Predicting HK internet industry stock price index movement

based on machine learning techniques

袁一公 南京大学

1. Introduction

This study tries to analyze the daily movement (increase or decrease) of HK internet industry stock price index based on some widely used techniques of machine learning. Similar studies have been conducted on American and Indian stock markets guiding Portfolio Managers to make trading decisions. However, the accuracy for continuous data prediction is generally lower, saying the future price. Therefore, we choose to analyze the moving direction.

The topic of HK internet industry corporation value study is rather meaningful at the current situation. Although Chinese internet tech corporation has experienced a sharp growth in the past decade, we all know that in the previous month Hongkong stock markets have seen a great decline, among which the internet industry suffers most. The reasons for the shock include but not limited to political condition, global economy and companies’ financial performance etc. At this moment of panic, investors are eager to know whether they should restore faith in the internet industry or hold short position. By utilizing technical analysis based on machine learning, we can give our own answer to this question.

Reference:

[1] Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques, Jigar Patel, Sahil Shah, Priyank Thakkar ⇑

, K. Kotecha

[2] Stock Closing Price Prediction using Machine Learning Techniques, Mehar Vijha , Deeksha Chandolab, Vinay Anand Tikkiwalb, Arun Kumar

[3] Forecasting stock index returns using ARIMA-SVM, ARIMA-ANN, and ARIMA-random forest hybrid models, Manish Kumar

[4] Predicting the Daily Direction of the S&P500, Udacity Machine Learning Engineer Nanodegree

[5] 千亿资金南下 借道ETF抄底港股, 葛瑶

1. Data & Method

To analyze the fluctuation of general market, internet technology industry and specific corporations, I select 2 stock price indices (Hongkong HSI, HSIII) and 2 stock () All the data comes from the Choice Financial Platform.

The study compares four prediction models, Artificial Neural Network (ANN), support vector machine (SVM), random forest and Naive-Bayes with two approaches for input to these models. The first approach for input data involves computation of ten technical parameters using stock trading data (open, high, low & close prices) while the second approach focuses on representing these technical parameters as trend deterministic data. Accuracy of each of the prediction models for each of the two input approaches is evaluated. Evaluation is carried out on 10 years of historical data from 2003 to 2012 of two stocks namely Reliance Industries and Infosys Ltd. and two stock price indices CNX Nifty and S&P Bombay Stock Exchange (BSE) Sensex.

首先，将数据加工为9组常用指标，并进行数据筛选分成3组用于模型拟合，每年取出占总样本20%的数据作为组外数据，留80%数据作为组内数据。进一步，比较四种预测模型，人工神经网络 (ANN)、支持向量机 (SVM)、随机森林和朴素贝叶斯，以及这些模型的两种输入方法。 输入数据的第一种方法涉及使用股票交易数据（开盘价、最高价、最低价和收盘价）计算十个技术参数，而第二种方法侧重于将这些技术参数表示为趋势确定性数据。 评估运用两种输入方法时每一种的预测模型的准确性。 评估的基础是Reliance Industries、Infosys Ltd. 两只股票以及 CNX Nifty、S&P Bombay Stock Exchange (BSE) Sensex 两个股票价格指数从 2003 年到 2012 年的 10 年历史数据。