of unsupervised learning. It is used to seperate data into different groups using their natural dissimilarities. Unlike supervised learning, this type of learning algorithms does not have any indicator for assessing the quality of results, and this means that it does not have any meaning or objective itself. Instead, it discovers the distribution of data and uses the definition given by people who have the specific domain knowledge. By giving the rules for partitioning data self-defined meanings, useful information can be obtained and utilised in different domains[2].

2.2 K-means Clustering

An example of a reference: [1].

3 Recurrent Neural Network

3.1 Background

3.2 Long Short-Term Memory

4 K-means Clustering on Price Fluctuations

4.1 Data Prepocessing

4.1.1 Data Formats

4.1.2 Normalisation

4.2 Data Analysis

4.2.1 Shifting Along Timeline

4.2.2 Model Training

4.3 Result

5 LSTM on Highest/Lowest Nextday Growth Prediction

5.1 Data Prepocessing

5.1.1 Data Formats

5.1.2 Normalisation

5.1.3 Training Set and Validation Set

5.2 Data Analysis

5.2.1 Activation Functions

5.2.2 Parameters

5.2.3 Regularisation

5.2.4 Loss Function

5.2.5 Model Training

5.3 Result

5.3.1 Back Test

6 Performance

6.1 K-means Clustering

6.2 Long Short-Term Memory

7 Conclusion and Evaluation

7.1 Visualisation

7.2 Further Work

References

[1] Adam Gibson and Josh Patterson. *Deep Learning - A Practitioner's Approach*. O'Reilly Media, first edition, 2017.

[2] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, second edition, 2009.