

RECURRENT NEURAL NETWORKS AND NONLINEAR PREDICTION IN SUPPORT VECTOR MACHINES

Dr. Jennifer S. Raj,

Professor,
Department of Electronics & Communication Engineering,
Gnanamani College of Technology, Namakal, India.
Email: jennifer.raj@gmail.com

J. Vijitha Ananthi,

Research Analyst, EJesus, India.
Email: vijithaananthi@gmail.com

Abstract: The nonlinear regression estimation issues are solved by successful application of a novel neural network technique termed as support vector machines (SVMs). Evaluation of recurrent neural networks (RNNs) can assist in pattern recognition of several real-time applications and reduce the pattern mismatch. This paper provides a robust prediction model for multiple applications. Traditionally, back-propagation algorithms were used for training RNN. This paper predict system reliability by applying SVM learning algorithm to RNN. Comparison of the proposed model is done with the existing systems for analysis of prediction performance. These results indicate that the performance of proposed system exceeds that of the existing ones.

Keywords: Support Vector Machine, Recurrent Neural Networks, pattern recognition, back-propagation algorithm

1. INTRODUCTION

Over the years, researchers are working towards prediction of nonlinear data for applications in stock market analysis, e-commerce, and traffic speed prediction and so on. RNNs have the ability to learn temporal sequence by processing arbitrary input sequences using internal memory units [2]. Traditional RNN models imposes two major issues namely: (1) it is essential to predetermine the time step count ahead for most RNN; (2) Traditional RNNs are unsuccessful in capturing the long temporal reliance for input sequence. It has been proven to be challenging to train RNN with 5-10 time lags due to exploding and vanishing gradient issues.

Propagation through time and dynamic back-propagation techniques are used for training recurrent neural networks classically. In order to determine the suitable architecture of neural network to be used, a sensitivity analysis can be performed. The traditional architectures in use for predictive analysis includes Autoregressive Integrated Moving Average (ARIMA) models and Multi-Layer Perceptron (MLP). Using dynamic back-propagation for training RNN is difficult as it has to trail paths of unruly systems. Large scale bound nonlinear

optimization techniques are to be used for large data sets. In regard to nonlinear modelling and time-series prediction, new perspectives are provided by recurrent SVMs.

Forecasting the failure time of hardware components and the reliability can be done by Infinite Impulse Response Locally Recurrent Neural Network (IIR-LRNN) modelling technique. The engine failure process and predictive performance is studied in several learning models like ARIMA, RBF, and MLP and so on. IIR-LRNN offers better performance compared to ARIMA and RBF. SVM can also be used for financial time series prediction based on structural risk minimisation principle based empirical error and risk function with regularized term. The structural risk minimisation principle reduces the generalization error's upper bound whereas traditional methods reduces misclassification errors.

SVM provides global optimum solution whereas other neural network models provides local optimal solutions. Both regression as well as classification can be performed by Support Vector Machines method. They are extremely useful in mathematical relations where the inputs of the underlying process maps to the unknown outputs as it is expensive to obtain these relations due to several independent factors that influence the output. Artificial intelligence, neural networks and fuzzy models are used for the processing of big data and improvement of learning abilities. These models consume more amount of time for training of models.

While considering multiple regression analysis, weather and total load based relationship is influenced and load is calculated. Weighted least square is estimated and computed with regression coefficients exponentially and equally. This process uses historical data in predefined amounts. When SVM is used alone without careful selection of parameters, the results yielded are poor and inefficient. Hence this paper proposes the application of optimization algorithms for improving the prediction quality in RNNs.

2. EXISTING LITERATURE

Trevor et al. [7] compared few classification algorithms including SVM, Particle Swarm Optimization, Artificial Neural Network and Naïve Bayes Classifier for minimising investor risk and maximizing capital. The paper provides an improved prediction method for robust stock forecasting that offers improved precision with a combination of particle swarm optimization and tuning support vector machine factors. In stock behaviour prediction applications, optimization of SVM parameters with properly tuned particle swarms can be done for high prediction accuracy. This system is compared with the ANN and naïve Bayes classifier models for prediction of accuracy. This system is of great use for spontaneous interday merchants.

Márcio et al. [8] performed a study on SVM regression compared to advance learning approaches like the traditional Multilayer Perceptron model, Radial Basis Function, Infinite Impulse Response Locally Recurrent Neural Networks and the Box-Jenkins autoregressive-integrated-moving average. The study shows that SVM offers better efficiency in comparison with other methodologies. This research assists in handling time series data that are smooth and wavy and offers noteworthy performance in this regard.

Suykens et al [9] uses SVM in combination with RNN. The paper considers a cost function based least square version that offers equality constraints to the RNN. Recurrent least square SVM (LS-SVMs) offer reliable outputs even when the training data sets are relatively small. In case of large data sets, more controlled nonlinear optimization method is to be used. Sayan et al [10] performed implementation and testing of chaotic time series based data based that compares the performance of local polynomial techniques, polynomial and rational approximation, Neural Networks, and Radial Basis Functions based approximation methodologies.

Jie Xu et al [11] studied the influence of dependent sampling techniques on incremental support vector machines (ISVM) learning performance. The paper performs a comparative analysis between Markov resampling (MR-ISVM) and randomly independent sampling (RIS-ISVM). MR-ISVM offers smaller misclassification rates, sparser classifiers and lesser training time. Xianxian Hou et al [12] used Grey Relational Analysis, Support Vector Regression, and fruit fly optimization algorithm for selection of technical indicators, making predictions and optimization respectively. It offers better stock forecasting accuracy, quicker convergence, more feasibility and improved efficiency.

SVM is broadly categorised into Support Vector Regression (SVR) and Support Vector Classification (SVC). Guillermo et al [14] proposed a time adaptive SVR method for data analysis. The work shows better precision in numerous non-stationary data analysis aspects such as hidden forcing profile reconstruction, input relevance estimation, modelling and prediction. Several ANN approaches has been outperformed by SVM in various researches. Current trends in research focus on applying these methodologies in big data analytics and energy analytics for monitoring large quantities of data. Comparison of performance of deep learning algorithm is also investigated in numerous researches.

Maciej et al [15] worked on developing a neutralisation model and a Model Predictive Control (MPC) nonlinear algorithm that is simple and precise. The use of pruned LS-SVM Wiener model for this purpose improves the computational complexity but offers improved accuracy. Duyu Tang et al [17] provides a neural model that indicates improved performance in comparison with traditional algorithm and a RNN that outperforms certain other neural network models.

3. PROPOSED WORK

On application of optimization algorithms such as Gravitational Search Algorithm (GSA), Artificial Bee Colony Algorithm (ABC) and Differential Evolution algorithm (DE), a hybrid model is created. GSA allows simulation of mass interactions in an iterative fashion. The evaluation of fitness function is done for analysing the agent. Factors like initial gravitational constant (G_0), maximum iteration parameters, β constant, and population size (N) and so on affect the performance of the algorithm.

In ABC, the solution to the optimization problem is offered by the food source position corresponding to the solution's fitness. DE was initially used for solving fitting problems of Chebyshev polynomials. It is comparatively easy to implement and has fast convergence rate. It also requires lesser parameters. This algorithm basically comprises of four steps namely- Parameter initialization, starting of evolution, preliminary calculations and offspring generation.

Figure 1 provides a schematic of RNN architecture with single hidden layer and time delay z^{-1} . There are 30 neurons hidden in every individual hidden layer.

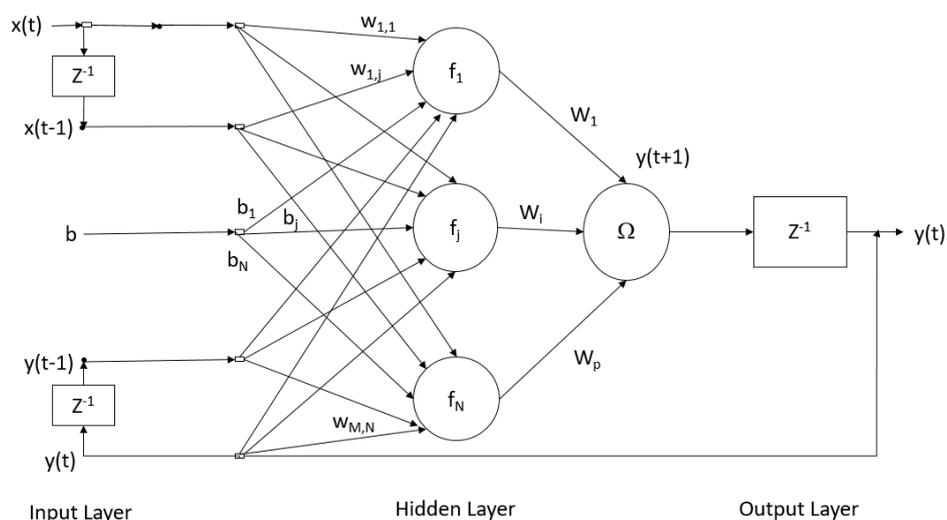


Figure 1: Architecture of RNN with single hidden layer (own illustration based on [5])

$$f_j(x_i) = \tanh(\sum_{i=1}^M w_{i,j} x_i + b_i) \text{ ----- (1)}$$

$$\Omega(f_j) = \sum_{j=1}^N W_j f_j + B \text{ ----- (2)}$$

Where f represents the activation function, u is the input, B is the bias. The input neuron and hidden neuron j is connected with weights w . W is the output neuron and b is the bias of the hidden neuron. We use the statistical criteria of Mean Absolute Percentage Error (MAPE), Mean Average Error (MAE) and Mean Squared Error (MSE) in order to assess the performance of the models and to measure its accuracy. MAPE analyses the accuracy of a forecast system. MAE represents the average absolute horizontal or vertical gap amongst the points in the scatter graph. MSE is also called as Mean Square Deviation (MSD) and is used for estimation of average error squares i.e. square of the difference between actual and estimated values.

$$MAPE = \frac{1}{n} \sum_{k=1}^n \left| \frac{t_k - \hat{t}_k}{t_k} \right| \times 100 \text{-----} (3)$$

$$MAE = \frac{1}{n} \sum_{k=1}^n |t_k - \hat{t}_k| \text{-----} (4)$$

$$MSE = \frac{1}{n} \sum_{k=1}^n (t_k - \hat{t}_k)^2 \text{-----} (5)$$

t_k and \hat{t}_k are the actual and predicted values of the k^{th} observation. The influence of MSE is stronger when compared to that of MAE. N is the total count of observations.

4. RESULT

On application of the optimization algorithms, the SVR parameters obtain optimal values as represented in the table below. C is a regularization parameter, ϵ represents the loss function and σ is the Radial Basis Kernel Function (RBF) kernel parameter. These values are analysed and the parameter that offers maximum accuracy is for most appropriate prediction and optimization of the algorithm. In this experiment, 15 samples are used for analysis of these learning parameters.

Table 1: Optimization algorithm based optimal values of learning parameters

| Optimal values | C | ϵ | σ |
|----------------|-----|------------|----------|
| | | | |

| | | | |
|-----------|-----|------|-------|
| GSA model | 2.3 | 2289 | 0.211 |
| ABC model | 2.4 | 2506 | 0.214 |
| DE model | 2.7 | 2606 | 0.215 |

The optimization of the algorithm is done by selecting the parameters and building the SVM model. Further, the fitness calculation is done and optimal parameters are obtained based on the fitness values. The optimal model of SVM is then built and its performance is analysed. There is a feedback provided from output to the input in RNN models. Hence, there is a higher error in time step and causes overfit of the network. When the error measurement and computation efforts are taken into consideration, the overall performance of the system is considerable when a single hidden layer is used.

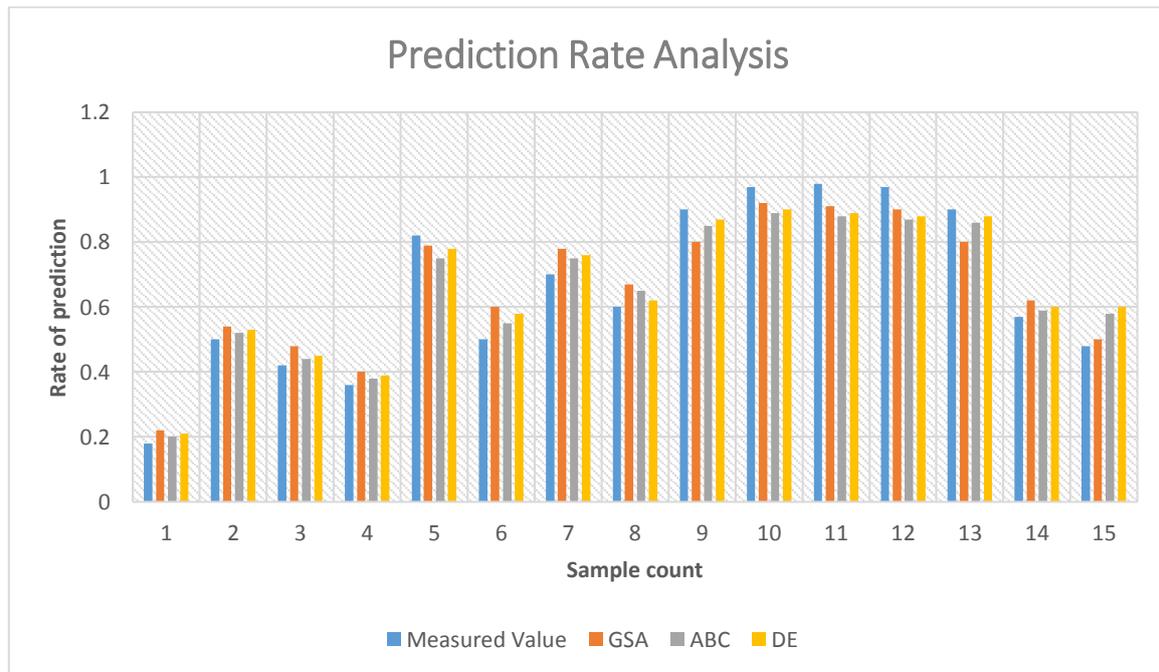


Figure 2: Prediction Rate Analysis graph of the 3 algorithms and measured values

5. CONCLUSION AND FUTURE SCOPE

In order to solve the nonlinear regression estimation issue, the paper proposes applying SVM optimization to RNNs. In this regard, this paper presents the application of three dynamic optimization algorithm namely Gravitational Search Algorithm (GSA), Artificial Bee Colony Algorithm (ABC) and Differential Evolution algorithm (DE) to create a hybrid prediction model. This model proves to improve the accuracy and speed in identification of optimal values of the parameters of SVM. Future work would depend on comparing more optimization algorithms and extracting their best features for developing a further more efficient system. Also, focus is to be done on application specific algorithm development.

References

- [1] Pai, Ping-Feng, and Chih-Sheng Lin. "A hybrid ARIMA and support vector machines model in stock price forecasting." *Omega* 33, no. 6 (2005): 497-505.
- [2] Ma, Xiaolei, Zhimin Tao, Yin Hai Wang, Haiyang Yu, and Yunpeng Wang. "Long short-term memory neural network for traffic speed prediction using remote microwave sensor data." *Transportation Research Part C: Emerging Technologies* 54 (2015): 187-197.
- [3] Kim, Kyoung-jae. "Financial time series forecasting using support vector machines." *Neurocomputing* 55, no. 1-2 (2003): 307-319.
- [4] Mohandes, Mohamed. "Support vector machines for short-term electrical load forecasting." *International Journal of Energy Research* 26, no. 4 (2002): 335-345.
- [5] Koschwitz, D., J. Frisch, and C. Van Treeck. "Data-driven heating and cooling load predictions for non-residential buildings based on support vector machine regression and NARX Recurrent Neural Network: A comparative study on district scale." *Energy* 165 (2018): 134-142.
- [6] Chniti, Ghassen, Houda Bakir, and Hédi Zaher. "E-commerce time series forecasting using LSTM neural network and support vector regression." In *Proceedings of the International Conference on Big Data and Internet of Thing*, pp. 80-84. ACM, 2017.
- [7] Sands, Trevor M., Deep Tayal, Matthew E. Morris, and Sildomar T. Monteiro. "Robust stock value prediction using support vector machines with particle swarm optimization." In *2015 IEEE Congress on Evolutionary Computation (CEC)*, pp. 3327-3331. IEEE, 2015.
- [8] Das Chagas Moura, Márcio, Enrico Zio, Isis Didier Lins, and Enrique Droguett. "Failure and reliability prediction by support vector machines regression of time series data." *Reliability Engineering & System Safety* 96, no. 11 (2011): 1527-1534.
- [9] Suykens, Johan AK, and Joos Vandewalle. "Recurrent least squares support vector machines." *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications* 47, no. 7 (2000): 1109-1114.

- [10] Mukherjee, Sayan, Edgar Osuna, and Federico Girosi. "Nonlinear prediction of chaotic time series using support vector machines." In *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, pp. 511-520. IEEE, 1997.
- [11] Xu, Jie, Chen Xu, Bin Zou, Yuan Yan Tang, Jiangtao Peng, and Xinge You. "New incremental learning algorithm with support vector machines." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* (2018).
- [12] Hou, Xianxian, Shaohan Zhu, Li Xia, and Gang Wu. "Stock price prediction based on Grey Relational Analysis and support vector regression." In *2018 Chinese Control and Decision Conference (CCDC)*, pp. 2509-2513. IEEE, 2018.
- [13] Peng, Yifan, Anthony Rios, Ramakanth Kavuluru, and Zhiyong Lu. "Chemical-protein relation extraction with ensembles of SVM, CNN, and RNN models." *arXiv preprint arXiv:1802.01255* (2018).
- [14] Grinblat, Guillermo L., Lucas C. Uzal, Pablo F. Verdes, and Pablo M. Granitto. "Nonstationary regression with support vector machines." *Neural Computing and Applications* 26, no. 3 (2015): 641-649.
- [15] Ławryńczuk, Maciej. "Modelling and predictive control of a neutralisation reactor using sparse support vector machine Wiener models." *Neurocomputing* 205 (2016): 311-328.
- [16] Xia, Youshen, and Jun Wang. "A one-layer recurrent neural network for support vector machine learning." *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 34, no. 2 (2004): 1261-1269.
- [17] Tang, Duyu, Bing Qin, and Ting Liu. "Document modeling with gated recurrent neural network for sentiment classification." In *Proceedings of the 2015 conference on empirical methods in natural language processing*, pp. 1422-1432. 2015.
- [18] Li, Zhangjun, Xujian Feng, Ziqian Wu, Cuiwei Yang, Baodan Bai, and Qunqing Yang. "Classification of Atrial Fibrillation Recurrence Based on a Convolution Neural Network With SVM Architecture." *IEEE Access* 7 (2019): 77849-77856.
- [19] Patel, Alpna, and Arvind Kumar Tiwari. "Sentiment Analysis by using Recurrent Neural Network." Available at SSRN 3349572 (2019).
- [20] Lee, Scott H., Drew Levin, Patrick D. Finley, and Charles M. Heilig. "Chief complaint classification with recurrent neural networks." *Journal of biomedical informatics* 93 (2019): 103158.
- [21] Gangwar, Satyam, Vikram Bali, and Ajay Kumar. "Comparative Analysis of Wind Speed Forecasting Using LSTM and SVM." (2019).
- [22] Niculae, Vlad, Joonsuk Park, and Claire Cardie. "Argument mining with structured SVMs and RNNs." *arXiv preprint arXiv:1704.06869* (2017).
- [23] Benrachou, Djamel Eddine, Filipe Neves dos Santos, Brahim Boulebtateche, and Salah Bensaoula. "Online vision-based eye detection: LBP/SVM vs LBP/LSTM-RNN." In *CONTROLO'2014—proceedings of the 11th Portuguese conference on automatic control*, pp. 659-668. Springer, Cham, 2015.
- [24] Peng, Yifan, Anthony Rios, Ramakanth Kavuluru, and Zhiyong Lu. "Chemical-protein relation extraction with ensembles of SVM, CNN, and RNN models." *arXiv preprint arXiv:1802.01255* (2018).