Asian Journal of Information Technology 12 (1): 1-6, 2013
ISSN: 1682-3915

Forecasting the General Examination Results Using Back Propagation

Nik Haslinda Abdul Halim and Salwani Abdullah
School of Computer Science, Faculty of Information Science and Technology,
University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

Abstract: The Penilaian Menengah Rendah (PMR) examination is one of the government's public examinations taken in form 3 by every school pupil during the duration of their secondary school education. PMR's result forecasting is necessary and it provides a crucial element for the educators to facilitate the classroom management after the pupil completes their form 3 studies. It is the usual practices of the educators to use the results of the tests and examinations to forecast PMR results, regardless of the pupil's background. The purpose of this study is to forecast the public examination results using the back propagation approach. This methodology is employed to develop a Neural Network Model in order to determine the forecast accuracy based on the input parameters from three factors, i.e., demographic, academic and a combination of demographic and academic factors. The performance is measured in terms of the accuracy of the forecast based on the denormalization of the computed results. The dataset used for this study is from the Sekolah Menengah Kebangsaan Puchong Utama (1) (SMKPU1).

Key words: Forecasting, back propagation, neural network, school, test, dataset

INTRODUCTION

Currently, all schools forecast the pupil performance based only on the results of the examinations and tests according to headcount process that was introduced by the Ministry of Education. It is a process to identify academic and career potential of each pupil. According to Hjh. Salmiah Mohd Salleh, the recipient of Kedah's 2003 Innovative Teacher award, Headcount process is a process to identify, rationalize and justify the determined learning capabilities characteristics of the pupil, aiding the teachers to project the performance targets of the subjects taught in the class. Initially, teachers will identify a measure called Take-Off Value (TOV) for each pupil, based on their final examination results for examples, form 2 final examination results for PMR pupils and form 4 final examination results for SPM pupils. Next, the Targeted Expected Result (ETR) is determined. ETR refers to the best public examination results at the school level. Over the course of the examination year, the Targeted Operational Increment (OTI) is carried out by evaluating whether the expected increase can be achieved by two or three assessments. Various academic activities are carried out during and outside of schooling hours. Finally, when the Actual Results (AR) for PMR and Sijil Pelajaran Malaysia (SPM) are announced, the comparison between ETR and AR is performed to determine the accuracy of the process.

The PMR achievement of SMKPU1 up till now is not as expected because the forecasting of the pupils' results relies only on the academic factor. Other factor such as demographic (background) of the pupils might also play a role in determining the PMR results prognosis of the school. The objective of this study is to realize whether there is any correlation in the demographic data that influenced the forecasted results.

Earlier neural network studies include the construction and comparisons of models with the linear regression method to make a forecast of the educational expenses of primary and secondary schools pupils in the United Kingdom and also the terraced houses price forecasting in Kuala Lumpur (Ku Ruhana *et al*., 1999). The results have indicated that the multi-level Neural Network Models constructed are suitable for use in forecasting the outcomes seek. This study will use a back propagation classification within the Neural Network Models adopted for examination results forecasting.

To get a good forecast value using the back propagation model, the selection of the network structure, especially the number of hidden nodes and suitability of the selected parameters such as learning (α) and

Corresponding Author: Nik Haslinda Abdul Halim, School of Computer Science, Faculty of Information Science and Technology, University Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

momentum (β) constants need closer attention. If the number of the hidden nodes used is less than the number of input nodes then α with a value 0.5 is sufficient to yield a relatively good estimate of the forecast. However, if the hidden nodes used either equal or exceed the number of the input nodes then larger value of α is required. Training and testing operations of the back propagation method will be faster if the data is in a matrix structure. The number of training data will greatly affects the time to attain the minimum target error.

There are many forecasting models produced that employ the back propagation classification for examples, the shares price prediction using artificial neural networks and the project performance monitoring and prediction package of Traditional General Contract (TGC) (Zin *et al*., 2006). The study of the aerodynamic characteristics forecasting using the neural networks has shown that the forecast using neural network with back propagation classification returns more accurate results than using neural network with radial basis function (Malmathanraj and Selvi, 2008). Experimental data from the wind tunnel tests included information of alpha, frictional drag and mach number (comparison of speed to the speed of sound), ranging between 0.6-3.

MATERIALS AND METHODS

Data procuration and preparation

Data sources: The data for this study were obtained from the Ministry of Education, Department of Schools Student Information System (SIS), SMKPU1 Internal Examination Coordinator and SMKPU1 PMR Coordinator as tabulated in Table 1. From Table 1, it can be seen that there are three types of data used spanning the years from 2006 to 2008:

- Demographic data; students' personal information such as gender, race, religion, number of siblings, residing location, family income, father's and mother's employment
- Academic data; test 1 results, mid-year, pre-trial and PMR trial examinations
- Actual PMR results

Data analysis: Data shown in Table 1 are analyzed and divided into three segments which are classified as the training, validation and testing segments. All the segments are made up of three types of survey data that have different input parameters but similar output as shown in Table 2.

Pre-processing: Prior to the classification and forecasting simulation, appropriate data pre-processing activities such as cleaning, transformation and reduction of the available data (Zin *et al*., 2006) are meticulously carried out.

Table 1: Sources of data for the study

Sources	Data	Years
Student information system	Personal information	2006 to 2008
Internal examination coordinator	First test	2006 to 2008
	Mid-year examination	
	Pre-trial	
	PMR trial	
PMR coordinator	PMR results	2006 to 2008

Table 2: Distribution of the data studied

Segments	Parameter (input)	Output
Training	Demographic data	PMR 2006 results
	Academic data	
	Combination of demographic and academic data	
Validation	Demographic data	PMR 2007 results
	Academic data	
	Combination of demographic and academic data	
Testing	Demographic data	PMR 2008 results
	Academic data	
	Combination of demographic and academic data	

In this research, data cleaning involves the process of removing incomplete data according to the pre-determined parameters. This is due to the fact that there are records of pupils who enrolled in the school but did not sit for the examination and there are those who have sat for the examination at the school but did not fully complete their personal information. Apart from removing the incomplete data those which contain minor errors are corrected by figuring out the available parameters in the information.

There are a variety of data transformation techniques which are available and commonly used such as min-max normalization, z-score normalization and decimal scaling (Han and Kamber, 2006). Before the data transformation process can be done, the data on each parameter are categorized according to numbers. Usually in categorizing the parameters, the numbers used are from 1-5 (Zin *et al*., 2006).

In this study, researchers have opted to use the min-max normalization technique. The data is converted by the technique into smaller values ranging from 0-1.0. This method of representing the data into smaller values will facilitate the speeding up of the training processes and also preventing the parameters with larger values from weighing down the parameters with smaller values during the welding processes (Zin *et al*., 2006). The min-max normalization technique equation is as follows:

$$V' = \left(\frac{V\text{-}min_a}{max_a\text{-}min_a}\right)\times(new\_max_a\_new\_min_a)+min_a \quad (1)$$

Where:

a = The original value of the data
new = The data new value

Table 3: Categories of the data parameters studied

Parameters	Catagories
Race	
Malay	1
Chinesse	2
Indians	3
Others	4
Religion	
Islam	1
Others	2
Gender	
Male	1
Female	2
Occupation	
Professional	1
Manufacturing	2
Services	3
Others	4
Unemployed	5
Number of siblings	
1 and 2	1
3 and 4	2
5 and 6	3
7 and above	4
Total income	
≤500	1
500 to ≥1000	2
1000 to ≥2000	3
2000 to ≥4000	4
>4000	5
Residency	
Bungalow	1
Condominium	2
Terrace	3
Flat	4
Village house	5
PMR result	
8A, 7A, 6A, 5A (Excellent)	1
4A, 3A (Satisfactory)	2
2A and below (Good)	3
Have E (Fail)	4

The value of the parameter with min_a and max_a ranges can then be transformed to a new value V'. From Eq. 1, the new values can only be in the range from 0-1.0. Table 3 shows the category of each parameter of the data studied represented by numbers.

Modeling: The prepared data are modeled using neural networks algorithm. The modeling process produces many models resulting from three different types of data parameters introduced. Three models which are considered to be the best are then selected and later reduced to only one model for the final work. The selection of only a single model out of the three uses the trial and error method to establish which model provides results with the minimum Mean Square Error (MSE). As elaborated by Al-Shalabi *et al.* (2006) the most appropriate method to determine the optimal design is still through trial and error method. The neural network toolbox or NNTool which is one of MATLAB's (Matrix Laboratory) toolboxes is used to find and establish the best Neural Network Model.

Only one hidden level is used in this study for the modeling of the problem since the performance of the more accurate model that relies on more than one hidden level does not guarantee good results (Zin *et al.*, 2006). On the contrary, the model with only one hidden level normally gives better results. For the choice of the number of nodes in the hidden level if only one hidden level is used the number of hidden nodes is recommended not to be >75% of the number of input nodes (Bailey and Thompson, 1990).

RESULTS AND DISCUSSION

After the networks have been trained and tested, the output will be restore back to the original value to produce the forecast value. This restoring process uses the denormalization equation as follows (Khare and Agarwol, 1997):

$$Y_{denormalization} = (Y_k - Y_{min})(Y_{max} - Y_{min}) + Y_{min} \quad (2)$$

where, Y_k is the value for the denormalization. The forecast values obtained are then compared with the actual value to evaluate the ability of the back propagation model acquired from the modeling process.

All the networks with different parameters are trained using the Levenberg-Marquardt backpropagation (TRAINLM) learning algorithm with the learning rate of 0.1 and a momentum constant of 0.9. In this research, when different transfer function is used the nodes in the hidden layer are also different. The three transfer functionsof the neural network toolbox used are TANSIG (hyperbolic tangent sigmoid transfer function), PURELIN (linear transfer function) and LOGSIG (log-sigmoid transfer function).

Parameters based on the demographic factor: In this study, the PMR results forecasting using only the demographic factor depends upon eight demographic parameters such as gender, race, religion, number of siblings, residency, total family income, father's and mother's occupations. The Neural Network Model built during the simulation with the least Mean Square Error (MSE) will be regarded as the best model. The number of nodes in the hidden layer can either be 2, 3, 4, 5 or 6 nodes while the output can has only 1 node.

A total of 9 models were constructed for each of the Neural Network Model structure to obtain the smallest

MSE. The configurations of the Neural Network Model structure constructed are 8-2-1, 8-3-1, 8-4-1, 8-5-1 and 8-6-1. The weightages and biases resulting from the models generated are recorded, especially from the ones with the smallest MSE, indicating their significance as the best models produced by the simulations.

The results of the experiments performed to find the model that fits the best of the best model's criteria, based on the number of different nodes in the hidden layer in the PMR results forecasting that relies only on the demographic factor are tabulated in Table 4. From Table 4, it can be deduced that the best Neural Network Model using back propagation algorithm based on demographic factor and its parameters is the model with MSE score of 0.083 and the prediction accuracy of 91.67%. The model structure is 8-6-1, meaning there are 8 input parameters and 6 nodes in the hidden layer with LOGSIG as the activation function in both the hidden and output layers.

The output value of the best model chosen will be transformed back to the original value using Eq. 2 to produce the forecast value. Graphically shown in Fig. 1 is the comparison between the forecast PMR's results using only the demographic factor in the input parameter and the actual PMR's results based on the categories tabulated in Table 3.

Table 4: Summary of the experimental results based on the demographic factor

	Hidden layer		Output layer			
Model structure	No. of nodes	Activation function	No. of nodes	Activation function	MSE	Accuracy (%)
8-2-1	2	TANSIG	1	TANSIG	0.085	91.51
8-3-1	3	TANSIG	1	TANSIG	0.087	91.35
8-4-1	4	LOGSIG	1	PURELIN	0.085	91.50
8-5-1	5	LOGSIG	1	PURELIN	0.087	91.26
8-6-1	6	LOGSIG	1	LOGSIG	0.083	91.67

Parameters based on the academic factor: In relying only upon the academic factor, PMR results forecast depends on four parameters as inputs which are the first test, mid year, pre-trial and trial examinations. Similar to the forecast based on the demographic factor, Neural Network Model that produces the least MSE during the simulation will be considered as the best model. The number of nodes used in the hidden layer can either be 2 or 3 nodes, whereas the output has only 1 node.

A total of 9 models were constructed for each of the Neural Network Model structure to obtain the smallest SE. The configurations of the Neural Network Model structure constructed are 4-2-1 and 4-3-1. The weightages and biases resulting from the models generated are recorded, especially from the ones with the smallest MSE, indicating their significance as the best models produced by the simulations.

Table 5 shows the results of the experiments performed to find the best model from the best models selected. The models are based on the number of different nodes in the hidden layer. It can be infer from Table 5 that the best model for neural networks using back propagation algorithm based on the academic factor has MSE of 0.060 with the percentage accuracy of 94.01%. The best model has a 4-3-1 structure, meaning four input parameters, 3 nodes in the hidden layer and 1 node in the output layer with both layers activated by the function LOGSIG.

Table 5: Summary of the experimental results based on the academic factor

	Hidden layer		Output layer			
Model structure	No. of nodes	Activation function	No. of nodes	Activation function	MSE	Accuracy (%)
4-2-1	2	TANSIG	1	TANSIG	0.069	93.09
4-3-1	3	LOGSIG	1	LOGSIG	0.060	94.01

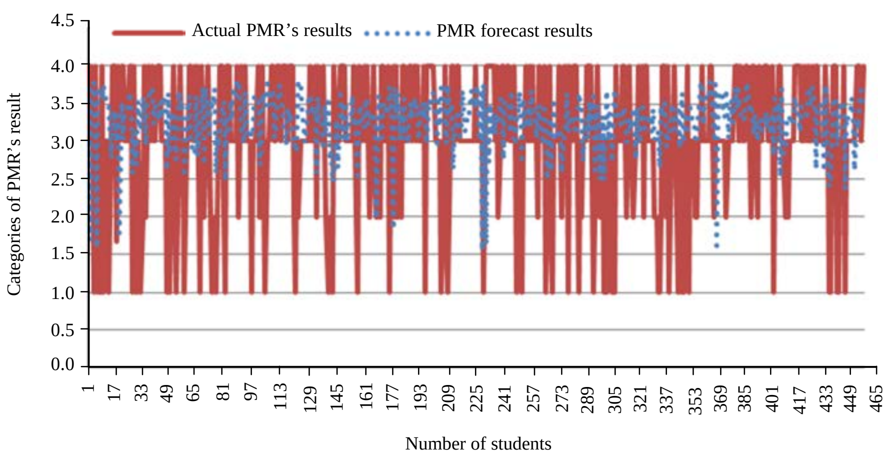

Fig. 1: Comparison between the PMR forecast results based on the demographic factor and the actual PMR's results

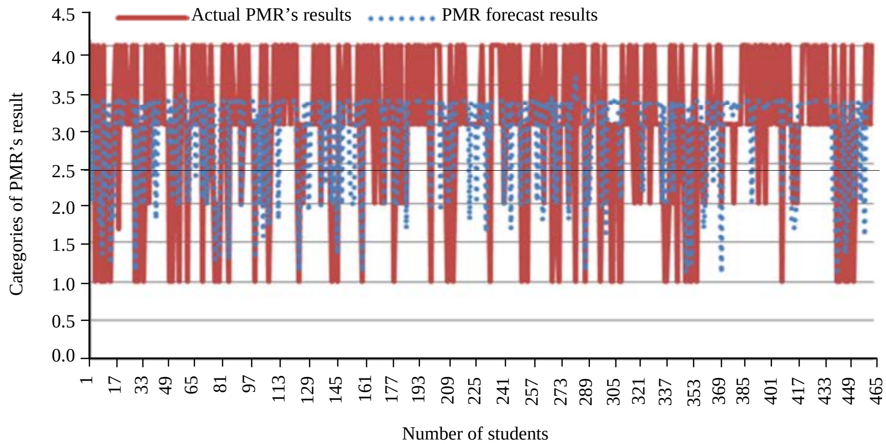

Fig. 2: Comparison between the PMR forecast results based on the academic factor and the actual PMR's results

Table 6: The best neural network models built with the specific input parameters from the different factors considered

	Hidden layer		Output layer			
Model structure	No. of nodes	Activation function	No. of nodes	Activation function	MSE	Accuracy (%)
8-6-1	6	LOGSIG	1	LOGSIG	0.083	91.67
4-3-1	3	LOGSIG	1	LOGSIG	0.060	94.01
12-5-1	5	TANSIG	1	LOGSIG	0.106	89.37

The generated output value from the best model selected will be transformed back to the original value by Eq. 2 for the forecast. Figure 2 shows the comparison graph between the forecast PMR's results using only the academic factor as the condition to the input parameter and the actual PMR's results based on the categories tabulated in Table 6.

Parameters based on the combination of the demographic and academic factors: In forecasting the PMR results based upon the combination of the demographic and academic factors, there are 12 input parameters such as gender, race religion, number of siblings, residency, total family income, father's and mother's occupation, first test, mid year, pre-trial and trial examinations to consider. As in the study, the Neural Network Model built by the simulation that produces the least MSE will be regarded as the best model. The number of nodes in the hidden layer can either be 2, 3, 4, 5, 6, 7, 8 or 9 nodes while the output can has only 1 node.

A total of 9 models were constructed for each of the Neural Network Model structure to obtain the smallest MSE. The configurations of the Neural Network Model structure constructed are 12-2-1, 12-3-3, 12-4-1, 12-5-1, 12-6-1, 12-7-1, 12-8-1 and 12-9-1. The weightages and biases resulting from the models generated are recorded, especially from the ones with the smallest MSE indicating their significance as the best models produced by the simulations.

Table 7: Summary of experimental results for the combination of the demographic and academic factors

	Hidden layer		Output layer			
Model structure	No. of nodes	Activation function	No. of nodes	Activation function	MSE	Accuracy (%)
12-2-1	2	PURELIN	1	LOGSIG	0.426	57.43
12-3-1	3	PURELIN	1	LOGSIG	0.146	85.38
12-4-1	4	LOGSIG	1	TANSIG	0.137	86.25
12-5-1	5	TANSIG	1	LOGSIG	0.106	89.37
12-6-1	6	TANSIG	1	TANSIG	0.113	88.68
12-7-1	7	LOGSIG	1	TANSIG	0.120	88.00
12-8-1	8	PURELIN	1	LOGSIG	0.115	88.48
12-9-1	9	LOGSIG	1	PURELIN	0.163	83.69

The results of the experiments performed to find the model that fits the best of the best model's criteria, based on the number of different nodes in the hidden layer in the PMR results forecasting that considers both the demographic and academic factors are tabulated in Table 7.

From Table 7, it can be concluded that the best model for neural networks using the back propagation algorithm based on the combination of the demographic and academic factors is the model with MSE of 0.106 and percentage accuracy of 89.37%. The best model has a 12-5-1 structure, meaning there are twelve input parameters, 3 nodes in the hidden layer and 1 node in the output layer. In this case, the layers are activated by different activation functions, TANSIG for the hidden layer LOGSIG for the output layer.

The generated output value from the best model selected will be transformed back to the original value by Eq. 2 for the forecast. Figure 3 shows the comparison graph between the forecast PMR's results using both the demographic and academic factors as the criterion to the input parameters and the actual PMR's results based on the categories tabulated in Table 6.

The modeling experiments performed using the back propagation algorithm has resulted in the construction of

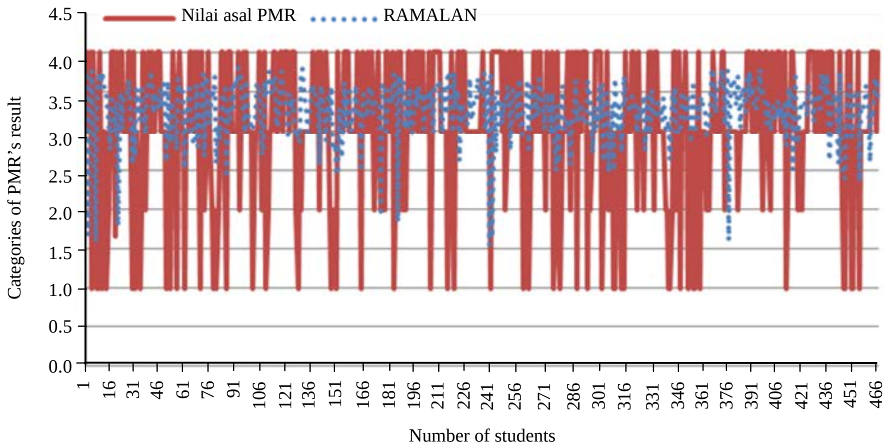

Fig. 3: Comparison between the PMR forecast results based on the combined demographic and academic factors and the actual PMR's results

three Neural Network Models selected to be the best models. All three models contain the number of input parameters that are very dependent on the factors selected. Table 6 lists the Neural Network Models chosen to be the best from the simulations done with regards to the factors selected.

From Table 7, it is observed that the model with the structure of 4-3-1 and input parameters only from the academic factor is the best model with MSE of 0.060 and the percentage accuracy of 94.01%. When compared with the model with structure of 8-6-1 resulting from the simulation that only considers demographic factor which yields MSE of 0.083 and the percentage accuracy of 91.67%, the difference is not that altogether large enough to disregard the influence of the demographic factor. The other resulting model with the structure of 12-5-1 which considered the combination of both factors, its performance hint at the inadequacies of the big number of input parameters compared to smaller number of nodes. The models listed in Table 6 also indicate that the academic factor is the major factor in influencing and affecting the accuracy of the forecasting for SMKPU1 PMR results.

CONCLUSION

Experimental results show that the academic factor obtained better results with an accuracy of 94.01% and the mean square error of 0.060 compared to the demographic and combination of the demographic and academic factors.

REFERENCES

Al-Shalabi, L., Z. Shaaban and B. Kasasbeh, 2006. Data mining: Preprocessing engine. J. Comput. Sci., 2: 735-739.

Bailey, D. and D. Thompson, 1990. How to develop neural-network application. AI Expert, 5: 38-47.

Han, J. and M. Kamber, 2006. Data Mining: Concepts and Techniques. 2nd Edn., Morgan Kaufmann Publisher, San Fransisco, USA., ISBN: 1-55860-901-6.

Khare, M.K. and A. Agarwol, 1997. Development of neural network software for the optimisation of cutting speeds and feed in single pass turning. Proceedings of 17th All India Manufacturing Technology Design and Research Conference, January 9-11, 1997, Warangal, India, pp: 219-227.

Ku Ruhana, K.M., A. Abu Bakar and M.N. Norwawi, 1999. Multi layer perceptron modelling in the housing market. Malaysian Manage. J., 3: 61-69.

Malmathanraj, R. and S.T. Selvi, 2008. Prediction of aerodynamic characteristics using neural networks. Asia J. Inform. Tecnol., 7: 19-26.

Zin, R., S.A. Mansur, A. Bakri and T.C.L. Caren, 2006. Predicting the performance of traditional general contract projects: A neural network based approach. Proceedings of the 6th Asia-Pacific Structural Engineering and Construction Conference, September 5-6, 2006, Kuala Lumpur, Malaysia.