ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

Iterative ARIMA-Multiple Support Vector Regression models for long term time series prediction

João Fausto Lorenzato de Oliveira and Teresa B. Ludermir

Federal University of Pernambuco - Center of informatics
Av. Jornalista Anibal Fernandes, s/n, Recife, PE, 50.740-560, Brazil

Abstract. Support Vector Regression (SVR) has been widely applied in time series forecasting. Considering long term predictions, iterative predictions perform many one-step-ahead predictions until the desired horizon is achieved. This process accumulates the error from previous predictions and may affect the quality of forecasts. In order to improve long term iterative predictions a hybrid multiple Autoregressive Integrated Moving Average(ARIMA)-SVR model is applied to perform predictions considering linear and non-linear components from the time series. The results show that the proposed method produces more accurate predictions in the long term context.

1 Introduction

A time series is a discrete-time stochastic process composed by a finite set of items correlated in time. Time series forecasting has many applications in areas such as planning, management, production, maintenance and control of industrial processes. The forecasting process is based on some knowledge of the past, which can be performed by several statistical methods [1] and by nonlinear methods such as Support Vector Regression (SVR) [2] and Artificial Neural Networks (ANNs) [3].

Some time series may present seasonal or\and trend patterns, which must be modeled by a method in order to produce better forecasts. Statistical methods such as ARIMA are able to map linear aspects such as trend, however nonlinear patterns are not easily captured. Different models can complement each other in capturing linear or nonlinear patterns of data sets in time series. Bates and Granger [4] explored several combinations of architectures for time series prediction, Pai and Lin used a hybrid ARIMA-SVR model to predict stock prices [5], Zhu and Wei used a hybrid ARIMA-Least Squares Support Vector Regression(LSSVR) to predict carbon price.

However, when performing iterative multi-step-ahead predictions, the accumulation of error may still be present at each forecast iteration. In order to address this problem, Zhang et al. used a multiple SVR [6] strategy where the SVR models are trained independently in different prediction horizons. This method reduces the number of iterations and may reduce the propagated error by performing many direct predictions.

The authors would like to thank FACEPE, CNPq and CAPES (Brazilian Research Agencies) for their financial support

In this paper we propose a Multiple ARIMA-SVR model in order to map linear and nonlinear patterns in time series and to improve the accuracy in iterative long term forecasts. The remainder of this paper is organized as follows: Section 2 presents a brief review of time series forecasting strategies, Section 3 shows the proposed method. Section 4 presents the experiments and the conclusions and future works are presented in section 5.

2 Time series forecasting

In this section, the iterated, direct and multiple approaches for time series forecasting are reviewed. For all three approaches consider a time series $\{z_t\}^N{}_{t=1}$, where N is the length of the time series.

In the one-step-ahead prediction the model depends on the d past values and has the form $z'(t+1) = f(z(t), z(t-1), ..., z(t-(d-1)))$, where d is the embedding dimension or time window.

The prediction of p-steps-ahead in the iterative approach takes the form $z'(t+h) = f(z'(t+h-1), z'(t+h-2), ..., z'(t+h-d))$ [7], where the past d predictions are taken into consideration in order to perform the p^{th} prediction. In the iterative approach, the model is trained to perform one-step-ahead predictions at time $t+1$ iteratively h times using past predictions as inputs for the next prediction. This process do not need information at the desired horizon h however it accumulates the error from all the past predictions. Only one model is required to perform predictions.

The prediction of p-steps-ahead in the direct approach has the form $z'(t+h) = f(z(t), z(t-1), ..., z(t-(d-1)))$ [8]. In other words, the model is trained directly at the horizon h, and the predictions are performed directly into that horizon. In this approach we need information at horizon h in order to train the model and the error is not accumulated. To perform predictions in different values of h, other models have to be trained at this horizon.

The multiple approach takes into consideration the iterative and direct methods [6]. Multiple models are trained independently at different horizons, and perform iterative predictions until the desired horizon is achieved. Given a horizon h, k models (with $k < h$) will be trained independently to perform predictions at the horizons $\{t+1, t+2, ..., t+k\}$. This method does not need to train h models as needed in the direct approach and produces less accumulation of error than the iterative method.

3 Proposed Method

The proposed method incorporates the linear and nonlinear mappings of the time series through a composition of a ARIMA model and multiple SVR models.

The ARIMA model was first introduced by Box and Jenkins [9] and has been widely applied in time series forecasting. The predictions in the ARIMA model are a linear combination of past values and past residual errors, and has the form presented on eq. 1.

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

$$\theta_p(\mathbf{B})(1-\mathbf{B})^s z_t = \phi_q(\mathbf{B})\epsilon_t \tag{1}$$

where z_t is the actual value and ϵ_t is the residual error at time t, $\mathbf{B}$ is a backward shift operator ($\mathbf{B}z_t = z_{t-1}$ and $\Delta z_t = z_t - z_{t-1} = (1-\mathbf{B})z_t$), ϕ_i and θ_j are the coefficients, p and q are degrees for the autoregressive and moving average polynomial functions. The integration part is due to the differencing performed to the time series in order to obtain a stationary series. The number of differences is determined by the parameter s.

The SVR method was proposed by Vapnik [10] and it is based on the structured risk minimization (SRM) principle, aiming at minimizing an upper bound of the generalization error. Let $\{\mathbf{z}_i, y_i\}_{i=1}^{l}$ be a training set where $\mathbf{z} \in \Re^d$, $y_i \in \Re$ is the prediction output of z_i, d is the embbeding dimension of the time series and l is the number of training samples. The objective of SVR is to find the best function from a set of possible functions $\{f|f(\mathbf{z}) = \mathbf{w}^T\mathbf{z} + \mathbf{b}, w \in \Re^d, b \in \Re\}$, where w is the weight vector and $\mathbf{b}$ is a bias or threshold.

In order to find the best function f, it is necessary to minimize the regularized risk $\frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{l} L(y_i, f(\mathbf{z}_i))$, where $C > 0$ is a regularization factor, $\|.\|$ is a 2-norm and $L(.,.)$ is a loss function. In order to induce sparsity in SVR the ϵ-insensitive loss function presented on eq. 2, which creates an ϵ-tube allowing some predictions to fall within the boundaries of this ϵ-tube.

$$L(y, f(\mathbf{z})) = \begin{cases} 0, & |f(\mathbf{z}) - y| < \epsilon \\ |f(\mathbf{z}) - y| - \epsilon, & \text{otherwise} \end{cases} \tag{2}$$

The nonlinearity of SVRs is achieved by mapping the input space into a higher dimensionality space. This mapping is possible through the use of kernels, and the regression function takes the form $f(\mathbf{z}) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)k(\mathbf{z}_i, \mathbf{z}) + \mathbf{b}$, where α and α^* are lagrange multipliers and $k(\mathbf{z}_i, \mathbf{z})$ is a kernel function.

In this study, the gaussian kernel is adopted ($k(\mathbf{z}_i, \mathbf{z}_j) = exp(\frac{-\|\mathbf{z}_i - \mathbf{z}_j\|^2}{2\gamma^2})$), where γ is a parameter of the gaussian kernel.

The proposed model uses the ARIMA method as a prepossessing step in order to map linear patterns in the time series.

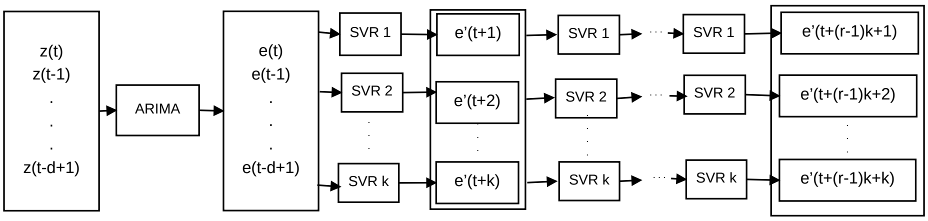

Fig. 1: Hybrid Arima Multiple SVR Method

Fig. 1 shows the framework of the ARIMA Multiple SVR method (ASVR-M). The r variable denotes the number of iterations of the method. The forecasting

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

procedure of this method performs less iterations than the iterative method, which runs as many iterations as the horizon size.

The residuals produced by the ARIMA $e_t = z_t - z'_t$, are used as an input for the Multiple SVR methods. In this step, k SVR models are trained independently, in order to perform iterative predictions. The final forecast can be achieved by adding the forecasts of the ARIMA and multiple SVR models $o'_t = z'_t + e'_t$.

4 Experiments

The experiments were performed in four benchmark datasets including Sunspot, Airline Passengers, IBM stock prices and Death. Our proposed approach is compared with other methods such as Iterative SVR (SVR-I), Direct SVR (SVR-D), Multiple SVR (SVR-M), ARIMA Iterative SVR (ASVR-I) and ARIMA Direct SVR (ASVR-D), implemented using the LIBSVM package [11] in a MATLAB 8.2 environment. The datasets were partitioned, using 80% for training and 20% for testing. First the p, q and s parameters of the ARIMA(p,s,q) model are defined with grid search 5-fold cross-validation in the training data with parameters ranging from $\{1, 2\}$, $\{0, 1, 2\}$ and $\{0, 1, 2\}$ respectively and the best ARIMA(p,s,q) model is selected based on the Aikake's Information Criterion (AIC) [12].

In all SVR models the parameter selection was performed by doing a 10-fold cross-validation grid search in the residuals produced by the selected ARIMA model with C, ϵ and γ within the range $\{10^{-1}, 10^{0}, 10^{1}, 10^{2}\}$, $\{10^{-4}, 10^{-3}, 10^{-2}\}$ and $\{2^{-5}, 2^{-4}, ..., 2^{2}\}$ respectively on the training set according to [6].

The annual Wolf's sunspot numbers dataset has 289 points and presents seasonal patterns. The monthly airline passengers dataset has 144 points and presents trend and seasonal patterns. The death dataset represents the number of deaths and serious injuries in UK road accidents each month and has 192 points. The daily IBM stock closing prices has 369 points and presents trend patterns. Fig 2 presents the autocorrelograms for all datasets, where the lags are set to unity. The sunspot, passengers, IBM and death datasets are available from [13].

The experiments for the sunspot, airlines passenger, death and IBM were performed considering an embedding dimension $d = 6$, prediction horizon $h = 10$ and number of SVR models in the multiple approaches $k = d$ based on [6]. All datasets had their values scaled to [0...1]. The experiments were executed for 10 iterations.

The performance measure used in this work as the mean squared error (MSE) described on eq. 3.

$$MSE = \frac{\sum_{i=1}^{N} (z_i - {z'}_i)^2}{N} \tag{3}$$

Table 1 contemplates the mean and standard deviations for the MSEs in all datasets. The results show that in most cases the hybrid ARIMA-SVR procedure

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7.
Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

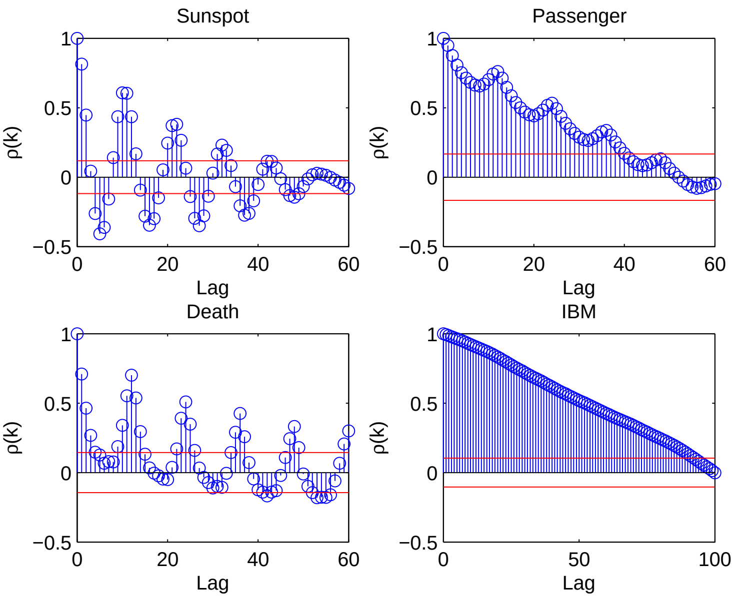

Fig. 2: Autocorrelograms of Datasets

Method	Sunspot	Passenger	Death	IBM
SVR-I	0.064(0.003)	0.136(0.028)	0.191(0.031)	0.004(0.000)
SVR-D	0.052(0.005)	0.020(0.004)	0.023(0.001)	0.018(0.004)
SVR-M	0.061(0.003)	0.153(0.030)	0.064(0.006)	0.010(0.003)
ASVR-I	0.030(0.004)	0.021(0.026)	0.125(0.003)	0.004(0.000)
ASVR-D	0.032(0.002)	0.003(0.000)	0.025(0.003)	0.007(0.000)
ASVR-M	0.045(0.005)	0.026(0.003)	0.060(0.006)	0.003(0.000)

Table 1: Mean square errors and Standard deviations for all datasets

improved the values achieved by the SVR strategies for long-term time series forecasting. An empirical analysis of the results achieved demonstrates that in the sunspot dataset the ASVR-I achieved the best results, followed by the ASVR-D. In the Passenger dataset, the ASVR-D outperformed the other hybrid approaches, however the ASVR-I presented a high variance in the results, while the results obtained by the ASVR-M were more stable. In the IBM dataset the ASVR-M achieve the best results, followed by the ASVR-I method. In the Death dataset the ASVR-D and SVR-D achieved the lowest errors, followed by the ASVR-M method.

A Wilcoxon signed-rank test was applied to perform a robust analysis with 95% of confidence. In the Sunspot dataset, the ASVR-I and ASVR-D achieved similar results and performed best. In the Passenger dataset the ASVR-D outperformed all techniques and the ASVR-I and ASVR-M performed similarly. In the Death dataset the ASVR-D and SVR-D achieved similar performance. In

ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014, i6doc.com publ., ISBN 978-287419095-7. Available from http://www.i6doc.com/fr/livre/?GCOI=28001100432440.

the IBM dataset the ASVR-M and SVR-I achieved similar results.

5 Conclusion

This paper realized a comparison between strategies (iterative, direct, multiple) to perform long-term time series forecasting using Support Vector Regression (SVR) models and proposed a hybrid model based on a Autoregressive Integrated Moving Average (ARIMA) model and on multiple SVRs.

The direct approaches (SVR-D, ASVR-D) achieved the best results in most datasets, which is expected since there is no iterative error accumulation. However the multiple approaches (SVR-M and ASVR-M) tends to reduce the accumulation error by running fewer iterations.

Under these settings the hybridization increased the performance of most methods, however the iterative method was the one with more increase of performance, in most cases achieving similar results to the ASVR-M.

For future works, we intend to analyze the performance of a hybridization between ARIMA and an ensemble of SVRs to perform prediction using the iterative, direct and multiple strategies.

References

[1] Spyros Makridakis, Steven C Wheelwright, and Rob J Hyndman. *Forecasting methods and applications*. Wiley. com, 2008.

[2] K-R Müller, Alex J Smola, Gunnar Rätsch, Bernhard Schölkopf, Jens Kohlmorgen, and Vladimir Vapnik. Predicting time series with support vector machines. In *Artificial Neural Networks—ICANN'97*, pages 999–1004. Springer, 1997.

[3] Guoqiang Zhang, B Eddy Patuwo, and Michael Y Hu. Forecasting with artificial neural networks:: The state of the art. *International journal of forecasting*, 14(1):35–62, 1998.

[4] John M Bates and Clive WJ Granger. The combination of forecasts. *OR*, pages 451–468, 1969.

[5] Ping-Feng Pai and Chih-Sheng Lin. A hybrid arima and support vector machines model in stock price forecasting. *Omega*, 33(6):497–505, 2005.

[6] Li Zhang, Wei-Da Zhou, Pei-Chann Chang, Ji-Wen Yang, and Fan-Zhang Li. Iterated time series prediction with multiple support vector regression models. *Neurocomputing*, 2012.

[7] Abdelhamid Bouchachia and Saliha Bouchachia. Ensemble learning for time series prediction. In *Proceedings of the International Workshop on Nonlinear Dynamic Systems and Synchronization*, 2008.

[8] Souhaib Ben Taieb, Antti Sorjamaa, and Gianluca Bontempi. Multiple-output modeling for multi-step-ahead time series forecasting. *Neurocomputing*, 73(10):1950–1957, 2010.

[9] George EP Box, Gwilym M Jenkins, and Gregory C Reinsel. *Time series analysis: forecasting and control*. Wiley. com, 2013.

[10] Vladimir Vapnik. *The nature of statistical learning theory*. springer, 2000.

[11] Chih-Chung Chang and Chih-Jen Lin. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27, 2011.

[12] Hamparsum Bozdogan. Model selection and akaike's information criterion (aic): The general theory and its analytical extensions. *Psychometrika*, 52(3):345–370, 1987.

[13] R.J. Hyndman. Time series data library, 2010.