International Journal of Electrical and Computer Engineering (IJECE)
Vol. 8, No. 5, October 2018, pp. 3341~3348
ISSN: 2088-8708, DOI: 10.11591/ijece.v8i5.pp3341-3348

Initial Optimal Parameters of Artificial Neural Network and Support Vector Regression

Edy Fradinata[1], Sakesun Suthummanon[2], Wannarat Suntiamorntut[3]
[1]Industrial Engineering Department, Syiah Kuala University, Banda Aceh, Indonesia
[1,2]Industrial Engineering Departement, Prince of Songkla University, Hatyai, Thailand
[3]Computer Engineering Departement, Prince of Songkla University, Hatyai, Thailand

Article Info

Article history:

Received Jan 16, 2018
Revised Mar 3, 2018
Accepted Mar 23, 2018

Keyword:

ANN
MSE
Optimization
Supervised
SVR

ABSTRACT

This paper presents architecture of backpropagation Artificial Neural Network (ANN) and Support Vector Regression (SVR) models in supervised learning process for cement demand dataset. This study aims to identify the effectiveness of each parameter of mean square error (MSE) indicators for time series dataset. The study varies different random sample in each demand parameter in the network of ANN and support vector function as well. The variations of percent datasets from activation function, learning rate of sigmoid and purelin, hidden layer, neurons, and training function should be applied for ANN. Furthermore, SVR is varied in kernel function, lost function and insensitivity to obtain the best result from its simulation. The best results of this study for ANN activation function is Sigmoid. The amount of data input is 100% or 96 of data, 150 learning rates, one hidden layer, trinlm training function, 15 neurons and 3 total layers. The best results for SVR are six variables that run in optimal condition, kernel function is linear, loss function is ε-insensitive, and insensitivity was 1. The better results for both methods are six variables. The contribution of this study is to obtain the optimal parameters for specific variables of ANN and SVR.

Corresponding Author:

Edy Fradinata,
Departement of Industrial Engineering,
Syiah Kuala University,
Teuku Nyak Arief, Darussalam, Banda Aceh, 23111, Indonesia.
Email: edinata69@gmail.com

1. INTRODUCTION

Artificial Neural Network (ANN) is a structure of learning systems where it is inspired by living organisms, especially to a human system. It consists of a very complex network that is equipped with some neurons which are interconnected each other, these neurons work to remember, to calculate, to generalize, to adapt, to get low dynamism and has high flexibility. SVR is a method to contribute the solution by small subset from the training points where produce the enormous computational advantages. The e-insensitive loss function pretends the existence of the global minimum solution and the optimization bound [1].

Support Vector Regression (SVR) can improve various interesting features and produce a better performance [2]. The calculation is constructed on the conception of minimization in structural risk. The concept of performance is better than the traditional Empirical Risk Minimization (ERM) where it was worked in conventional neural networks [3]. Actually, SVM has the purpose to solve the classification condition, but lately it can be used in the regression domain. Originally, it was designed for solving pattern recognition. Determination of hyperplane is separating the positive and negative environment value of them. This method is very command used in fundamental risk minimization and numerical learning theory [4]. The learning and training error rate were used for testing in the limited data error. ANN and SVR are the methods that could be run data to find the best model with their data's characteristic [5]. The model will represent the condition of data accuracy [6]. The model was treated the best accuracy if the combination of the hidden layer, the neuron, the activation function and the kind of training function contribute the smaller Mean Square Error (MSE) belong to the kinds of data that forecasted where it was compared to the original data. The combination of parameters that run is called architecture [7].

This paper is mainly propose an ANN and SVR approach to choose the best fit of parameters before it could be used to the specific steps of the network. ANN's parameters are a variety in percent of data, the hidden layer, the neuron, the transfer function, and training function. Some parameters of SVR are: kernel function, lost function and insensitivity. The architecture will influence the result of measurement of a network.

2. METHODOLOGY

The methodology will brief the view step of architecture, each parameter that representing both methods between ANN and SVR. In this research will focus in a backpropagation network and eisensitive to SVR[8]. The process of methodology is illustrated at Figure 1.

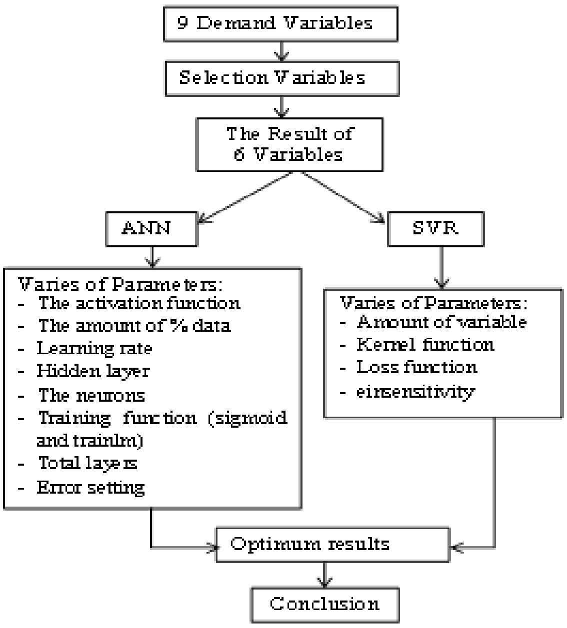

Figure 1. Methodology of study

3. RESULTS AND ANALYSIS

3.1. Data experiment

The variables determinatof demand are GDP growth (D1); Population (D2), A potential customer (D3),; Price (D4); Sales (D5); Advertising (D6); Quality (D7); Expectation future price (D8); Preference price,(Trend seasonal) (D9) [9]. The fluctuations of data show the characteristic of time series data set in monthly basis or in 8 years cement demand [10].

3.2. Design of ANN parameters

3.2.1. Test of input variable

The difference variable has been calculated above with selected data correlation to demand and the total dataset. This experiment shows the influence of the amount of input variables with sigmoid as a transfer function Table 1. Table 1 the variables from 2 variables were varied to 6 variables. When the amount of variables increase, the MSE tends to decreased. The smallest was 6 variables, with the MSE 3.78e-^{6} (Post processing value but it was not reple back to the initial scale, it is eligible to compared each other). The purpose model of ANN is shown in Figure 2(a) and Figure 2(b).

Table 1. Test Run the Amount of Variables

Amount of variables	MSE
2 (D3, D4)	4.20e-6
4 (D3, D4, D5, D6)	6.22e-6
6 (D3, D4, D5, D6, D7,D9)	3.78e-6

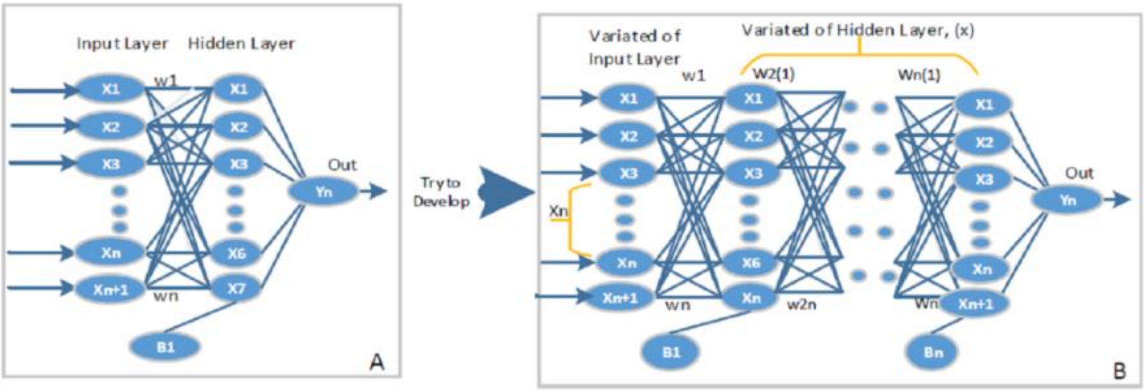

Figure 2. The purpose concept of backpropagation neural network

3.2.2. Test of entrance dataset

Six variable input data was varied from 40% to 100% then measured the MSE, the resulted can be seen in Table 2. Table 2 when percent of data increase the MSE decrease. The minimum MSE results 100% of data.

Table 2. Varying Percent Input of Data

Percent Data	Feed Data	MSE
40%	38	8.03e-7
50%	48	7.79e-7
60%	58	7.88e-7
70%	68	7.77e-7
80%	78	6.61e-7
90%	88	5.68e-7
100%	96	4.73e-7

3.2.3. Test difference of activation function

The test for this activation function threated 2 kinds of activation function. They were sigmoid and purelin. This activation aimed to pursue the activated of the data to process their range. Table 3 shows if the variable increased the MSE of sigmoid tend to the decreased weather has a peak at the 4 variables. Variable 6 is smallest for sigmoid.

Table 3. Run with different Activation Function

No	Variables	MSE	
		Sigmoid	Purelin
1	2	4.20e-6	5.30e-6
2	4	6.22e-6	1.13e-6
3	6	3.78e-6	4.26e-6

3.2.4. Test of learning rate

The learning rate tried some kinds of rate: 50,100,150 and 200. It can be seen in Table 4. Table 4 shows that learning rate increased to contribute the impact of the MSE decreased at point 150. This point was contributed the smallest error with 0.000189.

Table 4. Run with a different Learning Rate

Neuron	MSE
50	0.000201
100	0.000225
150	0.000189
200	0.000210

3.2.5. Test of a hidden layer

Hidden layer set try with 1, 2 and 3, (from: Sigmoid, 96 data). The test with random blocking, see in Table 5. The Table 5 shows the group layers combine in three observations. It tends to decrease in their pool. Then from the layer 1 MSE is the smallest at 0.000248. The 1 layer will be used [11].

Table 5. Run the Hidden Layer

Observation	(Group) Layer	Result (MSE)
1	1	0.000349
2	1	0.000248
3	1	0.000378
4	2	0.000256
5	2	0.000339
6	2	0.000363
7	3	0.000313
8	3	0.00037
9	3	0.000445

3.2.6. Test the amounts of neuron in the Layer

The test amount of neuron were tested with 3 different neurons, 6, 8 and 10. It shows in Table 6. Table 6 shows that the amount of neuron was increased will contribute the MSE was decreased and 10 neurons were the best contributed to error.

Table 6. Run with a different Amount of Neuron

Amount of Neuron	MSE
6	0.00481
8	0.00452
10	0.00310

3.2.7. Test of network training function

The various network training functions are applied in this experiment to see the effectivity each network training function, it can be seen in Table 7. Table 7 shows the variety of network training functions and the best training function is Trainlm and the second is Traingdm. The study is tried with six variables and shows in the Table 1 that the amount of variables are increased the MSE decreased with minimum 3.78e-6. The variable will influence the result output of prediction, in this research six variables are better amount than the smaller dataset, this is very reasonable for neural network powerful to simulate nonlinear belong the number of different variables in horizon terms of time [12].

Then at Table 2 shows the amount of data increase while the MSE decrease and the best percentage is 100% or 96 amounts of data with MSE 4.73e-7. This is reasonable for the bigger data should improve the better result of prediction from the output pattern of neural network for this characteristic of dataset. This is relevant to the theory of neural network that neural network is better working with big data then smaller, because the smaller data could not do the training process more accurately and the bias will be higher [13], [14].

Table 3 shows the test of activation functions are varied with Sigmoid and Purelin, the best activation function is sigmoid on six variables compare to each other on the same amount of variables so it would be used for the parameters to keep smoothly running to execute data on the range of 0 to 1. Table 4 shows the different learning rate from 50 to 200 and the best one was at 150. This learning rate will help the data to process in the overlap of the real data before testing the data to prediction. In this section, the rule defines the network weight on trial and error by an epoch. The error is updated to supervised learning until found the smaller network error [15].

Table 5 shows the varieties of layer and the observation of the best MSE with 0.000248 with layer 1, the layer to help the data running on the optimal to keep the over fitting process, because if too many layer will not take the long time process. It will also occure the over fitting with the data weather the layer should be obtain the better result but the overfitting will be stopped on the process of forecasting dataset. In this step, the number of weights do the iterate calculation to the hidden layers part. The numbers of weights are depending on the size of training set to the individual reflection of data and use for actual forecasting dataset. On the table shows, that more amount of hidden layers contributes unsatisfy result after it was run more amount of hidden layers. Some recommendations from other researchers very common to use one of hidden layer is better than more in a process of it.

Table 6 from this experiment show the “differences” amount of determination neurons in hidden layer, the step was starting from the smallest number to higher number of neurons where the contribution of the neuron significantly to get the smaller MSE. The random number sample are taken from 6 to 10 and this determination obtain the best MSE 0.00310 with 10 neurons [16]. Table 7 tells the variation of training function where the best result is Trainlm with MSE 0.000234 in algorithm of Levenberg Marquadt. In this case, Trainlm is better than sigmoid weather sigmoid is more common is used to train backpropagation algorithm [17].

Table 7. Different network training function

Training Function	MSE
Trainlm	0.000234
Traingdm	0.000428
Trainingda	0.000817
Traingdx	0.00110

From the discussion part, it can conclude that the result will be given the best fit of data if use the selected variables, meaning that the result from each training function will be influenced significantly and reduce the overfitting process to obtain the optimal condition.

3.3. Design of SVR’s parameters

There are some parameters in SVR to construct the SVM for predicting. However, the two dominant relevant are e-insensitivity and kernel function because both parameters could be increased the e-mean and decreased the error and increasing the accuracy of the process of data. It can decrease the number of SVs leading to data compression. The parameters of SVR are kernel function, ε-insensitive loss function, insensitivity, an upper bond. The test is using the different amount of data. The data will be used 6 variables. Kernel Function: Linear, Polynomial, Radial Basis Function, Tangent Hyperbolic, and Loss function’s parameters are e-insensitive, Quadratic, Laplace and Huber. Insensitivity is 1. *Kernel Function* is the classification problems in optimal condition σ can be computed based on Fisher discrimination. It is also to regression the problems in the basic of scale, space theory, and it is demonstrated the existence of a certain range of σ, within the generalization performance is stable. A certain important in the range of σ can be reached via dynamic evaluation. In conclusion, the lower bound of an iterating step size of σ is given. *Loss function* is the relationship function between error and the penalty to that error. The differences of loss function will produce the differences of SVR. Loss function ε-insensitive is the very common. The experiment starts from the 6 variables and measure the result of both parameters, such as:

3.3.1. Test of kernel function and loss function.

The kernel function and loss function were tested with linear, polynomial for Kernel, and e-insensitive for loss function. It can be seen in Table 8. Table 8, the linear is better than polynomial in a Kernel Function. It was the best choice for MSE. The other side loss function is better for einsensitive.

Table 8. Run different Kernel Function and Loss Function

Gaussian Kernel Function			Loss Function	
Statistic	Linear	Polynom	Statistic	e-insensitive
Means	0.2007	0.2007	Means	0.2007
MSE	0.0021	0.0257	MSE	0.0018
SD	0.0018	0.0018	SD	0.0021

3.3.2. Test the "upper bond"

Choose e-isensitivity as a focus on a variety of variables, the test with UpB 2 and 3 from Table 9, as follow: Table 9 the upper bond 2 and 3 are no changed at al. It can be chosen number 2, means that the result of the einsensitive whether it was changed, it would be no impact to the einsensitive.

Table 9. Run with a different Upper Bond in e-isensitive

UpB=2	einsensitive	UpB=3	einsensitive
Means	0.2007	Means	0.2007
MSE	0.0021	MSE	0.0021
SD	0.0018	SD	0.0018

3.3.3. Check the insensitive number 1 and 2

This test was varied of the insensitive: 1 and 2. The tested can be seen at Table 10. Table 10 insensitive 1 and 2 are tried with e-insensitive and both of them are the best. But usually better use the 1 insensitive. This also shows no effect to the result whether it is changed.Table 8 shows the variation of Kernel function and loss function. The kernel function variation is linear on good result and shown better than polynomial with 0.0021. The loss function is small enough to be used with einsensitive with MSE 0.0018. The kernel functions have function of constructing the nonlinear decision hyper-surface on the input space of SVR. Both of them must be selected correctly where the structure was defined on the dimensional feature space and order complex to end solution [18]. Other researcher uses the same Gaussian kernel function for predict the performance [19] but in this research try two kind of Gaussian kernel functions, they are linear and polynomial. Table 9 shows the upper bond try with 2 and 3 numbers, but it shows that no change whether it have been changed for both numbers, it can be seen in Table 9, this function to keep the accuracy in the hyperplane area where it was placed on the points of training dataset [20]. Generally, it uses one as the upper bonds for the experimental.

Table 10. Run with different i-nsensitivity 1 and 2

Ins=1	einsensitive	Ins=2	einsensitive
Means	0.2007	Means	0.2007
MSE	0.0021	MSE	0.0021
SD	0.0018	SD	0.0018

Table 10 shows the insensitivity with 1 and 2 with MSE 0.0021 and this matter also no changes the result of MSE from the different number, choose insensity 1, the insensitive have the function of to fit the training data from Table 10. As originally, the purpose use svm was for solving the pattern recognition cases, but lately has been extended to solve nonlinear regression estimation cases such as in academic and industrial platforms e-insensitive loss function [21]. For svr the result from each parameter will be influenced significantly by the result. Because the svr will transform the data to be linier separable in the feature space of hyperplane to be the best regression. This method has promised the good methods in the future.

4. CONCLUSION

Based on this study, this is the initial step to the next step for the future experiment and the varieties of parameters of demand could be influenced on the artificial neural network and support vector regression methods. It can be concluded as follow: ANN could be an effective run on the "differences" parameters with six input variables, each condition has the optimal point itself. The result of this study was as follow: the activation function was Sigmoid. The amount of feed data was 100% or 96, 150 learning rate, 1 hidden Layer, 10 neurons, trinlm for training function, 3 layers for total layer, set up error 0.001 and work with a network of feed-forward backpropagation. Furthermore, the SVR as well as the amount of variables were 6. The general parameters were used with linear kernel function, e-insensitive loss function, and one insensitivity. Some variables are tried in this study for svr but not show the significant changes, means that the svr does not need to identify the initial parameter especially for upper bond and insensitive.

If these initial condition of parameters are used to do the next step for other purpose the result of training and simulation should be quickly and easy to get the optimal condition of network process data due to the characteristic of neural network able to work in nonlinearity and produce the suitable task for other purposes performance. Actually there is no specify way to get the best result of the network process data, mostly, it is do it with trial and error but at least with this study the way to define the optimal condition first before to do many trial and error methodology. This study finding the way how to get the starting initial condition to start neural network process.

ANNs and SVR are very talented method to better performance of the network result for many purposes, such as for forecasting, robotic, automotive, medical equipment's and many things else. Some researchers have many compared the performance of traditional methods which is study in statistical major to these methods, specially neural network methods but for SVR the study still limited and need to develop more knowledge finding many information about this methods.

Both of these methods are very special case because they do not need the statistical testing method specifically. Linear and nonlinear can do with this methods, parametric and nonparametric as well. Even, these methods will be better working with the big dataset, because it can easy to train the dataset and give the better result. The Suggestion to the next study is a development of these optimization condition's parameters for ANN and SVR to do the further study such as forecasting of determinant of demand with development of other method or hybrid method.

ACKNOWLEDGEMENTS

This research was under scholarship of 2012 Kemristekdikti of Indonesia. Thank you very much to Kemristek Dikti of Indonesia and Prince of Songkla University, Hatyai, Thailand.

REFERENCES

[1] T. B. Trafalis and B. Santosa, "Predicting monthly flour prices through Neural Networks, RBFs and SVR", *Intelligent Engineering Systems Through Artificial Neural Networks,* vol. 11, pp. 745-750, 2001.
[2] H. Drucker, *et al*., "Support vector machines for spam categorization", *Neural Networks, IEEE Transactions on,* vol. 10, pp. 1048-1054, 1999.
[3] K. Muller, *et al*., "An Introduction to Kernel-based Learning Algorithms", *Neural Networks, IEEE Transactions on,* vol. 12, pp. 181-201, 2001.
[4] I. B. Tijani and R. Akmeliawati, "Support Vector Regression based Friction Modeling and Compensation in Motion Control System", *Engineering Applications of Artificial Intelligence,* vol. 25, pp. 1043-1052, 2012.
[5] B. Shan, *et al*., "Application of Online SVR on the Dynamic Liquid Level Soft Sensing", in *Control and Decision Conference (CCDC), 2013 25th Chinese*, 2013, pp. 3003-3007.
[6] H. Esen, *et al*., "Modeling a Ground-coupled Heat Pump System by a Support Vector Machine", *Renewable Energy,* vol. 33, pp. 1814-1823, 2008.
[7] S. J. Hanson, *et al*., "Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a "face" area?", *Neuroimage,* vol. 23, pp. 156-166, 2004.
[8] B. Santosa, Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis vol. 978, 2007.
[9] "Determinants of Demand", in *http://market.subwiki.org/wiki/Determinants_of_demand*, ed, 22 December 2012, collected on 12 December 2015.
[10] E. Fradinata, *et al*., "Forecasting Determinant of Cement Demand in Indonesia with Artificial Neural Network", *Journal of Asian Scientific Research,* vol. 5, pp. 373-384, 2015.
[11] E. Fradinata, *et al*., "ANN, ARIMA and MA Timeseries Model for Forecasting in Cement Manufacturing Industry: Case Study at Lafarge Cement Indonesia—Aceh", in *Advanced Informatics: Concept, Theory and Application (ICAICTA), 2014 International Conference of*, 2014, pp. 39-44.
[12] P. Bunnoon, "Electricity Peak Load Demand using De-noising Wavelet Transform integrated with Neural Network Methods", *International Journal of Electrical and Computer Engineering,* vol. 6, p. 12, 2016.
[13] M. N. Rao, *et al*., "A Predictive Model for Mining Opinions of an Educational Database Using Neural Networks", *International Journal of Electrical and Computer Engineering,* vol. 5, 2015.
[14] B.-H. Adil and G. Youssef, "Hybrid Method HVS-MRMR for Variable Selection in Multilayer Artificial Neural Network Classifier", *International Journal of Electrical and Computer Engineering,* vol. 7, p. 2773, 2017.
[15] I. A. Basheer and M. Hajmeer, "Artificial Neural Networks: Fundamentals, Computing, Design, and Application", *Journal of Microbiological Methods,* vol. 43, pp. 3-31, 2000.
[16] B. Sutijo, *et al*., "Forecasting Tourism Data using Neural Networks-Multiscale Autoregressive Model", *Jurnal Matematika & Sains,* vol. 16, pp. 35-42, 2011.
[17] M. T. Hagan and H. B. Demuth, "Neural networks for control," in *American Control Conference, 1999. Proceedings of the 1999*, 1999, vol. 3, pp. 1642-1656.
[18] K. Duan, *et al*., "Evaluation of Simple Performance Measures for Tuning SVM Hyperparameters", *Neurocomputing,* vol. 51, pp. 41-59, 2003.
[19] A. Smola, *et al*., "Asymptotically Optimal Choice of ε-loss for Support Vector Machines", in *ICANN 98*, ed: Springer, 1998, pp. 105-110.
[20] N. Cristianini and J. Shawe-Taylor, "An Introduction to Support Vector Machines and other Kernel-based Learning Methods", Cambridge University Press, 2000.

[21] V. Vapnik, *et al*., "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing", in *Advances in Neural Information Processing Systems 9*, 1996.

BIOGRAPHIES OF AUTHORS

Edy Fradinata received his bachelor and master degree at Institute Teknologi Sepuluh Nopember-Surabaya (ITS). Currently, He is studying for PhD program at Prince of Songkla University, Hatyai, Thailand. He has some work experinces at Petrochemical plant, chemical, INGO/UNs. Interesting researchs are in Optimalization linear and non-linear, Data mining and heuristic, big data, SCM, Perf.Mgnt, MCDM, chemical process and manufacture industry, GIS, etc.

Sakesun Suthummanon received his M.B.A (Business Administration), Prince of Songkla University, B.Eng. (Industrial Engineering), Maha Nakorn Prince of Songkla University, then he continued his study for M.Sc. and Ph.D. (in Industrial Engineering), University of Miami, Fiorida, USA. He interesting research on Engineering Economics, Production and Operations Management, Quality Management, Logistics and Supply Chain Management, etc.

Wannarat Suntiamorntut, he was from 1 April 1998 - 30 June 1999 Researcher at Embedded System Lab, Computer Engineering Dept. KMITL 1 April 1998 – 2000 Master.Eng(com) at Chulalongkron University. 1 August 1999 – Present, Lecturer, Computer Engineering Dept. Prince of Songkla University. 1 January 2002 was as Ph.D. at University of Manchester, "Low-Power Asynchronous Digital Signal Processor. Now, she is Department Head and Associate Department Head for Student Affairs at Prince of Songkla University. She is interesting research in Design and Verification Microprocessor Using VHDL on FPGA, Testing and Verification, Asynchronous Design and Low-Power Circuit Design, etc.