Department of Computer Science
University of Hong Kong
Final Year Project 2015-2016

Detailed Project Plan

Title: Financial Data Forecaster

Supervisor: Dr Yip, Chi Lap
Team: Lo Chun Wing (3035074969)
Matsumoto Keisei (2013718993)
Date: 4th October, 2015

1 Table of Contents

2 Abstract

Financial Data Forecaster is a program or a software which helps in predicting future market trend. Behind the forecaster, there are algorithms, database and more mathematical calculations. In light of it, this project plan outlined the idea of how this project will be carried out.
Stock market is a huge market and different scope will have different properties on the market so this project limited the scope of market which is only in Hong Kong with a fixed number of stocks only. With a limited scope, more accurate calculation should be obtained. After various studies on data-mining methods and financial indicators, artificial neural network is chosen to be the backbone of the project. In this project plan, there will be more explanations of different part of this project.

3 Introduction

The introduction gives an overview of the whole project including why this project is held and what this project is about. Predicting market trend is not a new stuff in this world yet this issue is kept being discussed by various parties. Being able to predict accurately the future financial outcome is equivalent to earning big money. This project aims at analyzing this problem in an academic way which provides a different way of prediction on the market trend.

3.1 Project Overview

Stock market prediction has always been regarded as a challenging task in the business field. There are financial models trying to describe the trends of stock market price also data-mining methods trying to find out non-random movements of stock price based on historical stock data. In contrast, some proposed that the stock market trend is non-predictable because the trend of stock is governed by random walk[i].

In this project, it is believed that market trend is a financial time-series prediction problem which historical data is able to give some hints on predicting future price of stock market. There are studies trying to solve this problem by means of artificial neural networks (ANNs), support vector machine (SVM) or other data-mining methods which attain a certain extend of success. However, there are also limitations in those studies like over-fitting problem.

This project aims at finding out an algorithm which can most fit Hong Kong's stock market prices using machine learning and financial models. We hope that by hybridizing different algorithm constructed by past studies, this project can eliminate limitations from each method. Different sets of real stock data will be training sets and experiments will be carried out in order to verify the accuracy of the algorithm.

3.2 Project Deliverables

This project will create a financial data forecaster program based on the algorithm to be developed by this project. There will be backend database storing all historical stock prices available for supporting the forecaster and data will be treated as training sets for the program. To predict the stock prices in Hong Kong, this project aims at developing a program which serve several functions including predicting trends, predicting tomorrow's closing price and formulating the best solution for buying stock according to the predicting result.

4 Objectives

4.1 Scope of Project

In this project, 15-20 stocks from Hang Seng Index (HSI) components will be selected for prediction. Prediction will base on historical market data only hence news or company background will not be included in the prediction.

4.2 Algorithm

The prediction will not only base on a single financial indicator or just a few indicators. In contrast, the algorithm to be developed aims at combining multi financial indicators and different market strategies by machine learning methods such as artificial neural networks (ANNs) and finding out the complex correlations between indicators.

In addition to correlations between indicators, the algorithm aims to predict the closing price of the next day and also the rough market trend of that particular stock. After the prediction, the best action to be taken (buy or sell) will be generated for reference of client.

5 THEORETICAL BACKGROUND

5.1 TECHNICAL ANALYSIS

Technical analysis has been widely used by investors to predict the direction of future market prices through studying past market data. Technical analysis is based on financial indicators that is mathematically transformed from prices and volume. These indicators include relative strength index (RSI), sample/exponential moving average, moving average convergence/divergence (MACD), Williams %R, stochastic oscillator, etc. One fundamental principle of technical analysis is that it only makes use of market prices and trend as the trace of prediction as it is believed that market price presents all information that reflects the stock and that information is unbiased[ii].

As technical analysis does not consider external factors such as news, business's financial statements or state of the economy, it can be seen as the candidate to be modelled with machine learning approaches. Technical analysis implies that financial modelling can be achieved by using historical market data as input to the approximation model and output predicted result, and it also provides clues to the usage of different technical indices which can be applied such as setting the initial condition of the learning model. As all data is numerical, they can be easily normalized and manipulated as input values to the approximation functions.

5.1.1 Financial Indicators

Technical analysis techniques rely on analysis of the trend of financial indicators. Following are some financial indicators selected which helps in predicting the future stock prices.

Financial Indicators	Formula	Description
Absolute Breadth Index (ABI)	abs(# of Advancing Stocks - # of Declining Stocks)	Market momentum indicator
Accumulation/Distribution Line (AD)	$\frac{(CLOSE - LOW) - (HIGH - CLOSE)}{HIGH - LOW} \times VOLUME + 1$	Momentum indicator
Advance Decline Line (ADL)	(# of Advancing Stocks - # of Declining Stocks) + Previous Period's A/D Line Value	Market momentum indicator
Directional Movement Index (ADX/DMI)	Depends on +DI & -DI	Lagging indicator
Average True Range (ATR)	Depends on three different true range	Volatility indicator
Breadth Thrust Index (BTI)	MA(Up / Up + Down, N)	Momentum indicator
Commodity Channel Index (CCI)	(Typical Price - 20-period SMA of TP) / (.015 x Mean Deviation)	Versatile indicator
Chaikin Oscillator	EMA(A/D, n) - EMA(A/D,m)	Momentum indicator
Moving Average Convergence/Divergence (MACD)	MACD = EMA[12] – EMA[26] Signal = EMA[9] of MACD Histogram = MACD - Signal	Lagging indicator
Market Facilitation Index (BW MFI)	Range*(High – Low)/Volume	Willingness of the market to move the price

Money Flow Index (MFI)	Typical price -> Raw money flow -> Money flow ratio -> MFI	Momentum indicator
Momentum Indicator	Closing – Closing(n)	Momentum indicator
Percentage Price Oscillator (PPO)	{(12-day EMA - 26-day EMA)/26-day EMA} x 100	Momentum indicator
Percentage Volume Oscillator (PVO)	((12-day EMA of Volume - 26-day EMA of Volume)/26-day EMA of Volume) x 100	Momentum indicator
Relative Momentum Index (RMI)	RMI(m,n) = RM(n)/1+RM(n)	Momentum indicator
Relative Strength Index (RSI)	RSI(n) = 100 – 100/1+RS(n)	Momentum indicator
Stochastic Oscillator (KD)	%K = 100*(Closing – Low(%K)/High(%K)-Low(%K))	Momentum indicator
William's %R	100 – (Closing-Low(n)/Hign(n)-Low(n))	Momentum indicator

5.2 Data Mining Methods

5.2.1 Artificial Neural Network (ANNs)

Artificial neural network (ANN) is an artificial intelligence modeling method that can model complex linear and non-linear function or relation between given input variables. The principle of artificial neural network is to model the complex processing power of the biological nervous system to achieve statistical learning and approximation of functions. The general idea of ANN is to create a network of interconnecting computing element which models the biological neuron, that will exchange information among each other. Each connection contains a weight which is a simplification of the complex structure of the network. The structure of the network can be altered by changing the numerical weight, making ANN adaptive and capable for machine learning by reflectively altering the weight in each connection.

5.2.2 Training algorithm

One major learning algorithm to train a neural network is the backpropagation algorithm[iii]. As financial data are time series data, input data of a particular time will have the market price of near future as the desired output. As the desired output is known, the error of the approximation can be calculated. After the initial network is randomly setup (or setup up based on knowledge), the backpropagation algorithm will correct the structure of the network in terms of the connection weight depending on the error of the feed-forward computation. The backpropagation algorithm can be roughly divided into the following steps:

1. Feed-forward computation
2. Propagate error to the output
3. Propagate error to the hidden layer
4. Update weight

After repeating the steps until the output error has become sufficiently small, the algorithm stops and the model should have a desired accuracy.

6 METHODS, TOOLS AND TECHNIQUES

Among many machine learning and data mining methods, Artificial Neural Network (ANNs) is chosen to be the machine learning method in this project. As financial modelling involve analysis on very complex relationships between individual input index, ANN is a suitable candidate for this project[iv]. In addition to support ANN, python is chosen to be the programming language to use.

6.1 APPLICATION OF ARTIFICIAL NEURAL NETWORK

Collected datasets are planned to be divided into training and testing sets. Possible division can be using 80% of the dataset as training data and remaining 20% as testing data. The training datasets will be used to define the structure of our network by applying training algorithms. During the training phase, different parameters such as the training rate, number of hidden neurons, network weights and biases are adjusted. After the training phase, the testing data will be fed in to the network and the error and accuracy of the model is calculated.
Different aspects of the ANN being built have to be considered:

- Number of layers
- Training algorithm to use
- Structure of the network (recursive or not)

For the initial model, a single hidden layer network that trained with the backpropagation algorithm will be built. After the results of the initial model are obtained, other different configuration of the network will be tried in order to compare and find out the best solution.

6.2 PROGRAMMING LANGUAGE AND TOOLS

Python will be the programming language used in this project due to its platform independency and abundance of machine learning library. Pybrain and Scikit tool are two libraries planned to be used in implementation state as both library provide algorithms for ANN. In addition, Scikit tool provides a wider variety of machine learning algorithms while Pybrain focuses more on ANN development.

7 PROJECT ORGANIZATION

This project can be mainly divided into four major parts including collection and organization of data, review of related materials, development of new algorithm, experiment and refine of the algorithm.
For collection and organization of data, all information related to Hang Seng Index (HSI) components will be collected and stored in a format suitable for database use. Data will be collected every 15 minutes so that a more detailed trend can be obtained. This process will be done by all teammates to ensure no missing of data.
Besides this routine job, review of related materials will be done in the early stage of the project. Materials will be mainly related to two aspects which are financial modeling and data-mining. All teammates should be familiar with both aspects in order to carry out the best outcome for this project yet a small division of focus will be necessary for more in depth understanding. Justin will focus more on data-mining methods and usage while Louis will focus more on financial indicators and financial modeling.

Combining all knowledge after review of materials, new algorithm will be developed together. With the slightly difference in review of materials, algorithm maybe slightly varies for each teammate yet the backbone will be mainly the same. In the development stage, the implementation part of ANN will be mainly done by Justin while each calculation of financial indicators and the database management will be mainly done by Louis. After developing the algorithm, each teammate will be responsible for implementing it and testing it.
For experiment, past cases will be used for testing also continuously assessment on future prediction will be evaluated. As there should be two slightly different programs at the time of experiment, so refinement will be done based on evaluation of both programs.

8 SCHEDULE

This project will last roughly a year and different tasks are identified and set with a date to be completed. Routine work like collection of data and review of materials will be done through out the project. The following table shows the detailed schedule with different stages identified.

Date	Stages	Details
August - September	Review of materials	Materials related to data-mining methods, financial modeling and stock market analysis will be studied
4th October	Phase 1 (Inception)	Detailed project plan and project web page should be finished
October – January	Development of algorithm (1st)	First draft of algorithm will be created with least information to be input to ensure the algorithm works
11-15th January	First presentation	Able to present the first draft of the algorithm and give a whole picture on the method used
January	Organization of data	Data will be organized in the best way for inputting into database
24th January	Phase 2 (Elaboration)	A rough implementation of first draft of algorithm should be runnable to test and detailed interim report should be done
January - February	Development of algorithm (2nd)	Second draft of algorithm will be created with more types of information to be input so that a more complex algorithm can be formed
February	Implementation of program	Program will be created according to the second draft of algorithm
March	Experiment on program	Real data will be input into the program to test the accuracy
March - April	Amendment on algorithm and program	According to the experimental result, amendments will be made on both algorithm and program
17th April	Phase 3 (Construction)	Final implementation of the program should be done with a well tested result also final report should be done

List of appendices

[i] Cootner, Paul H (1964), The random character of stock market prices
[ii] Basu, S. (1977). Investment performance of common stocks in relation to their price- earnings ratios: A test of the efficient market hypothesis. The journal of Finance, 32(3), 663-682.
[iii] R. Rojas (1996), Neural Networks - A Systematic Introduction, trans. J. Feldman, Springer-Verlag, Berlin, New-York
[iv] S. Ward and M. Sherald (1995), The Neural Network Financial Wizards, Technical Analysis of Stocks and Commodities, Reprinted, Technical Analyses Inc.,Seattle, Washington 1995