

Predicting Time Series Using Integration of Moving Average and Support Vector Regression

Chihli Hung, Chih-Neng Hung, and Szu-Yin Lin

Abstract—Time series prediction is one of the major tasks in the field of data mining. The approaches of time series prediction can be divided into statistical techniques and computational intelligence techniques. Most researchers use one specific approach and compare the performance with other approaches. This paper proposes a novel hybrid approach, which integrates traditional moving average models with support vector regression for the prediction of ATM withdrawals in England. The use of moving average modeling is not only for the purpose of smoothing but also for time series prediction. We treat a weekly median moving average as the benchmark. Based on experimental results, our proposed approach consistently outperforms the benchmark.

Index Terms—Support vector regression, cash withdrawal analysis, time series prediction, data mining.

I. INTRODUCTION

This paper proposes a novel approach, which integrates traditional moving average models with support vector regression [1] for the prediction of ATM withdrawals in England. Time series prediction is one of the major tasks in the field of data mining [2]-[4]. Mining useful knowledge from large amounts of data has attracted a great deal of interest. A large portion of data in the real world is presented as time series, in which data items are time dependent and obtained over repeated time intervals [5]. Time series prediction is beneficial for many applications, such as predictions for financial returns, stock market prices, retail transactions, river floods, electrical consumption, and so on. There are several time series prediction competitions, which not only group researchers from different fields together, but also provide valid and reliable evaluation platforms for the empirical performance of different approaches [6], [7].

Generally speaking, the approaches to time series prediction include statistical techniques and computational intelligence techniques. Most researchers use one specific approach and compare the performance with other approaches. This paper combines the computational intelligence technique, i.e. the support vector regression (SVR) [8], with statistical moving average models, for eleven time series from the NN5 time series prediction competition

dataset¹.

For the prediction of these time series, we use two different time-based moving average approaches, i.e. weekly moving average and daily moving average, in order to reduce the amount of variation in the data. We use the median of different weekly and daily moving averages instead of the traditional moving average. As it is not easy to predict the target variable, Y , directly, we treat the difference between the actual value and the median of moving average as the prediction target for SVR.

The remainder of this paper is organized as follows. The related work, focusing on computational intelligence approaches, is described in Section II. Section III introduces our methodologies, which include support vector regression and our proposed prediction approach. The experimental design, including the dataset and the evaluation, is presented in Section IV. The experimental results are given in Section V and finally, the conclusion is provided in Section VI.

II. RELATED WORK

A vast amount of research has been devoted to time series prediction. The approaches to time series prediction can be broadly divided into two groups. The first group uses statistical techniques and the second group uses computational intelligence techniques. The major methods in the statistical group contain moving average (MA), linear regression, multiple regression, logistic regression, cluster analysis, multiple group discriminant analysis, autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and so on [9].

Even though such approaches have performed well for several time series tasks, the fast development of computer hardware and software techniques provides an efficient simulation environment recently. Thus, researchers in the computational intelligence field use many techniques including the multilayer perceptron neural network (MLP) [10], recurrent neural network (RNN) [11], self-organizing map (SOM) [12], support vector machine (SVM) [13], genetic algorithm (GA) [14], fuzzy logic [15], etc. for time series prediction.

For example, Hill *et al.* [16] compared six statistical methods with the artificial neural network (ANN) technique on 111 time series. Based on the experimental results, the ANN outperforms the statistical methods. Xie and Hu [17] predicted Shanghai housing price index using ARIMA, ANN and SVM, and found that ANN and SVM are superior to ARIMA. Many ANN models have been proven useful for

Manuscript received April 10, 2014; revised June 18, 2014. This work was supported in part by the Ministry of Science and Technology, Taiwan under Grant of NSC 102-2410-H-033-033-MY2.

Chihli Hung and Szu-Yin Lin are with the Department of Information Management, Chung Yuan Christian University, Chungli 32023, Taiwan, ROC (e-mail: chihli@cycu.edu.tw, stan@cycu.edu.tw).

Chih-Neng Hung is with the Department of Information and Technology Education, Ministry of Education, Taipei, Taiwan, ROC (e-mail: hcn0619@gmail.com).

¹ <http://www.neural-forecasting-competition.com/NN5>

time series, e.g. [18], [19]. However, some researchers have reached the opposite conclusion. After extensive analysis of the forecasting accuracy of ANN and statistical forecasters, Foster *et al.* [20] found that linear regression and the simple average of the exponential smoothing method are superior to their ANN counterpart. Some work was based on a hybrid approach, which combines the ANN and statistical methods, such as ARIMA [21], or several techniques in computational intelligence [22], [23] in order to achieve better performance.

In this paper, we focus on using computational intelligence techniques and integrate moving average methods with support vector regression for eleven time series in the NN5 time series prediction task.

III. METHODOLOGIES

A. Support Vector Regression

The support vector machine (SVM), based on the computational approach termed the principle of structural risk minimization, proposed by Vapnik [1], was originally mainly used for binary classification tasks. This principle is related to the statistical learning theory in order to develop a novel type of neural network. The main idea of SVM is to find the hyper plane with the biggest margin to divide data into two groups. This hyper plane is called the optimal separating hyper plane. For linearly indivisible data, SVM should map them onto a higher dimensional feature space by the transformation of its kernel function. In other words, the optimal separating hyper plane can be found by the transformation from linearly indivisible data in a lower dimensional feature space to linearly divisible data in a higher dimensional feature space.

Based on the same concept of SVM, Vapnik *et al.* [8] proposed another version of SVM for regression, which is called support vector regression (SVR). SVR has been proven valuable for modeling and predicting time series in several practical applications [24]. However, the quadratic programming solver is the only known training algorithm for SVM for a long time. This solver causes SVM to suffer from consuming long processing time while dealing with a large dataset. Several training algorithms have recently been proposed for SVM. In this paper, we use a sequential minimal optimization (SMO) algorithm [25]. More theoretical analysis can be found in the technical report of Smola and Schölkopf [26].

B. The Prediction Approach

It is necessary to model time series by analyzing the mechanism that produces the time series before prediction. The data used in the NN5 time series prediction task contains about two years of daily cash withdrawals at various automatic teller machines (ATMs) which are located throughout England. As such withdrawals from ATMs are mainly related to different day of the week, and the weekly fluctuation rate, it is necessary to deal with trends and seasonal variations. The withdrawals are also influenced by special events or structural system breaks, which may cause the time series to include outliers, zeros and missing values. That is, these time series also contain trends, cyclic variations, seasonal variations and random movements. The time series

prediction target variable, Y , can be illustrated as $Y = T \times C \times S \times I$ or their sum, where T , C , S and I indicates trend, cyclic seasonal and random movements, respectively.

Even though there are several approaches which are able to model and predict time series usefully, some of them may not be easy to build. For example, Rojas *et al.* [21] argued that obtaining the structure of the ARMA model is a problem itself, and it seems difficult to develop an automatic system without the direct intervention of a human expert in the Box-Jenkins methodology. Conversely, a simple approach such as moving average tends to reduce the amount of variation present in time series and may have a better performance than other very complex algorithms [27].

Moving average (MA) is an approach for predicting the target variable, Y_t , by its k following sequence of arithmetic mean, MA_k as in (1). Thus moving average can be treated as a smoothing approach for time series as it is able to eliminate unwanted fluctuations of time series. However, this smoothing feature also makes moving average react to an event with some time lag. For example, in Fig. 1, the solid line illustrates the actual value and the dotted line is moving average of three previous actual observations. The difference between actual value and its moving average at the event A is made due to this smoothing feature of MA. Therefore, the prediction target made by moving average is a lagging indicator, which responds to the information behind the target.

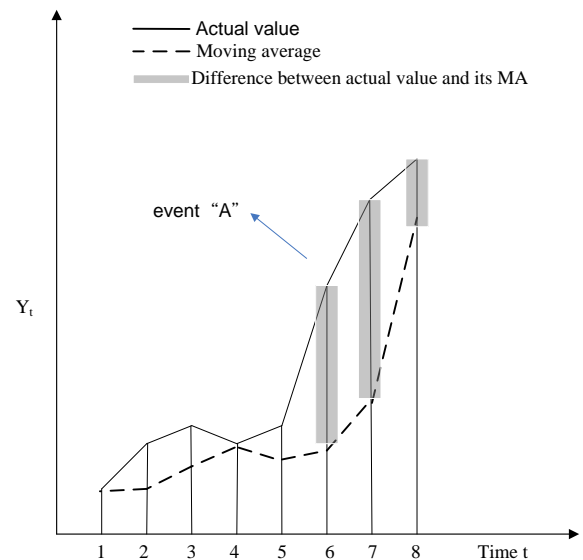


Fig. 1. The difference between actual value and moving average at the event "A".

$$Y_t \approx MA_k = \frac{1}{K} (Y_{t-1} + Y_{t-2} + \dots + Y_{t-K+1}) \quad (1)$$

The NN5 dataset generated from ATM cash withdrawals, which are mainly related to different day of the week. In other words, the relationship of the same day of the week in different weeks is significant. Thus we consider moving average on a weekly base. Moreover, the median of several weekly moving average values is more stable than a single moving average for the prediction of the difference between the actual value and the MA median. According to our experiments, the median of two week MA, three week MA and four week MA is quite useful.

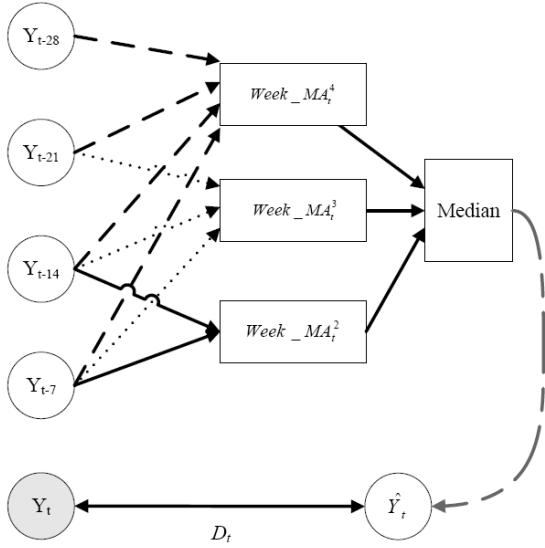


Fig. 2. The conceptual diagram of our proposed approach. D_t indicates the difference between median of MAs and the actual value at time t .

The equations of these MAs are shown as (2)-(4), respectively. In this paper we use the support vector regression (SVR), based on the sequential minimal optimization (SMO) algorithm [25] with a polynomial kernel function and an epsilon of 0.001, to model and predict time series by evaluating the difference between actual value and its median weekly moving average. This difference is named as D_t . We can then get the prediction target, Y_t , as in (5) and the conceptual diagram is shown in Fig. 2.

$$Week_MA_2 = \frac{1}{2}(Y_{t-7} + Y_{t-14}) \quad (2)$$

$$Week_MA_3 = \frac{1}{3}(Y_{t-7} + Y_{t-14} + Y_{t-21}) \quad (3)$$

$$Week_MA_4 = \frac{1}{4}(Y_{t-7} + Y_{t-14} + Y_{t-21} + Y_{t-28}) \quad (4)$$

$$Y_t \approx D_t + median(Week_MA_2, Week_MA_3, Week_MA_4) \quad (5)$$

The target output variable used by SVR is the difference between actual value and its median weekly moving average, i.e. D_t . By analyzing the features of the NN5 dataset, cash withdrawals are influenced by three groups of factors. The first group contains previous actual withdrawals, median of weekly MA and median of daily MA. Input factors in the second group are fluctuation rates. The purpose of factors in the third group is to track deviation of historic observations. We name this approach as SVR-median.

On the other hand, the traditional MA approach is not only for the purpose of smoothing but also for time series prediction. We directly use weekly median MA to predict time series without using any computational intelligence algorithm. This approach is our benchmark model and is named as MA-median.

IV. EXPERIMENTAL DESIGN

A. Dataset

The 2008 NN5 time series competition for artificial neural

networks and computational intelligence extends the earlier 2007 NN3 time series competition. Both competitions provide two datasets: a complete 111 time series dataset and its reduced subset, which contains 11 time series. In this paper, we evaluate our approach based on the reduced subset as these two datasets contain similar complexity. The NN5 reduced dataset includes about two years of daily cash withdrawals at various automatic teller machines (ATMs) located in different parts of England. Each time series is formed by cash withdrawals from an ATM. More specifically, the daily cash withdrawals of each time series start from 18 March 1996 to 22 March 1998, and contain about 735 known observations as the dataset. However, since the cash withdrawals are influenced by special events and some structural system breaks, each individual time series contains the different numbers of missing values, zeros and outliers. We use the last 97 observations from the dataset as the test set and the rest as the training set so as to evaluate the performance of our proposed approach. The missing values are replaced by their associated median weekly MA. When their associated median cannot be found, then their associated moving average of the previous two days is used instead.

B. Evaluation

As the time series prediction target variable, Y , is numeric, it is more difficult to evaluate its performance than other nominal prediction or classification target variables used in a classification task. That is, the performance of time series prediction cannot be evaluated solely on the criterion of accuracy. We should evaluate the similarity between actual value and its prediction target value. On the other hand, as different individual time series may contain a different number of observations, it is necessary to consider the different scale between individual time series. We follow the evaluation criterion defined in the NN5 time series competition, which uses the symmetric mean absolute percent error (SMAPE) [28] as in (6).

$$SMAPE = \frac{1}{n} \sum_{t=1}^n \frac{|X_t - Y_t|}{\frac{1}{2}(X_t + Y_t)} \times 100 \quad (6)$$

V. EXPERIMENTAL RESULTS

The NN5 reduced dataset consists of eleven time series. Each series consists of 735 known observations. The last 97 observations from the original training set are used as our test set. We use the weekly moving average approach, i.e. MA-median as the benchmark for our proposed approach, i.e. SVR-median. Table I shows the performance of these two approaches evaluated by SMAPE. Our proposed approach performs consistently better than the benchmark on both training set and test set for all time series.

A t -test is used to test whether or not a difference between the proposed SVR-median and the MA-median prediction models achieves statistical significance. According to the t -test result, this proposed SVR-median prediction model is significantly better than the MA-median one due to 0.000 of the p -value for both training and test sets.

TABLE I: COMPARISON BETWEEN MA-MEDIAN AND SVR-MEDIAN FOR 11 TIME SERIES

Time Series	Training Set		Test Set	
	MA-median	SVR-median	MA-median	SVR-median
(101)	22.9%	16.4%	28.5%	22.0%
(102)	23.2%	18.0%	33.4%	25.1%
(103)	23.7%	17.7%	26.4%	22.1%
(104)	26.3%	21.9%	25.2%	20.5%
(105)	23.2%	18.8%	31.2%	26.1%
(106)	24.4%	19.4%	27.2%	24.1%
(107)	22.0%	16.7%	34.3%	27.8%
(108)	21.8%	16.5%	31.7%	26.8%
(109)	16.6%	13.5%	26.1%	20.7%
(110)	27.6%	19.1%	30.0%	24.2%
(111)	24.0%	18.8%	26.4%	21.7%
Mean SMAPE	23.3%	17.9%	29.1%	23.7%

VI. CONCLUSION

This paper has proposed a novel hybrid approach, by the integration of statistical moving average models and support vector regression in the computational intelligence field. These time series are generated from daily cash withdrawals of different automatic teller machines, which are influenced by different day of the week, special events and occasional system structural breaks. This paper uses the sequential minimal optimization algorithm for SVR to predict the difference between the actual withdrawal and its median of various weekly moving averages. We consider a weekly median moving average as the benchmark for this research. Experimental results show that our proposed approach performs better than the benchmark for all time series evaluated by the criterion of the symmetric mean absolute percent error.

REFERENCES

- [1] V. Vapnik, "Principles of risk minimization for learning theory," *Proceedings of Neural Information Processing Systems*, pp. 831-838, 1991.
- [2] A. K. Palit and D. Popovic, *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*, London: Springer-Verlag London Limited, 2005.
- [3] T.-C. Fu, "A review on time series data mining," *Engineering Applications of Artificial Intelligence*, vol. 24, pp. 164-181, 2011.
- [4] P. Esling, and C. Agon, "Time-series data mining," *Computing Surveys*, vol. 45, no. 1, paper 12, 2012.
- [5] E. Keogh, "A decade of progress in indexing and mining large time series databases," in *Proc. the 32nd International Conference on Very Large Data Bases*, 2006, pp. 1268-1268.
- [6] J. McNames, J. Suykens, and J. Vandewalle, "Winning entry of the K.U. Leuven time-series prediction competition," *International Journal of Bifurcation Chaos*, vol. 9, no. 8, pp. 1485-1500, 1999.
- [7] A. Lendasse, E. Oja, O. Simula, and M. Verleysen, "Time series prediction competition: The CATS benchmark," *Neurocomputing*, vol. 70, no. 13-15, pp. 2325-2329, 2007.
- [8] V. Vapnik, S. E. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation, and signal processing," in *Advances in Neural Information Processing Systems*, vol. 9, M. C. Mozer et al., Eds. Cambridge: MIT Press, 1997, pp. 281-287.
- [9] B. L. Bowerman, R. T. O'Connell, and A. B. Koehler, *Forecasting, Time Series, and Regression: An Applied Approach*, Belmont, CA: Thomson Learning, 2005.
- [10] R. Yadav, P. Kalra, and J. John, "Time series prediction with single multiplicative neuron model," *Applied Soft Computing*, vol. 7, no. 4, pp. 1157-1163, 2007.
- [11] R. C. Staudemeyer and C. W. Omlin, "Evaluating performance of long short-term memory recurrent neural networks on intrusion detection data," in *Proc. the South African Institute for Computer Scientists and Information Technologists Conference (SAICSIT)*, pp. 218-224, 2013.
- [12] G. Barreto, "Time series prediction with the self-organizing map: A review," *Studies in Computational Intelligence*, vol. 77, pp.135-158, 2007.
- [13] A. Sorjamaa, J. Hao, N. Reyhani, Y. Ji, and A. Lendasse, "Methodology for long-term prediction of time series," *Neurocomputing*, vol. 70, no. 16-18, pp. 2861-2869, 2008.
- [14] M. M. M. Fuad, "Differential evolution versus genetic algorithms: Towards symbolic aggregate approximation of non-normalized time series," in *Proc. the 16th International Database Engineering & Applications Symposium*, pp. 205-210, 2012.
- [15] H. G. Seedig, R. Grothmann, and T. A. Runkler, "Forecasting of clustered time series with recurrent neural networks and a fuzzy clustering scheme," in *Proc. the 2009 International Joint Conference on Neural Networks (IJCNN)*, 2009.
- [16] T. Hill, M. O'Connor, and W. Remus, "Neural network models for time series models forecasts," *Management Sciences*, vol. 42, no. 7, pp. 1082-1092, 1996.
- [17] X. Xie and G. Hu, "A comparison of Shanghai housing price index forecasting," in *Proc. the Third International Conference on Natural Computation*, pp. 221-225, 2007.
- [18] T. H.-K. Yu and K.-H. Huarng, "A bivariate fuzzy time series model to forecast the TAIEX," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2945-2952, 2008.
- [19] G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," *Information Sciences*, vol. 177, pp. 5329-5346, 2007.
- [20] W. R. Forster, F. Collopy, and L. H. Ungar, "Neural network forecasting of short, noisy time series," *Computers and Chemical Engineering*, vol. 16, no. 2, pp. 293-297, 1992.
- [21] I. Rojas, O. Valenzuela, F. Rojas, A. Guillen, L. J. Herrera, H. Pomares, L. Marquez, and M. Pasadas, "Soft-computing techniques and ARMA model for time series prediction," *Neurocomputing*, vol. 71, pp. 519-537, 2008.
- [22] H.-J. Kim and K.-S. Shin, "A hybrid approach based on neural networks and genetic algorithms for detecting temporal patterns in stock markets," *Applied Soft Computing*, vol. 7, pp. 569-576, 2007.
- [23] K. W. Lau and Q. H. Wu, "Local prediction of non-linear time series using support vector regression," *Pattern Recognition*, vol. 41, pp. 1556-1564, 2008.
- [24] L. Cao, "Support vector machines experts for time series forecasting," *Neurocomputing*, vol. 51, pp. 321-339, 2003.
- [25] J. Platt, "Fast training of support vector machines using sequential minimal optimization," in *Advances in Kernel Methods-Support Vector Learning*, B. Schoelkopf et al. Eds., Cambridge: MIT Press, pp. 185-208, 1999.
- [26] A. Smola, and B. Schölkopf, "A tutorial on support vector regression," Technical Report, Royal Holloway College, NeuroCOLT Technical Report (NC-TR-98-030), University of London, UK, 1998.
- [27] R. T. Clemen, "Combining forecasts: A review and annotated bibliography," *International Journal of Forecasting*, vol. 5, pp. 559-583, 1989.
- [28] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, vol. 8, pp. 69-80, 1992.



Chihli Hung received a PhD degree at the School of Computing and Technology from the University of Sunderland, UK in 2004. He is an associate professor at the Department of Information Management, Chung Yuan Christian University, Taiwan, R.O.C. His current research interests include text mining, intelligence systems, computational intelligence, and data mining.



Chih-Neng Hung received a PhD degree at the Institute of Management Science from Tamkang University, Taiwan in 2013. His current research interests include data mining and information security governance.



Szu-Yin Lin received his Ph.D. degree in information management from National Chiao Tung University, Taiwan, 2012. He is an assistant professor in the Department of Information Management at Chung Yuan Christian University, Taiwan. His expertise is on service-oriented architecture and computing, intelligent dynamic data analysis, multi-agent systems, and internet technology.