Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 127 (2018) 161–169

Procedia Computer Science

www.elsevier.com/locate/procedia

The First International Conference on Intelligent Computing in Data Sciences

Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange

Anass Nahil[a,*], Abdelouahid Lyhyaoui[a]

[a] *Laboratory of Innovative Technologies, Abdelmalek Essaadi University, Tangier 90000, Morocco*

Abstract

Stock market prediction is a key problem to the financial field. An accurate prediction model may yield profits for investors. However, the volatile nature of the stock market makes it difficult to develop efficient models. This study attempts to improve the prediction capacity of the stock price via an integrated prediction model based on Kernel Principal Component Analysis (KPCA) and Support Vector Machines for Regression (SVR). KPCA is firstly introduced to reduce the feature dimensions. On that basis, SVR is used to build a short-term investment decision system. 23 technical indicators are calculated for five major Moroccan banks listed on the Casablanca Stock Exchange (Casablanca S.E) and used as input of the models. A comprehensive parameters setting is performed to improve the prediction performance. The simulation results show that, through KPCA attribute reduction, the structure of the investment decision system can be simplified significantly with improvement of the model performance. The average performance of the integrated model that uses KPCA and SVR is significantly better than that of SVR model, which verifies the effectiveness and accuracy of the proposed method.

Keywords:
Support vector machines; Kernel principal component analysis; Prediction; Stock price movement

1. Introduction

Predicting future values of a time-series transcends a range of disciplines. Time-series analysis finds application in numerous and diverse domains such as stock price predicting. This domain is one of the most challenging applications of time series prediction. Even though many researches have been conducted on stock price predicting, most empirical findings are associated with the developed markets. Few researches have studied the stock price evolution in emerging financial market, especially in Moroccan stock market.

* Corresponding author. Tel.: +212-660-615-004
E-mail address: anassnahil@gmail.com

10.1016/j.procs.2018.01.111

It is of interest to study the extend of stock price index movement predictability using data from emerging markets such as that of Morocco. Established in 1929, the Casablanca S.E is one of the oldest exchanges in Africa. Major reforms providing an improved legal framework for a growing number of domestic investors took place in 1967; in 1993 reforms were undertaken to refine investor protection, promote transparency and introduce electronic trading. Those reforms culminated in 1997 by the launch of Maroclear, the clearing house of Casablanca S.E. An indicator of the success of reforms can be found in the market capitalization rate: The ratio of stock market capitalization to GDP jumped from 23.8% in 2002 to 86.1% in 2006. The Casablanca S.E is characterized with high volatility in the market returns. Such volatility attracts many local and foreign investors as it provides high return possibility [1].
The financial market is a complex, evolutionary, and nonlinear dynamical system [2]. The financial time series are non-stationary, noisy, and deterministically chaotic [3]. This means that their distribution is changing over the time. Their stationarity is not only in the sense of the mean and variance of the series, but the relationship of the data series to other related data series may also be changing. Modeling such dynamical and non-stationary time series is expected to be a challenging task.
Neural networks have been successfully used for modeling financial time series ranging from stock index trading [4], corporate bond rating [5] to currency exchange [6]. Neural networks can map any nonlinear function without a priori assumption about the data [7]. Unlike traditional statistical models, neural networks are data-driven, they let the data speak for themselves. So neural networks are more powerful in describing the dynamics of financial time series than traditional statistical models [8]. Support vector machines (SVM), has been successfully applied to predict stock price index and its movements. Wen, Q. et al. (2010) [9] proposed a new intelligent trading system based on oscillation box prediction by combining stock box theory and support vector machine algorithm. G. Lin, Y., Guo, H., and Hu, J. (2013) [10] applied a correlation-based SVM filter to rank and select a good subset of financial indexes. Yu, H., Chen, R., and Zhang, G. (2014) [11] used a PCA-SVM stock selection model that, significantly, outperform the A Share Index of Shanghai Stock Exchange. Gong, X. et al. (2016) [12] proposed a novel approach for locating chart patterns in financial time series using and extended UCR suite and SVM.
Feature extraction is crucial in developing an efficient SVR forecastor. The stock price data often contain a lot of valuable information such as investment intentions and behaviors. At the same time, with larger stock trading data volume, there are usually more irrelevant indices, making the data contain a certain amount of redundant information. This affects effective prediction of stock price trend. In the modeling, all available indicators can be used as the inputs of SVR, but irrelevant or correlated ones could deteriorate the generalization performance of SVR. Principal Component Analysis (PCA) is a well-known method for feature extraction. By calculating the eigenvectors of the covariance matrix of the original inputs, PCA linearly transforms a high-dimensional input vector into a low dimensional one whose components are uncorrelated. KPCA is a nonlinear PCA developed by generalizing the kernel method into PCA [13]. KPCA Firstly maps the original inputs into a high-dimensional feature space using the kernel method and then calculates principal component (PC) in the high-dimensional feature space. The linear PCA in the high-dimensional feature space corresponds to a nonlinear PCA in the original input space.
The purpose of this paper is to predict the price movement in the Casablanca S.E using KPCA and SVR. The major contributions of this study are to demonstrate and verify the predictability of stock price using KPCA and SVR and to evaluate the added value of the KPCA for feature extraction in the context of SVR.
The remainder of this paper is organized into 5 sections. Sections 2 and 3 provide a brief overview of the theoretical literature about KPCA and SVR. Section 4 describes the proposed methodology. Finally, Section 5 contains results and discussion.

2. Kernel principal component analysis

PCA is a linear dimension reduction method, which can only handle linearly correlated data. KPCA is developed to over-come such weakness by conducting nonlinear process monitoring [14, 15, 16]. KPCA projects original data space into a high-dimensional feature space before implementing PCA operation. The input matrix can be obtained as $X = [x_1, x_2, , x_n]^T \in R^{n\times m}$ where x_i is the observation vector at time i. Mapping from dataspace to feature space is implemented with the following nonlinear mapping function $\phi(.)$:

$$R^m \xrightarrow{\phi(.)} F^h \tag{1}$$