Financial Time Series Analysis

by

Yin Jiangling

Master of Science in Software Engineering

2011

Faculty of Science and Technology

University of Macau

University of Macau

Abstract

FINANCIAL TIME SERIES ANALYSIS

by Yin Jiangling

Thesis Supervisor: Dr. SI Yain Whar

Master of Science in Software Engineering

Financial time series analysis and prediction has been one of active research problems in financial data mining area due to the commercial applications and the attractive benefits that it offers. A number of algorithms have been introduced into this area for analysis and prediction on the time series, such as classification, clustering, segmentation, indexing and discovering rules from historical time series. However, financial time series data is characterized as large in data size, dynamic and non-linear. Therefore it poses a great challenge for analysts to efficiently discover required information. As for data mining tasks, most of the execution time spent by algorithm is to move data from disk into main memory. This is acknowledged as the major bottleneck in Data Mining because many algorithms require multiple accesses of the data. Moreover, when analyzing or predicting the movement of time series, financial analysts usually capture the trends, shapes or patterns contained within the time series instead of the exact value of each time series data point. Thus time series representation on dimensionality reduction becomes the necessary step before mining information in this area.

In this thesis we study the property of the maximum or minimum points on the time series, which have the property of maintaining the trends and fluctuations of the time series. These points are called turning points in our thesis. However, not all turning points from the original time series equally contribute to the overall movement of the time series. Thus we introduce three independent methods with different perspectives to select turning points for approximating time series with

dimensionality reduction. The first method is to detect turning points at different levels of details based on the heuristics. This approach is extremely useful for analyzing stock data in top-down fashion. The second method is to select turning points based on a sliding window. The third method is to select turning points based on the degree of importance. The importance of a turning point is calculated according to its contribution on preserving the time series trends and shape. These turning points with different importance are stored in an optimal binary search tree. Such storage scheme allows more important points to be retrieved first while the average search cost is kept at minimum.

Keywords: Financial time series, Dimensionality reduction, Turning points, Trends, Sliding window, Optimal binary search tree

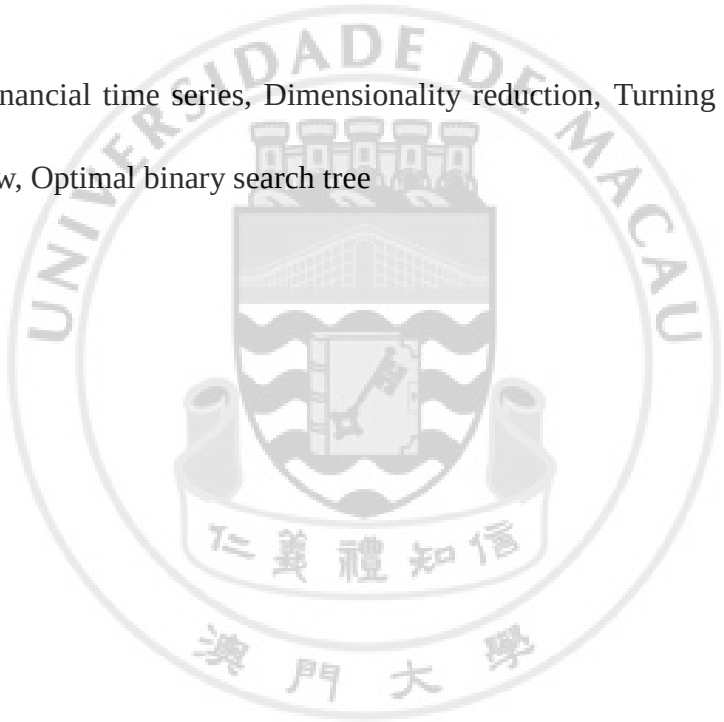


TABLE OF CONTENTS

TABLE OF CONTENTS ..... **i**

LIST OF TABLES ..... **v**

ACKNOWLEDGMENTS ..... **vi**

Chapter 1 Introduction **1**

1.1 Financial Time Series Analysis on Dimensionality Reduction Representation 1
1.2 Turning Points (TPs) 2
1.3 Issues for Time Series Representation 4
1.4 Problem Statement and Contribution 5

Chapter 2 Related Works **8**

2.1 Discrete Fourier Transform (DFT) Method 8
2.2 Piecewise Linear Approximation (PLA) Method 9
2.3 Piecewise Aggregate Approximation (PAA) and Adaptive Piecewise Constant Approximation (APCA) Methods 11
2.4 Perceptually Important Points (PIP) Method 12

Chapter 3 Heuristic Based Turning Points (TPs) Selection **14**

3.1 Turning Points (TPs) Identification at Different Levels 14
3.2 Experimental Results 18

Chapter 4 Sliding Window Method for Turning Points (TPs) Selection **21**

4.1 Turning Points (TPs) Selection 21
4.2 Properties of Sliding Window Method 26
4.3 Experimental Results 32
4.3.1 Reconstruction Error Comparison *32*
4.3.2 Properties Comparison *35*

Chapter 5 Turning Points (TPs) Storing and Retrieval Based on Degree of Importance **39**

5.1 Identification and Evaluation 39
5.1.1 Importance Evaluation *42*
5.2 Storing and Retrieving 45
5.2.1 Optimal Binary Search Tree (Optimal-BST) *45*
5.2.2 Storing Turning Points (TPs) into Optimal-BST *47*
5.2.2 Retrieving Time Series with Dimensionality Reduction from Optimal-BST *50*
5.3 Experimental Results 52

Chapter 6 Comparison of Turing Points (TPs) Selection through Heuristics, Sliding Window and Degree of Importance **58**

6.1 Reconstruction Error Comparison 58
6.2 Properties Comparison 61

Chapter 6 Conclusion and Future Works **65**

References **67**

VITA **70**

UNIVERSIDADE DE MACAU
仁義禮知信
澳門大學