Sliding-Window based Stock price forecasting expert system using LSSVR

NAVEEN KUMAR M -2015103519

UMAPATHI C -2015103613

ROHITH KUMAR K -2015103527

PROJECT GUIDE **Dr.ANGELIN GLADSTON**

Introduction

- **Financial markets are highly volatile and generate huge amounts of data daily**
- **It is the most popular financial market instrument and its value changes quickly**
- **Stock prices are predicted to determine the future value of companies stock or other financial instruments that are marketed on financial exchanges**
- **However the stock market is influenced by many factors such as political events, economic conditions and traders expectation**

Overall Objectives

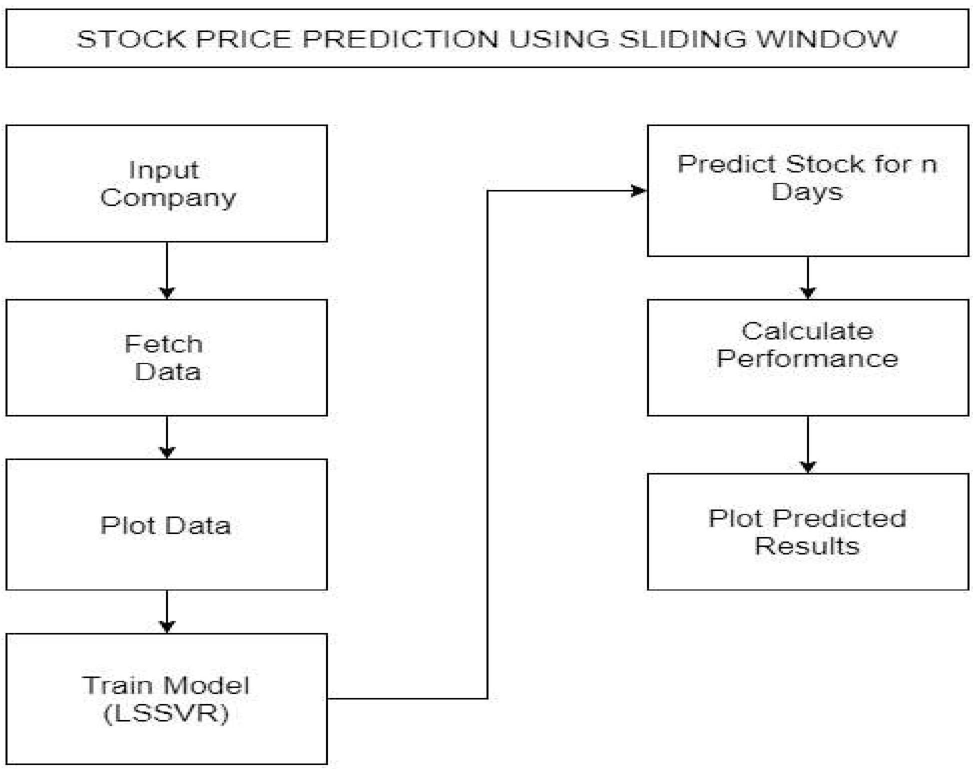
- **The main objective of this project is to achieving the best accuracy in predicting the stock prices in forecasting expert system using LSSVR**
- **Applying some selected Machine Learning algorithms for the limited dataset .**
- **Predicting the stock market involves predicting the closing prices of a company's Stock for any given number of days ahead.**
- **The field of machine learning is vast and plays a key role in a wide range of critical applications**

Literature Survey

TITLE	AUTHOR	YEAR	DESCRIPTION
Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information	J.-S. Chou, A.-D. Pham	2017	This article puts forward a revised reliability analy- sis based on SVM, namely LSSVR-MCS method. The LSSVR-MCS method transforms the inequality con- straint of SVR-MCS into equality constraint so as to change the solving algorithm of the support vector machine from quadratic programming to a linear equa- tion set and make the solving approach easier.
Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis	. M. Anish,B. Majhi	2016	A feedback type of the functional link artificial neural network (FFLANN) with recursive least square (RLS) training is proposed as a potential prediction model.
Efficient stock price prediction using a self evolving recurrent neuro-fuzzy inference system optimized through a modified differential harmony search technique	R. Dash and P. Dash	2016	The novelty of the model is based on the fact that the internal temporal feedback loops and time delayed output feedback loops are used for further enhancing the prediction capability of traditional neuro-fuzzy system in handling more dynamic financial time series data

Secondary factor induced stock index time-series prediction using self-adaptive interval type-2 fuzzy sets	D. Bhattacharya, A. Konar, and P. Das	2016	(i) employing a new strategy to induce the main factor time-series prediction by its secondary factors and (ii) self-adaptation of membership functions to properly tune them to capture the sudden changes in the main-factor time-series.

Block Diagram



Details of Modules

- **Phase Space Reconstruction**
- **Sliding-window method**
- **Least Squares Support Vector Regression**

Phase Space Reconstruction

- **In time series prediction, the time series are typically expanded into three or higher dimensional space to exploit the information that is implicit in them.**
- **Selecting a suitable pairing of embedding dimension m (lag) and time delay τ is very important for phase space reconstruction.**

- **Consider a time series x={x1,x2,x3…xn}.The time-delay vectors can be reconstructed as follows, where X is the input matrix and Y is the corresponding output matrix**
- **The output of the analysis is fed back to the input and future values are predicted from previous values in the time series.**

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \xrightarrow{m} X = \begin{bmatrix} x_1 & x_2 & \cdots & x_{m-1} & x_m \\ x_2 & x_3 & \cdots & x_m & x_{m+1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{N-m-\tau+1} & x_{N-m-\tau+2} & \cdots & x_{N-\tau-1} & x_{N-\tau} \end{bmatrix}, \quad Y = \begin{bmatrix} x_{m+\tau} \\ x_{m+1+\tau} \\ \vdots \\ x_N \end{bmatrix}$$

Sliding-window method

- **The forecast horizon is 1. In the first validation, the working window includes p historical observations (1,2,3, …) which are used to forecast the next value +1.**
- **In the second validation, the oldest value 1 is removed from the window and the latest value +1 is added, keeping the length of the sliding window constant at p. The next forecast value will be +2.**

- **The window continues to slide until the end of the dataset is reached. If the number of observations is N, then the total number of validations is (N-p).**

```
Algorithm : slidingWindow
Input   : data          [stock data]
Output  : A data frame of a lagged stock data
    LAG ← 1
    y ← remove first LAG rows from data
    reset row indices of y
    x ← remove last LAG rows from data
    train ← merge (x,y) into a dataframe
    rename column name to x and y
    return train
```

Least Squares Support Vector Regression

- **The LSSVR approach is a well-developed ML technique with many advanced features that support a high generalization capacity and fast computation.**
- **The LSSVR training process entails the use of a least squares cost function to obtain a linear set of equations in a dual space to minimize the computational cost**
- **Accordingly, iterative methods, such as the conjugate gradient method are typically used to derive a solution by efficiently solving a set of linear equations.**
- **To reduce the computational burden of the LSSVR for function estimation, the regression model in this study uses a quadratic loss function.**

- **The least squares version of the SVM classifier is obtained by reformulating the minimization problem as:**

$$\min J_2(w,b,e) = \frac{\mu}{2} w^T u$$

- **For the kernel function K(,) one typically has the Radial Basis Function:**

$$K(x, x_i) = \exp\Big(-\|x - x_i\|$$

- **The LSSVR involves equality instead of inequality constraints and works with a least squares objective function.**
- **The LSSVR approach considerably reduces computational complexity and increases efficiency compared to standard SVM.**
- **LSSVR solves linear equations instead of a quadratic programming problem.**

Performance Measures

- **The performance metrics that were used to assess the predictive accuracy of the proposed system included the RMSE, the MAE, the MAPE, the MSE, the correlation coefficient (R), the nonlinear regression multiple correlation coefficient (R2),and the synthesis index (SI).**
- **These indexes are used to measure whether the predicted values are close to the actual values.**

- **Table summarizes the formulas for these indexes, where y is the actual value, y is the predicted value, ¯ yi is the average across data samples, n is the number of data samples, m is the number of performance measures, and Pi is the ith performance measure.**
- **The SI ranges from 0 to 1 and an SI value of close to 0 indicates a highly accurate predictive model**

TABLE I
MATHEMATICAL FORMULAS FOR PERFORMANCE MEASURES

Measure	Formula
RMSE	$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y' - y)^2}$
MAE	$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \lvert y - y' \rvert$
MAPE	$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left\lvert \frac{y - y'}{y} \right\rvert$
MSE	$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n} (y - y')^2$
The correlation coefficient (R)	$R = \frac{n\sum y \cdot y' - \left(\sum y\right)\left(\sum y'\right)}{\sqrt{n\left(\sum y^2\right) - \left(\sum y\right)^2}\sqrt{n\left(\sum y'^2\right) - \left(\sum y'\right)^2}}$
Nonlinear regression multiple correlation coefficient (R^2)	$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_{i+1} - y'_i)^2}{\sum_{i=1} (y_i - \bar{y}_i)^2}$
SI	$\mathrm{SI} = \frac{1}{m}\sum_{i=1}^{m} \left(\frac{P_i - P_{\min,i}}{P_{\max,i} - P_{\min,i}}\right)$

DataSet

- **Historical daily prices were taken from Yahoo! Finance, a publicly accessible website, as they were by Six years (October 5, 2011 to May 31, 2017) of daily data on five company stocks were downloaded from Yahoo! Finance.**
- **The data were closing stock price**

References

1. **C. M. Anish and B. *Majhi Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis* J. Korean Statist. Soc. 2016.**
2. **R. Dash and P. *Dash Efficient stock price prediction using a self evolving recurrent neuro-fuzzy inference system optimized through a modified differential harmony search technique Expert Syst*.2016.**
3. **Bhattacharya, A. Konar, and P. Das,*Secondary factor induced stock index time-series prediction using self-adaptive interval type-2 fuzzy sets*,2016.**
4. **J.-S. Chou, K.-H. Yang, J. P. Pampang, and A.-D. Pham, *Evolutionary metaheuristic intelligence to simulate tensile loads in reinforcement for geosynthetic-reinforced soil structures, Comput. Geotechnics*, 2015.**
5. **J.-S. Chou and A.-D. Pham, *Smart artificial firefly colony algorithm based support vector regression for enhanced forecasting in civil engineering*, Comput.-Aided Civil Infrastructure Eng, 2015.**
6. **D. Saini, A. Saxena, and R. C. Bansal, *Electricity price forecasting by linear regression and SVM,* in Proc. Int. Conf. Recent Adv. Innov. Eng, 2016.**
7. **J. Wang, R. Hou, C. Wang, and L. Shen, *Improved v-support vector regression model based on variable selection and brain storm optimization for stock price forecasting,* Appl. Soft Comput, 2016.**
8. **A. Jindal, A. Dua, K. Kaur, M. Singh, N. Kumar, and S. Mishra, *Decision tree and SVM-based data analytics for theft detection in smart grid*, IEEE Trans. Ind. Informat,2016.**