JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

Supportive Vector Regression (SVR) Hybridization and Evolutionary Genetic Algorithm in Modeling for Prediction / With Practical Application

Sahera Hussein Zain Al-ThalabiHuda Abd El-Sadah
Email: sahera.zain@uobasrah.edu.iqEmail: aser33588@gmail.com
University of Basra / Faculty of Administration and Economics / Department of Statistics

Received: 2022 March 15; **Revised:** 2022 April 20; **Accepted:** 2022 May 10

Abstract:
The forecasting of electrical loads critically supports energy management policy decisions. The aim of this study is to develop methods for predicting electrical load, as vector regression has been applied to predict electrical load, and the quality and stability of the SVR model depends greatly on the selection of optimal parameters, and the study proposes a new approach: Evolutionary SVR algorithms that solve problems of optimizing all parameters (SVR) To evaluate the values of three typical parameters, the supporting vector regression was combined with the genetic algorithm and chaotic genomic algorithm (SVR), and the electrical load 1980-2019 was used as a dataset, and compared the results (GASVR) with (CGASVR) to choose the best form for predicting the electrical load where the results show that more superior and efficient CGASVR model based on MSE, MAE, MAPE and MPE criteria to predict validity.

Key terms of study: support vector regression, genetic algorithm, fitness function, chaotic genetic algorithm

Introduction:

Electric power is one of the main sources of energy that contribute significantly to all sectors that promote the development process of the country's progress. Electricity is one of the main sources of

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

energy that contributes significantly to all sectors that promote the development process of the country's progress, as a result of the increased use of the basic requirements in the economy, it plays an important role in the process of economic and cultural growth and prosperity.

The use of electric energy is one of the most important aspects of civilization and development. It is a measure of progress and prosperity in any society because of the services that this energy provides. The accuracy of forecasting future electricity loads has thus received increased attention for its important role in future planning state economic eectors. The study aims to find the best statistical method used to predict electricity consumption by selecting the best statistical model for forecasting. Many researchers have used artificial intelligence techniques to improve prediction of electric loads where the neural network showed improved performance is acceptable for improving the accuracy of forecasting electricity loads, but she has problems as she takes a long time training and also has difficulty finding the optimal solution, to overcome its shortcoming (NN) has received many hybrid applications along with statistical methods, SVR was used for its outstanding non-linear performance to solve prediction problem a model (SVR) used with both evolutionary hybrid algorithms makes the resulting model an ideal and efficient tool for increasing the predictive accuracy of an electrical load,also, non-linear CGA use is excellent performance for searching for ideal parameters in a form (SVR). Three key parameters (C, σ, ε) play a critical role in the performance of SVR, since prediction accuracy suffers from a lack of knowledge in the selection of parameters in (SVR). To create a form (SVR), you must set and choose parameters (SVR) carefully, the choice of parameters in the (SVR) model greatly affects the accuracy of prediction. So this study proposes evolutionary algorithms to determine the supporting vector regression parameters that solve the problems of optimizing all parameters (SVR).

Thestudy[4] hybridized a supporting vector regression model with the (GA) to predict the daily tourist flow of holidays. Daily tourist flow data are used during holidays from 2008 to 2012, results have shown that the (GA-SVR) model is more effective and reliable than other alternative models.

A team of researchers [15] created an improved model (SVR) for predicting damping ratio in the cable package of particles, and the use of (GA) to identify optimal variables for improved predictability, the results found that the (SVR)that supports the proposed(GA) provides better predictability.

A study [7] provided a model for predicting electrical load based on (SVR) and (CGA) to improve prediction performance. To solve premature local optimization in choosing three parameters of the

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

(SVR) model, the proposed (CGA) is used, which is based on the chaos optimization algorithm and (GA), according to the research results, the(SVRCGA) model provides more accurate and reliable prediction results than(ARIMA) and other alternative models.

Researchers [1] presented with a model of (SVR) and (GA) using the crystal network skeletal parameter to estimate the magnetic temperature of manganite-based materials, whereas the (GA_SVR) was compared to the (GSA_SVR) model on a fault-tolerant basis, the study found that the (GA_SVR) model showed better performance combined with less computational time.

A study [5] provided a model application (SVR) for traffic flow prediction (CGA), If you use (CGA) to defeat local optimization in selecting three form parameters (SVR), the results have shown that (CGA) reduces the prediction error and raises the prediction accuracy.

The current study dealt with a theoretical aspect that dealt with statistical methods used in forecasting and how to improve the accuracy of forecasting and another applied aspect in which the results were analyzed and the selection of the best model for predicting the electric load in the southern region.

support vector regression (*SVR*)

Vapnik support vector regression was developed in 1995, by introducing an alternative loss function, the regression version of (SVM) appears as an alternative and effective way to solve regression problems. This version is called vector support regression (SVR).

Instead of using the prediction error in the training set (experimental risk reduction principle), the support vector regression formulation (SVR) follows the structural risk reduction principle, which is to lower the upper limit of the generalization line by entering strokes on a structure or curve a collection of functions to which an estimate is made(SVR) provides greater potential for generalizing the relationship between inputs and outputs acquired during the training process, allowing it to make accurate predictions using new input data.[3]

(SVR) is a nonlinear technique based on basic kernel, and to prove the principle of (SVR), consider the following regression problem:We assume that a set of training data has been obtained

$$s = \{(x_1, y_1), \dots \dots .., (x_1, y_1)\}$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

Whereas: x_i: is the input vector ‘y_i: Is its output values associated with xi.

Regression function (SVR) which is

$$f(x) = w \times \Phi(x) + b \dots \dots \dots (1)$$

Whereas: Φ(x): refers to the function of nonlinear mapping,:w:is the weight vector,

b:represents the word "bias". [6]

It can be calculated by reducing the regular risk function using the following formula:

$$Risk = \frac{1}{n} C \sum_{i=1}^{n} L_s\big(y_i‘f(x)\big) + \frac{1}{2}\|w\|^2 \dots \dots (2)$$

Whereas:

$$L_s\big(y_i, f(x)\big) = \begin{cases} |y_i - f(x_i)| - \varepsilon, |y_i - f(x_i)| \geq \varepsilon \\ 0 \qquad\qquad\qquad\qquad otherwise \end{cases}$$

Denotes the loss function (ε-insensitive);$\frac{1}{2}\|w\|^2$:represents the organizer that controls the trade-off between complexity and approximate accuracy regression model to ensure that the model can improve generalized performance, C: represents a function of the cost of measuring experimental risks.

Non-negative stagnation variables(ζ_i, ζ^*) can be used to represent the distance between real values to ε values in equation (2), so that the constrained model can be formulated according to the following:[13]

$$\min \quad \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{m} (\zeta_i, \zeta^*) \dots \dots (3)$$

$subject\ to$

$$y_i - [w \times \Phi(x)] - b \leq \varepsilon + \zeta_i$$

$$[w \times \Phi(x)] + b - y_i \leq \varepsilon + \zeta^*$$

$$\zeta_i, \zeta^* \geq 0; \;\; i = 1,2,3, \dots \dots l$$

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

Restricted optimization problem using the lagrangian model as follows:

$$\text{Maximize} \sum_{i=1}^{l} y_i(\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^{l} (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} (\alpha_i - \alpha_i^*) \left(\alpha_j - \alpha_{ii}^*\right) K\left(x_i, x_j\right) \ldots (4)$$

$subject\ to$

$$\sum_{i=1}^{l} (\alpha_i - \alpha_i^*) = 0 ، \alpha_i, \alpha_i^* \ \in [0, C]، \quad i = 1,2,3, \ldots \ldots .. l$$

Whereas: α_i, α_i^* : are the lagrangian multipliers,$K\left(x_i, x_j\right)$: is the kernel function.and by solving the problem the best solution can be found as follows: [12]

$$f(x) = \sum_{i=1}^{l} (\alpha_i^* - \bar{\alpha}_i) K\left(x ، x_i\right) + \bar{b} \ldots \ldots .. (5)$$

The kernallfunction is defined as follows:

$$K\left(x_i, x_j\right) = \emptyset(x_i), \emptyset\left(x_j\right) \ldots \ldots . (6)$$

Using kernel functions, the support vector regression (SVR) can be calculated without explicit mapping in the feature space. [6]

There are four types of kernel functions used in the slope of the supporting vector (SVR): linear, polynomial, root base function (RBF), and sigmoid. performs three parameters such as regular parameter (C), kernel parameter (σ), and loss function parameter (ε),these parameters play an important role in support vector regression (SVR) performance. Consideration should be given to determining the appropriate parameters. [14]

Genetic algorithm

(GA) is a random search strategy that uses natural genetics and evolutionary theory to search vast and complex areas.(GA) deals with a series of chains and independent (chromosomes), each of which constitutes a possible solution to the problem. The results of the fitness function are used to set the fitness score of each individual. (GA)is ideally suited to solve large-scale nonlinear optimization problems, and tends to find the optimal solution. [15]

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

Use genetic algorithm (GA) to optimize support vector regression parameters (SVR).

(SVR)parameters must be set carefully to build an effective (SVR)model.Increasing fit occurs as a result of the three incorrect parameters used in the supporting vector regression model (SVR) and the three parameters are(C, ε, σ).[15]Below are the steps to improve the parameters (SVR) using a genetic algorithm (GA):

Step 1 (binary string encoding):An initial set of solutions is generated randomly representing parameter values in a model (SVR).To form a binary string, the parameters (SVR) are specially encrypted. C is set to avalue [1,100], The ε range is set to [0.0001, 0.01], and the parameter σ is set to [0,1].

Step 2 (Evaluate Fitness):To assess the efficiency of the set of parameters (C, ε, σ), it is carried out using the average square root of error (RMSE). [10]In the training data set is written to the fitness function as (RMSE), which is defined as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{N}} \ldots\ldots\ldots.(7)$$

Whereas: y_i:actual value,$\hat{y}_i$: expected value, N: is the number of forecast periods.[2]

Step 3 (GA operators): To choose excellent solutions for repetition operators to use the roulette wheel procedure for each pair of individual chains(chromosomes), the probability of finding new solutions is (0.5). After the mutation process is a crossover process that determines whether the solutions intersect with the next generation.

Step 4 (Stopping criteria): to improve the solution until the condition of cessation is achieved, The above steps are repeated for each of the parameters (C, ε, σ) corresponding to the lowest error in the model, with the best parameter (C, σ, ε) determined by the optimal structural fitness function.[10]

fitness function

A fitness function must be constructed to measure the efficiency of all solutions before searching for optimal (SVR) values,several estimation criteria (MAPE, RMSE, MSE) have been proposedwhich are used to compare the performance of the new model, and calculate the accuracy of the model estimation.

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - f_i}{y_i}\right| \times 100\% \text{ ،} MSE = \frac{\sum_{i=1}^{n}(\sigma_t - \hat{\sigma}_t)^2}{n} \text{ ،} MAE = \left|\frac{\sum_{i=1}^{n} x_i - x}{n}\right|$$

Whereas: σ_t :Is the observed volatility of the period t,$\hat{\sigma}$:Is the expected volatility of the period t,n: Total expected time period,$y_i,\ f_i$: represents actual values and prediction,N:Number of forecast periods.

To measure the results, the average percentage of error is usedthe prediction model is more accurate if the (MAPE)values are lower, the lower the value, the closer the results are to the actual data. [6]

chaotic genetic algorithm

Genetic algorithm (GA) is a flexible and powerful way to solve the optimization problem,However,there are two main disadvantages in (GA) the slow convergenceand their occurrence at the local optimal level, Hybrid chaos sequences and (GA)are therefore combined to overcome the problem of convergence confined to local optimization which results mainly from reducing population diversity.

The chaotic genetic algorithm (CGA) is an effective method for solving nonlinear and complex optimization problems. Its basic idea is to transform the problem variable from the solution space to the chaos space and then research the solution using three properties of chaotic variables: randomness, ergodicity, and regularity. [8]

Using chaotic genetic algorithm (CGA) to optimize support vector regression parameters (SVR)

The correct configuration of the three parameters (C, σ, ε) is critical to the efficiency and accuracy of the supporting vector regression (SVR).Often, getting each parameter to an optimal point does not produce good results for the support vector regression (SVR), so we can only get good performance when their combination reaches an optimal value. [9]

Choosing the correct values for the supporting vector parameters has a significant effect on the performance and quality of the support vector regression (SVR). It represents one of the important problems in SVRthe accuracy of the support vector regression model (SVR)depends largely on the determination and selection of the parameters of the model (C, σ, ε).So you try different combinations of the three parameters by sampling the search space intermittently. Once the composition with

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

minimal quadratic error is found, it can be searched by reducing the sample interval around the set, the procedure is replicated until the validation precision does not change significantly, as a result, it is suggested that the chaotic genetic algorithm (CGA) be used to boost the parameters of (SVR).[11] [9]

Statistical analysis

We are using a real data set for electricity load consumption in the southern area for the period (1980-2019)which includes the dependent variable (annual electricity load consumption in the southern area(MW, h)) the independent variables that measure the impact of thesupporting vector regression model are both,(Gross Domestic Product (GDP) in Iraqi dinars, population(thousand persons),average per capita income in Iraqi dinars,the number of participants in the southern region,kilowatt price in dinars /kwh).To predict through the best model using (GA),(CGA), this is done by estimating parameters (SVR) which are (C, σ, ε),and use the (cross-valinda) method to divide the data differently and then divide it into three chapters to be estimated k=3.

Table (1) estimation of parameters (SVR) and genetic algorithm (GA)

Optimum parameters	The average square of the error	Volatility model
$\boldsymbol{\delta}$	2.362	3.265
C	6.214×10^{3}	9.632×10^{5}
$\boldsymbol{\varepsilon}$	0.31	0.56
Number of inputs	K=3	p=1, q=2
Random error smaller	2.362	3.951

Source: from the work of the researcher using the program (Matlab 7.11)

The table above shows the estimation of parameters (C, ε, σ), and the effect of the genetic algorithm model (GA) after multiple iterations to obtain a better technique for the model through the least random error،The results indicated that the average error of the variance (2.362) and in the model (3.265), and the value of the average error in the constant (6.214×10^{3}) and the model (9.632×10^{5})،and the value of the average error in the standard error (0.31) and the model (0.56).

JOURNAL OF ALGEBRAIC STATISTICS
Volume 13, No. 3, 2022, p. 5025-5037
https://publishoa.com
ISSN: 1309-3452

Table (2) estimation of parameters (SVR) with chaotic genetic algorithm (CGA)

Optimum parameters	The average square of the error	Volatility model
$\boldsymbol{\delta}$	1.963	3.251
C	4.263×10^3	6.952×10^5
$\boldsymbol{\varepsilon}$	0.23	0.59
Number of inputs	K=3	p=1, q=2
Random error smaller	2.563	3.842

Source: from the work of the researcher using the program (Matlab 7.11)

The above table shows the estimation of (SVR) parameters by the chaotic genetic algorithm (CGA) after multiple iterations to obtain a better technique for the model with the least standard error, the results indicated that the average error for the variation (1.963) and in the model (3.251), the value of the mean error for the constant (4.263×10^3), for the model (6.952×10^5), and the value of the average error for the standard error and model (0.23), (0.59), which confirms that the use of the (CGA) affects the (SVR) model and has high and effective techniques through the standard error values because it dropped after the process was repeated.

From the two above models, it is evident that each model has values that indicate its quality and accuracy in selecting the required information. Given statistical equations and laws, we find that the model (CGA-SVR) is the best in terms of the accuracy required in analyzing the data. So a support vector regression method was chosen with the chaotic genetic algorithm to predict for eight futures years, electrical energy consumption.

Table (3) future forecast values for electric power consumption (2022-2030)

Forecast values of consumption using an algorithm CGA+SVR(Gigawatt hours)	the years

9236057	**2022**
12790829	**2023**
16353621	**2024**
19568089	**2025**
21317833	**2026**
22847328	**2027**
22389174	**2028**
24915560	**2029**
9236057	**2030**

Source: from the work of the researcher using the program (Matlab 7.11)

Conclusions:

The support vector regression was used with the anarchic genetic algorithm in this analysis to propose a new approach to predicting electrical loads, which showed several conclusions:

- A chaotic genetic algorithm is an excellent tool for searching for the ideal parameters of the support vector regression and it works better than a standard genetic algorithm.
- Combining the supporting vector regression with the chaotic genetic algorithm makes the resulting model (CGASVR) more efficient and accurate to predict electrical energy consumption.
- The model (CGASVR) is valid for predicting electrical energy consumption and gives better results for forecasting than the model (GASVR).

References:

[1] Abdullah, A.,Taoreed O. O.,Kabiru O. A.,Sunday O. O.,Nahier.A., (2018) Hybrid Support Vector Regression and Genetic Algorithm Model for Tuning Magnetic Ordering Temperature of Manganite Refrigerant,Research Journal of Applied Sciences 13 (1): 87-93.

[2] *Cao, G., & Wu, L., (2016). Support vector regression with fruit fly optimization algorithm for seasonal electricity consumption forecasting. Energy, 115, 734–745. doi: 10.1016/j.energy.2016.09.065*

[3] Chen, K.-Y., & Wang, C.-H., (2007). *Support vector regression with genetic algorithms in forecasting tourism demand. Tourism Management, 28(1), page 215–226.*

[4] Chen, R., Liang, C.-Y., Hong, W.-C., & Gu, D.-X., (2015). *Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm. Applied Soft Computing, 26, 435–443.* doi: 10.1016/j.asoc.2014.10.022

[5] Deng, Y., Xiang, J., &Ou, Z., (2012). *SVR with hybrid chaotic genetic algorithm for short-term traffic flow forecasting. 2012 8th International Conference on Natural Computation.* doi:10.1109/icnc.2012.6234768

[6] Fong-Ching Yuan, (2012). Parameters Optimization Using Genetic Algorithms in Support Vector Regression for Sales Volume Forecasting ‹Department of Information Management, Yuan Ze University, page 1481, Chung-Li, Chinese Taipei.

[7] Hong, W.-C., Dong, Y., Zhang, W. Y., Chen, L.-Y., & K. Panigrahi, B. (2013). *Cyclic electric load forecasting by seasonal SVR with chaotic genetic algorithm. International Journal of Electrical Power & Energy Systems, 44(1), 604–614.* doi: 10.1016/j.ijepes.2012.08.010

[8] Hong, W.-C., Dong, Y., Chen, L.-Y., & Wei, S.-Y. (2012). *Seasonal Support Vector Regression with Chaotic Genetic Algorithm in Electric Load Forecasting. 2012 Sixth International Conference on Genetic and Evolutionary Computing.* doi:10.1109/icgec.2012.128

[9] *Li Xuemei, Ding Lixing, Li Yan, Xu Gang, & Li Jibin. (2010). Hybrid Genetic Algorithm and Support Vector Regression in Cooling Load Prediction. 2010 Third International Conference on Knowledge Discovery and Data Mining. doi:10.1109/wkdd.2010.136*

[10] Liu, S., Tai, H., Ding, Q., Li, D., Xu, L., & Wei, Y. (2013). *A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Mathematical and Computer Modelling, 58(3-4), page 458-465.*

[11] Wang, W., Xu, D., &Qiu, L. (2010). *Support vector machine with chaotic genetic algorithms for annual runoff forecasting. 2010 Sixth International Conference on Natural Computation,* page 671-675.

[12] Wang, Y., Wang, B., & Zhang, X. (2012). *A New Application of the Support Vector Regression on the Construction of Financial Conditions Index to CPI Prediction. Procedia Computer Science, 9, 1263–1272.*

[13] Wei-Chiang Hong, 2011, Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm, Department of Information Management, Oriental Institute of Technology, 58 Sec. 2, Sichuan Rd., Panchiao, Taipei County 220, Taiwan, ROC, page 5570.

[14] Wenjun Wu, 2016, Hybrid Genetic Algorithm and Support Vector Regression Performance in CNY Exchange Rate Prediction, School of Statistics and Mathematics Zhongnan University of Economics and Law Wuhan, page 0137, 430073, China.

[15] Xia, Z., Mao, K., Wei, S., Wang, X., Fang, Y., & Yang, S. (2017). *Application of genetic algorithm-support vector regression model to predict damping of cantilever beam with particle damper. Journal of Low Frequency Noise, Vibration and Active Control, 36(2), 138–147.*

[16] Turki Jalil, A., Hussain Dilfy, S., Oudah Meza, S., Aravindhan, S., M Kadhim, M., & M Aljeboree, A. (2021). CuO/ZrO2 nanocomposites: facile synthesis, characterization and photocatalytic degradation of tetracycline antibiotic. *Journal of Nanostructures*. https://doi.org/10.22052/JNS.2021.02.014

[17] Sarjito, Elveny, M., Jalil, A., Davarpanah, A., Alfakeer, M., Awadh Bahajjaj, A. &Ouladsmane, M. (2021). CFD-based simulation to reduce greenhouse gas emissions from industrial plants. *International Journal of Chemical Reactor Engineering*, 20210063. https://doi.org/10.1515/ijcre-2021-0063

[18] Marofi, F., Rahman, H. S., Al-Obaidi, Z. M. J., Jalil, A. T., Abdelbasset, W. K., Suksatan, W., ... &Jarahian, M. (2021). Novel CAR T therapy is a ray of hope in the treatment of seriously ill AML patients. *Stem Cell Research & Therapy*, *12*(1), 1-23. https://doi.org/10.1186/s13287-021-02420-8

[19] Jalil, A. T., Shanshool , M. T. ., Dilfy, S. H. ., Saleh, M. M., & Suleiman, A. A. . (2022). HEMATOLOGICAL AND SEROLOGICAL PARAMETERS FOR DETECTION OF COVID-19. *Journal of Microbiology, Biotechnology and Food Sciences*, e4229. https://doi.org/10.15414/jmbfs.4229

[20] Vakili-Samiani, S., Jalil, A. T., Abdelbasset, W. K., Yumashev, A. V., Karpisheh, V., Jalali, P., ... &Jadidi-Niaragh, F. (2021). Targeting Wee1 kinase as a therapeutic approach in Hematological Malignancies. *DNA repair*, 103203. https://doi.org/10.1016/j.dnarep.2021.103203

[21] NGAFWAN, N., RASYID, H., ABOOD, E. S., ABDELBASSET, W. K., Al-SHAWI, S. G., BOKOV, D., & JALIL, A. T. (2021). Study on novel fluorescent carbon nanomaterials in food analysis. *Food Science and Technology*. https://doi.org/10.1590/fst.37821

[22] Marofi, F., Abdul-Rasheed, O. F., Rahman, H. S., Budi, H. S., Jalil, A. T., Yumashev, A. V., ... &Jarahian, M. (2021). CAR-NK cell in cancer immunotherapy; A promising frontier. *Cancer Science*, *112*(9), 3427. https://doi.org/10.1111/cas.14993