Trip attraction model using radial basis function neural networks

Joni Arliansyah\textsuperscript{a,}*\textsuperscript{,} Yusuf Hartono\textsuperscript{b}

\textsuperscript{a}Department Of Civil Engineering, University of Sriwijaya Inderalaya Sumatera Selatan 30662 Indonesia
\textsuperscript{b}Department Of Math Education, University of Sriwijaya Inderalaya Sumatera Selatan 30662 Indonesia

Abstract

Trip Attraction model with seven independent variables, i.e., population size, number of schools, number of students, number of teachers, areas of school buildings, number of offices, and number of houses applying Radial Basis Function Neural Networks (RBFNN) is presented in this paper. The data used in this study were derived from the origin destination survey in Palembang and the model was developed using 85 sets of land use - trip attraction data. A comparison was made between RBF model and regression model. The results show that RBF model performs better than regression model in predicting trip attraction and important variables are number of students, number of teachers, total areas of school buildings and number of offices.

Keywords: Trip Attraction, Prediction, Radial Basis Function Neural Networks

1. Introduction

In city transportation modeling, trip attraction model is very important to predict the attraction of the trip in the future, which in turn is used to plan the need for city transportation facilities and infrastructures.

The trip attraction modeling in a city can be implemented by connecting the trip attraction derived from the origin destination survey result and land use parameters in each determined zone. In Palembang the origin-destination matrix has been used in urban transportation model for the development of urban transportation facilities and infrastructures [1, 2]. An improved accuracy of trip generating model is essential to get a better result of trip predictions.

* Corresponding author. Tel.: +062 813 6755 3790; fax: +062 711 362 388.
E-mail address: joniarliansyah@yahoo.com
The model development of trip generation and trip attraction in Indonesia has been widely reported, particularly using regression models [3, 4, 5, 6, 7, 8]. The trip generation model using Artificial Neural Network (ANN) with Back Propagation Learning Algorithm has also been reported [9].

This paper discusses the RBFNN application in modeling trip attraction in Palembang, given the shortage of ANN models which take a long time to achieve convergent condition and can be trapped in minimum local condition in selecting the optimal criteria during learning procedure of the network [10]. Contrary to ANN, the RBFNN requires swift time to reach the convergent condition and ensures global convergent conditions. The RBFNN model has also been successful in the field of engineering applications [11, 12, 13, 14, 15].

The aims of the study are: (1) modeling the trip attraction in Palembang by using the Radial Basis Function Neural Network, and (2) comparing the RBFNN modeling results with the regression analysis model.

2. Methodology

In this study the trip attraction model in Palembang was developed using the origin destination matrix and data of land use in Palembang in 2009 and the model was developed using RBFNN and regression analysis. The data used are shown in Appendix A. The results of the attraction model with RBFNN were then compared with the results of the regression analysis model. The methodology and parts of RBFNN models, variables and analysis used are described in the following discussion.

2.1. Radial Basis Function Neural Networks

2.1.1. The topology of RBFNN

A RBFNN is a feedforward neural network that consists of three layers: input layer, hidden layer and output layer. Fig. 1 shows a typical architecture of a RBFNN. In the topology of networks, a RBFNN is similar to a special case of multilayer feedforward neural networks, but different in terms of node characteristics and learning algorithm.

There is no calculation in input layer nodes. The input layer nodes only pass the input data to the hidden layer. The input layer consist of \( n_s \) nodes where input vector \( x = (x_1, x_2, ..., x_{n_s}) \). The hidden layer consists of \( n \) nodes and each hidden node \( j = 1, 2, ..., n \) has a center value \( c_j \). Each hidden layer node performs a nonlinear transformation of the input data onto new space through the radial basis function. The most common choice for the radial basis function is a Gaussian function, given by:

\[
\phi_j(x) = \exp\left(-\frac{\|x - c_j\|^2}{2\sigma_j^2}\right)
\]  

(1)
where \( \| x - c_j \| \) represents the Euclidean distance between input vector \( (x) \) and the radial basis function center \( (c_j) \). 

\[ r_j \] is the width of radial basis function.

The output layer operation is linear, given by

\[ y(x) = \sum_{j=1}^{n} w_{j} \phi_{j}(x) \]  

(2)

where \( w_{j} \) are the connection weight of hidden layer to output layer and \( n \) is number of hidden node.

Since the RBFNN output is a simple linear combination, the parameter solution can be obtained using linear optimization methods. Therefore, it has fast convergence time and is guaranteed to converge to global optimum parameter. Moody et al. [16] demonstrated that the radial basis function networks learn faster than multilayer perceptron network. Park et al. [17] proved theoretically that radial basis function network are capable of universal approximation and learning without local minima, therefore it is guaranteed to converge to global optimum parameter.

Training of RBFNN involves determination of the following parameters.

- Number of hidden layer nodes.
- The center and the width of each radial basis function in each node.
- The connection weight of hidden layer to output layer.

2.1.2. Training Methodology

The orthogonal least squares (OLS) learning algorithm [10] is usually used to determine the center and the optimum number of hidden nodes. The OLS algorithm is operating in a forward selection manner. The procedure chooses the radial basis function center one by one in a rational way until an adequate network has been constructed. Once the optimum numbers of the hidden nodes and their centers are found, the connection weights can be determined.

2.2. Variables and Analysis

The dependent variable in this research is trip attraction defined as the number of vehicles (cars, motorcycles, bicycles) entering a region observed during morning rush hours. Observations were made in 85 kelurahan in Palembang. Seven independent variables used as predictors were:

- number of residents registered in the kelurahan \( (x_1) \)
- number of schools and universities in the kelurahan \( (x_2) \)
- number of students registered in schools or universities in the kelurahan \( (x_3) \)
- number of teachers teaching at the schools or universities in the kelurahan \( (x_4) \)
- number of offices in the kelurahan \( (x_5) \)
- total areas of school and university buildings in the kelurahan \( (x_6) \)
- number of houses in the kelurahan \( (x_7) \)

Data were collected from local central bureau of statistics. Table 1 displays descriptive statistics of all variables used in this research.
Table 1. Descriptive statistics

|             | N   | Minimum | Maximum | Mean   | Std. Deviation |
|-------------|-----|---------|---------|--------|----------------|
| residents   | 85  | 2228    | 46031   | 14664.72 | 7508.579       |
| schools     | 85  | 1       | 19      | 5.87   | 4.056          |
| students    | 85  | 300     | 28890   | 3213.98 | 3709.886       |
| teachers    | 85  | 12      | 1709    | 219.07 | 255.812        |
| school areas| 85  | 70      | 82961   | 10666.18 | 15811.643     |
| offices     | 85  | 1       | 26      | 3.00   | 4.006          |
| houses      | 85  | 12      | 7391    | 2587.75 | 1648.396       |
| trip        | 85  | 35      | 9400    | 1315.66 | 1336.249       |
| Valid N     | 85  |         |         |        |                |

In analysis, comparison between ordinary regression and radial basis function (RBF) models were made using the SPSS statistical package. Stepwise procedure was used to develop regression model based on ordinary least squares approach. Data were standardized using the following equation:

$$Z = \frac{X - \mu}{\sigma}$$  \hspace{1cm} (3)

where $X$ is observed value, $\mu$ is the mean, and $\sigma$ is the standard deviation. This standardization process yields variable $Z$ with zero mean and unit variance.

3. Results and Discussion

Regression analysis gives coefficients of all seven predictors as given in Table 2. The most significant predictors ($p<0.10$) are number of students ($x_3$), teachers ($x_4$), and offices ($x_5$), and total areas of schools ($x_6$). It is clear why stepwise procedure gives the following equation:

$$Y = 629.18 + 0.522X_3 - 4.518X_4$$ \hspace{1cm} (4)

with $R^2 = 0.414$ ($F = 28.98, p = 0.000$). This means number of students and number of teachers alone can explain 41.4% of variation in trip attraction. Inclusion of $X_5$ (number of offices) in the model increases $R^2$ only by 0.02. Notice that regressing trip attraction over all seven predictors gives $R^2 = 0.459$, which means the other five predictors contribute less than 5% in the model.

Table 2 Coefficients of predictors in regression analysis

| Model | Unstandardized Coefficients | Standardized Coefficients |
|-------|-----------------------------|---------------------------|
|       | B   | Std. Error | Beta | t   | Sig. |
| 1     | (Constant) | 427.576    | 269.849  |       |      |
|       | residents | -.001      | .027    | -.008 | -.052 | .959 |
|       | schools   | 5.702      | 42.026   | .017  | .136  | .892 |
|       | students  | .500       | .138     | 1.389 | 3.627 | .001 |
|       | teachers  | -3.679     | 1.998    | -.704 | -1.841 | .069 |
|       | school areas | -.018  | .010    | -.215 | -1.857 | .067 |
|       | offices   | 57.438     | 29.414   | .172  | 1.953 | .054 |
|       | houses    | .037       | .107     | .046  | .346  | .730 |

Dependent Variable: trip

For radial basis function (RBF) analysis, 50 cases (58.8%) were assigned to the training sample, 24 cases (28.2%) to the testing sample, and 11 cases (12.9%) to holdout sample. Network information is given in Table 3.
Table 3. RBF network information

| Input Layer | Covariates | 1 | residents |
|-------------|------------|---|-----------|
|             | 2          | schools |
|             | 3          | students |
|             | 4          | teachers |
|             | 5          | school areas |
|             | 6          | offices |
|             | 7          | houses |
| Number of Units | 7          |
| Rescaling Method for Covariates | Standardized |

| Hidden Layer | Number of Units | 5a |
|--------------|-----------------|---|
| Activation Function | Softmax |

| Output Layer | Dependent Variables | 1 | trip |
|--------------|---------------------|---|------|
| Number of Units | 1 |
| Rescaling Method for Scale Dependents | Standardized |
| Activation Function | Identity |
| Error Function | Sum of Squares |

*a. Determined by the testing data criterion: The "best" number of hidden units is the one that yields the smallest error in the testing data.*

Table 4. RBF parameter estimates

| Predictor | Predicted | Output Layer |
|-----------|-----------|--------------|
|           | H(1) | H(2) | H(3) | H(4) | H(5) | y |
| Input Layer | x1 | 4.048 | -529 | .863 | .296 | .176 |
|             | x2 | 3.393 | -354 | .409 | 1.364 | -1.57 |
|             | x3 | 5.623 | -294 | -067 | 2.270 | -0.55 |
|             | x4 | 4.753 | -335 | .004 | 2.862 | -0.52 |
|             | x5 | 2.304 | -372 | -020 | 3.628 | .738 |
|             | x6 | -281 | -254 | -183 | .038 | 3.480 |
|             | x7 | 1.990 | -596 | 1.050 | .500 | .611 |
| Hidden Unit Width | .596 | .596 | .834 | 1.114 | 1.039 |
| Hidden Layer   | H(1) | 5.492 |
|                | H(2) | -238 |
|                | H(3) | .018 |
|                | H(4) | -483 |
|                | H(5) | .297 |

*a. Displays the center vector for each hidden unit.*

The RBF network structure has five units of hidden layers as displayed in Fig. 2 and Table 4 gives RBF parameter estimates. Furthermore, Table 5 displays RBF model summary. It seems there were fewer errors in training than that in testing and holding samples. This must be due to the smaller sizes of testing and holdout samples. However, relative errors in testing and holdout samples are quite consistent indicating the RBF model can be used in future with high consistency.

Table 5. RBF model summary

| Training | Sum of Squares Error | 8.434 |
|----------|----------------------|------|
| Relative Error | 344 |
| Training Time | 00:00:00.078 |
| Testing | Sum of Squares Error | 5.824* |
| Relative Error | .906 |
| Holdout | Relative Error | .883 |

*a. The number of hidden units is determined by the testing data criterion: The "best" number of hidden units is the one that yields the smallest error in the testing data.*
Since dependent variable was rescaled by standardization, it has zero mean and unit variance, and, hence, sum of squares total is just equal to \( n - 1 \), where \( n \) is the sample size. Corresponding coefficient of determination \( R^2 \) can be easily computed as follows. For testing, we have

\[
R^2 = 1 - \frac{0.424}{49} = 0.828
\]

and for training

\[
R^2 = 1 - \frac{5.824}{23} = 0.747
\]

This shows that RBF performs better than ordinary regression in terms of the amount of variation in dependent variable explained by the model. This is due to flexibility of RBF model.

Finally, from analysis of independent variable importance, see Fig. 3, it is clear that number of students (\( x_3 \)), number of teachers (\( x_4 \)), total areas of school and university buildings (\( x_6 \)) and number of offices (\( x_5 \)) are the most important predictors. This agrees with output of stepwise procedure in regression analysis.

![Fig 2. RBF Network Structure](image)

![Fig. 3. Importance analysis of predictors](image)

4. Conclusions

The radial basis function neural network model trained to predict trip attraction using seven predictor performs better than ordinary regression model using least square approach. Violation of assumptions for regression model, such as normality of error items and linearity, must be one of the reasons for worse performance of regression model. However, the results from both models show that number of students, number of teachers, total areas of school buildings and number of offices are the most significant predictors for trip attraction.

### Appendix A. The Data Used in Analysis

| No. | X1  | X2  | X3  | X4  | X5  | X6  | X7  | Y    |
|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 1   | 3944| 6   | 2919| 109 | 8713.00| 3   | 906 | 2692 |
| 2   | 9035| 5   | 1326| 66  | 10406.00| 1   | 271 | 812  |
| 3   | 16589| 7   | 1388| 31  | 4414.00| 2   | 1630| 2317 |
| 4   | 19079| 12  | 4720| 228 | 28000.00| 9   | 371 | 2971 |
| 5   | 11155| 5   | 2140| 128 | 6760.00| 1   | 548 | 211  |
| 6   | 7747| 11  | 3817| 267 | 17200.00| 3   | 1754| 1283 |
| 7   | 10474| 9   | 2271| 109 | 10900.00| 1   | 1256| 156  |
| 8   | 12315| 4   | 436 | 28  | 22000.00| 2   | 583 | 372  |
| 9   | 16991| 7   | 1668| 135 | 70000.00| 1   | 2637| 332  |
| 10  | 46131| 19  | 28090| 1309 | 45972.51| 2   | 938 | 5400 |
| 11  | 5518| 7   | 1483| 33  | 12458.68| 1   | 469 | 1060 |

| No. | X1  | X2  | X3  | X4  | X5  | X6  | X7  | Y    |
|-----|-----|-----|-----|-----|-----|-----|-----|------|
| 64  | 11455| 7   | 1584 | 281 | 10980.32| 1   | 287 | 4660 |
| 65  | 14912| 5   | 3521| 210 | 9276.78| 1   | 2751| 1100 |
| 66  | 15348| 3   | 1521| 43  | 1678.69| 1   | 2804| 657  |
| 67  | 15849| 7   | 4055| 245 | 6829.20| 1   | 2124| 1253 |
| 68  | 10210| 7   | 3591| 226 | 1598.77| 1   | 920 | 3574 |
| 69  | 10149| 4   | 2172| 90  | 4800.00| 1   | 3955| 588  |
| 70  | 10397| 8   | 3707| 172 | 4691.00| 1   | 2142| 776  |
| 71  | 2698| 5   | 1092| 103 | 1125.00| 1   | 1379| 1678 |
| 72  | 13707| 7   | 2968| 130 | 3797.00| 1   | 2294| 1271 |
| 73  | 1684| 3   | 1015| 65  | 2790.00| 1   | 1786| 831  |
| 74  | 20271| 11  | 3727| 235 | 15841.00| 1   | 4749| 399  |
