Applied Mathematics and Computation 320 (2018) 444–451

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Short Communication

Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression

Salim Lahmiri

ESCA School of Management, 7, Abou Youssef El Kindy Street, BD Moulay Youssef, Casablanca, Morocco

ARTICLE INFO

Keywords:
Intraday stock price
Time series
Singular spectrum analysis
Support vector regression
Particle swarm optimization
Forecasting

ABSTRACT

Time series modeling and forecasting is an essential and hard task in financial engineering and optimization. Various models have been proposed in the literature and tested on daily data. However, a limited attention has been given to intraday data. In this regard, the current work presents a model for intraday stock price prediction that uses singular spectrum analysis (SSA) and support vector regression (SVR) coupled with particle swarm optimization (PSO). In particular, the SSA decomposes stock price time series into a small number of independent components used as predictors. The SVR is applied to the task of forecasting and PSO is employed to optimize SVR parameters. The performance of our proposed model is compared to the performance of four models widely used in financial prediction: the wavelet transform (WT) coupled with feedforward neural network (FFNN), autoregressive moving average (ARMA) process, polynomial regression (PolyReg), and naïve model. Finally, the mean absolute error (MAE), mean absolute percentage error (MAPE), and the root mean of squared errors (RMSE) are used as main performance metrics. By applying all models to six intraday stock price time series, the forecasting results from simulations show that the presented SSA-PSO-SVR largely outperforms the conventional WT-FFNN, ARMA, polynomial regression, and naïve model in terms of MAE, MAPE and RMSE. Therefore, our proposed predictive system SSA-PSO-SVR shows evident potential for noisy financial time series analysis and forecasting.

1. Introduction

Time series modeling and prediction is an important task in different applications [1–6]. In particular, several multiresolution decomposition methods have been adopted to analyze time series; including the traditional Fourier transform which is a basic tool in signal processing used to approximate the signal in terms of sinusoids and the wavelet transform [7] that simultaneously decomposes a signal into subsequences at different resolution time scales. In recent years, the singular spectrum analysis (SSA) of time series [8] has received a growing attention as a non-parametric time series modeling technique where an observed time series is unfolded into the column vectors of a Hankel structured matrix, known as a trajectory matrix [9]. In particular, the purpose is to unfold a time series into a trajectory matrix whose singular values are then determined to reconstruct a smoother time series which can be used for explaining structure and for forecasting [9]. Indeed, it decomposes the original time series into a sum of a small number of components: slowly varying trend, oscillatory components, and noise [10].

E-mail address: slahmiri@esca.ma

https://doi.org/10.1016/j.amc.2017.09.049
0096-3003/

Since forecasting stock market time series is receiving a large attention in the literature [11–15] because it is a major issue in economics and business applications and decision making, the purpose of this paper is to apply the SSA technique in forecasting intraday stock market prices to enrich the literature related to SSA application in this topic. Indeed, in this paper, we present a prediction system that uses SSA for time series decomposition, support vector regression (SVR) [16] for learning and prediction, and particle swarm optimization (PSO) which is a global optimization method [17] for SVR initial parameters optimization. As a version of the support vector machine (SVM) [18], the main advantage of the SVR is applying the structural risk minimization principle to minimize an upper bound on the generalization error rather than implementing the empirical risk minimization principle to minimize the training error [18]. Therefore, it theoretically guarantees to achieve the global optimum. In this regard, recently, the SVM and SVR have been successfully employed to solve non-linear regression and classification of time series problems in finance and engineering [19,20].

The performance of the presented forecasting system will be compared to that of conventional system used in the literature which is based on wavelet transform (WT) [7] for financial data decomposition and feedforward neural network (FFNN) [21] for forecasting [22]. In addition, the autoregressive moving average (ARMA), the polynomial regression (PolyReg) and naïve models are also considered in this study for comparison as well.

In sum, the contributions and highlights of this paper can be summarized as follows. First, we develop a hybrid approach based on the SSA financial time series processing technique and support vector regression for time series forecasting. Second, particle swarm optimization which is a global optimization heuristic technique is adopted to optimize SVR initial parameters. Third, the presented forecasting model is applied to a set of individual stocks rather than stock indices to gain further insight on the applicability of the proposed predictive system. Fourth, results are compared against three benchmark models: FFNN trained with WT coefficients (WT-FFNN), polynomial regression (PolyReg), naïve model, and the classical ARMA process.

The remainder of this paper is organized as follows. In Section 2, the technical methods are presented. The forecasting results are presented in Section 3. Finally, Section 4 concludes the paper.

2. Methods

2.1. Singular spectrum analysis

In this section we provide a brief review of the SSA as adapted from [23–26]. The SSA is based on the singular value decomposition (SVD) of the trajectory matrix, derived from the original time series [25]. In particular, it decomposes the time series into an additive set of independent principal components [27]. The basic SSA method consists of two complementary steps [24]; namely the decomposition and the reconstruction step; where each step includes two separate steps. The original signal is decomposed and reconstructed respectively in the first and second step. For instance, the decomposition step includes an embedding operation followed by singular value decomposition (SVD). Finally, the reconstruction step includes grouping and diagonal averaging operation. Each step is described following [24] and [25] notations.

In the embedding procedure, the purpose is to map a one-dimensional time series f of length N into an $l \times k$ matrix with rows of length l as follows:

$$\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_k] = \begin{bmatrix} f_0 & f_1 & \cdots & f_{k-1} \\ f_1 & f_2 & \cdots & f_k \\ \vdots & \vdots & \ddots & \vdots \\ f_{t-1} & f_t & \cdots & f_{r-1} \end{bmatrix} \tag{1}$$

where the trajectory matrix $\mathbf{X}$ is a Hankel matrix, vectors $\mathbf{x_i}$ are called l-lagged vectors, $k = r - l + 1$ is the number of windows ($1 \le l \le r$). The embedding operation is followed by applying SVD to the trajectory matrix X to represent it as a sum of rank-one bi-orthogonal elementary matrices. The SVD of the trajectory matrix is given by:

$$\mathbf{X} = \sum_{i=1}^{d} \mathbf{X}_i = \sum_{i=1}^{d} \sqrt{\lambda_i \mathbf{u}_i \mathbf{v}_i^t} \tag{2}$$

where λ_i ($i = 1,\ldots,l$) are the eigenvalues, in decreasing order of magnitude, of the covariance matrix $\mathbf{C}_\mathrm{X} = \mathbf{X}^t\mathbf{X}$, d_i ($i = 1,\ldots,l$) are the corresponding orthogonal eigenvectors, subscript t denotes the transpose of a vector, and v_i is given by:

$$\mathbf{v}_i = \mathbf{X}^t \mathbf{u}_i / \sqrt{\lambda_i} \tag{3}$$

In the grouping step, the elementary matrices X_i are split into several groups and a summation of matrices is performed within each group. For instance, the indices $J = 1,\ldots,d$ are grouped into M disjoint subsets $I_1,\ldots,I_M$ corresponding to split the elementary matrices $X_i = 1,\ldots,d$ into M groups; where Each group contains set of indices as $I = \{i_1,\ldots,i_p\}$. As a result, the matrices X_I and X are respectively given by:

$$X_I = X_{i_1} + \cdots + X_{I_p} \tag{4}$$

$$X = X_{I_1} + \cdots + X_{I_M} \tag{5}$$