SECOND JOINT CONFERENCE INDONESIA MALAYSIA ON MATHEMATICS AND STATISTICS

Surabaya, East Java, Indonesia
January 11-12, 2007

The Organizing Committee
Certifies that

Brodjol Sutijo
has participated in:
SECOND JOINT CONFERENCE INDONESIA MALAYSIA ON MATHEMATICS AND STATISTICS
as
Contributed Speaker
With paper entitled
Selecting Input Factor for Units of Radial Basis Function Networks

Surabaya, January 12, 2007

Prof. Dr. Maman A. Djaunari
President MSMSSEA

Dr. Muhammad Mashuri
Chairman Organizing Committee
SELECTING INPUT FACTOR FOR UNITS OF RADIAL BASIS FUNCTION NETWORKS

Brodjol Sutijo\(^1\), Subanar\(^2\), Suryo Guritno\(^2\)

\(^1\) Statistics Department, Sepuluh Nopember Institut of Technology, Surabaya
Kampus ITS, Kepuh Sukolilo Surabaya, 60111
Sutijo.b@yahoo.com

\(^2\) Mathematics Department, Gajah Mada University, Yogyakarta
Jl. SKIP Utara Yogyakarta
subanar@yahoo.com, guritno0@gmail.com

Abstract

Among all existing forecasting methods, Radial Basis Function (RBF) networks have been widely used, because it is capable of deducing hidden input-output relationship in data. RBF networks decompose a function into units or clusters of Gaussian functions. The number of units in networks was determined by using \(k\)-mean method. The networks use some input factors to approximate various data features, but there are unimportant input factor existing in some units. The unimportant input factors in a unit will generate unnecessary parameter which may result in situation that a network may represents the underlying data. In this paper mean and variance of each input factors in a unit are found to capture sense of connection input to units in hidden layer. Based on this finding, a method to identify and eliminate unimportant input factor is developed. Implementation of this method is to predict rate of non star hotel occupation in Yogyakarta. The result show that RBF network has four units in hidden layer with six input factors. Mean Square Error (MSE) of this network with full connection is 981.3187. Based on variance test of input factors, this research demonstrate that there are three unimportant links, which represent the deleted candidates. The performance of new network, network without unimportant link, increases with MSE 11.3706.

Keywords: RBF, Cluster, \(k\)-mean, variance test, MSE

1. Introduction

Radial Basis Function (RBF) networks were introduced into the Neural Networks (NN) literature by Broomhead and Lowe (1988). The RBF networks model is motivated by local tuned response observed in biologic neurons. The networks have been subject to intensive research over recent years and have successfully been employed to various problem (Moody and Darken, 1989; Lowe, 1999; Girosi et al. 1995). The network has become one of most popular feedforward NN with applications in regression, classification and function approximation problem (Liver, 1997; Haykin, 1999).

RBF network approximates nonlinear mapping by weight sum of Gaussian kernels. Therefore, an RBF learning algorithm must estimate the centers of units, their variance and weights of output layer. RBF network training usually proceeds in two steps: First, the basis function parameters are determined by clustering. Second, the final layer weight is determined by least square which reduces to solving a simple linear-system. One of the advantages of RBF networks, compare to type of another NN take possibility of choosing suitable parameter for units of hidden layer without having to perform a non linear optimization of the networks parameter. However, the problem select the appropriate number of basis function remain critical issue for RBF network.

Among all existing forecasting method, RBF networks have been widely used, because they are capable of deducing hidden input-output relationship in data. For Gaussian RBF network, it is capable in inherited from property that Gaussian RBF networks decompose function into localized Gaussian function (unit or cluster). Units in RBF network use the same input to approximate various data. However, there are unimportant input factors existing in same units. Unimportant factors in a unit will generate unnecessary adjustable parameter which may result in the situation that a network misrepresent the underlying data. The
The above situation is called over-fitting. The problem is to identify and eliminate insignificant input in each unit.

The criterion to check whether a parameter has a sense of connection depend on whether the associated link can be disconnected if the parameter is set to zero. Centers and standard deviations parameter do not meet the criterion. The inverses standard deviations are being used instead to identify insignificant input factors. The mean of Gaussian function can be viewed as parameter shifting a univariate in the exponential and inverses of standard deviation is viewed as a coefficient weighting the shifting variable. The value of this coefficient being too small implies in significant of shifted variable, conversely the value of the coefficient too large lead to the value or Gaussian function being close to zero, so the Gaussian function becomes insignificant.

II. Radial Basis Function Networks

1. Gaussian Radial Basis Function

The RBF network consists of some of parameters (weights) to be estimate. To find best model of RBF network necessary best combination for number of input variables, number of hidden units, centre and width of each hidden units. This is a central topic in some literature RBF network such as Bishop (1995), Ripley (1996), Fine (1999), Haykin (1999), Reed and Marks II (1999), and Lazaro et al. (2002), Lendase (2005).

The network can be designed to perform a nonlinear mapping from the input space to the hidden space, and a linear mapping from the hidden space to the output space. Accoring theory of multivariable interpolation:

\[ F: \mathbb{R}^N \rightarrow \mathbb{R} \quad \text{hold on:} \]

\[ F(x_i) = d_i, \quad i = 1, 2, 3, \ldots, N. \]

The RBF technique consists of choosing a function that has the following form :

\[ F(x) = \sum_{i=1}^{N} w_i \phi(||x - x_i||) \]  

(2)

Where, \{\phi(||x - x_i||) i=1, 2, 3, \ldots, N\} is a set of \(N\) random (usually nonlinear) functions, known as radial basis functions, and ||.|| represents a norm that is generally Euclidean. Consider the \( \mathbb{R}^n \rightarrow \mathbb{R}^m \) mapping implemented by the model

\[ \hat{y} = f(a_0 \sum_{i} w_i g_i(||x - v_i||)) \]  

(3)

The model (3) describes an RBFNN with inputs from \( \mathbb{R}^n \), e radial basis functions, and output units \( y \). The RBFNN network use \( x \) as input and \( y \) as response with \( v \) and \( z \) as prototype. Figure 1, show RBFNN architecture where,

\[ \phi_i = g_i \left(||x - v_i||\right). \]

Figure 1. RBFNN design
If the interpolation conditions equation (2) is inserted in (1), the following set of simultaneous linear equations can be obtained for the unknown coefficients (weights) of the expansion \( \{w\} \):

\[
\begin{bmatrix}
\phi_{i1} & \phi_{i2} & \cdots & \phi_{iN} \\
\phi_{j1} & \phi_{j2} & \cdots & \phi_{jN} \\
\vdots & \vdots & \ddots & \vdots \\
\phi_{N1} & \phi_{N2} & \cdots & \phi_{NN}
\end{bmatrix}
\begin{bmatrix}
w_1 \\
w_2 \\
\vdots \\
w_N
\end{bmatrix}
=
\begin{bmatrix}
d_1 \\
d_2 \\
\vdots \\
d_N
\end{bmatrix}
\quad (4)
\]

Where, \( \phi_{ij} = \phi(||x_i - x_j||) \) \( i, j = 1, 2, 3, \ldots, N \)

Let \( d = [d_1, d_2, \ldots, d_N] \), \( w = [w_1, w_2, \ldots, w_N] \) and \( \phi = \phi_{ij} \). The matrix \( \phi \) is called interpolation matrix, and equation (4) can be written in the compact form:

\[
\phi \cdot w = d
\quad (5)
\]

If the data points are all distinct, the interpolation matrix is positive definite and weight vector \( w \) can be formed as follows:

\[
w = \phi^{-1} d
\quad (6)
\]

In practice, equation (5) cannot be solved when matrix \( \phi \) is close to singular. Regularization theory can solve this problem by perturbation the matrix \( \phi \) to \( \phi + \lambda I \).

A Gaussian radial basis function \( \phi_j = g_j \left( \|x - v_j\|^2 \right) \) centered at \( v_j \) and with norm weighting matrix \( \Sigma \) may be expressed as:

\[
\phi_j = \exp \left[ -\frac{1}{2} (x - v_j)^\top \Sigma^{-1} (x - v_j) \right]
\quad (7)
\]

2. K-Mean Clustering

For the self-organized learning process, clustering algorithm that partitions the given set of data points into subgroups, each of which should be as homogeneous as possible. The K-mean clustering algorithm proceeds as:

1. Initialization. Choose random values for the initial centre \( v_i(0) \).
2. Sampling. Draw a sample vector \( x \) from input space, it is input of algorithm at iteration \( n \).
3. Similarity matching. Let \( j(x) \) best matching of centre for input \( x \).
   
   \[
   j(x) = \arg \min_{k} \| x(n) - v_k(n) \|_2, \quad k = 1, 2, \ldots, m
   \]
4. Updating. Adjust the centre of RBF:
   
   \[
   v_k(n+1) = v_k(n) + \eta \sum_{x \in C_k} (x - v_k(n))
   \]
5. Continuation. Increment \( n \) by 1, go back to 2, until no change in centre.

3. Analysis on Inverse Standard Deviation

To identify insignificant input factors in a unit by using \( \sigma^{-1}_\eta \), the output of Gaussian RBF can be expressed:

\[
\phi_i = \exp \left[ -\frac{1}{2} (x_i - v_{i\eta}) / \sigma_{i\eta} \right]^2
\quad (8)
\]

When \( \sigma^{-1}_\eta \) is very small, that is \( \sigma_{i\eta} \) very large, it imply wide and flat receptive field that contains no valuable information. Conversely \( \sigma_{i\eta} \) very small, it suggests a sharp and narrow receptive field that cannot capture information beyond a small neighborhood around the function's center. So, we need a threshold. Statistical Hypothesis and confidence interval for sample variance can be used to determined the threshold.
III. Empirical Study

In this study is used rate of non star hotel occupation data in Yogyakarta from January 1991 until December 2003. The data is collected from monthly report of Statistic Central bureau. On modelling, data was divided into two part, training and testing data set. Training data is used to parameterized RBF model and testing data to test the model. Training data consist of 120 data, from January 1991 until December 2000, the remain data is used to testing.

There are two main model in this study. The first, RBF model with input is $y_{t-1}, y_{t-2}, y_{t-3}, y_{t-4}, y_{t-5}$ and $y_{t-6}$. Second, RBF model with input is $y_{t-1}, y_{t-2}, y_{t-3}$ and $y_{t-4}$. First model is assumed there is not seasonal factors and number of input has to be given a priori. Second model included seasonal factors and input has to be given based on value of autocorrelation.

Number of units in hidden layer can be fix by clustering input data. One of cluster methods is K-means clustering. Input data is divided into k cluster by the method. Data have similar characteristics are included into one cluster, but data have dissimilarity on separate cluster. First model has four unit in hidden layer and second model has three unit in hidden layer.

Link of RBF network usually are all input connected to unit in hidden layers. Based on analysis on inverses standard deviation, there are some of link must be disconnected. First model, links of input variables 2 to unit 2, input variables 3 to unit 1 and 3 are disconnected candidates. Second model, all link to unit 2, input variable 4 to unit 3 disconnected candidates.

Table 1 show performances RBF networks: first and second models for full connection and one or more links of input to unit are disconnected on testing data. The performances (MS, MAPE and MAD) are characteristic of forecast value.

Table 1. Performance RBF network for first and second model.

<table>
<thead>
<tr>
<th>Model</th>
<th>Testing</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MS</td>
<td>MAPE</td>
<td>MAD</td>
</tr>
<tr>
<td>1</td>
<td>Full</td>
<td>712.876</td>
<td>0.326</td>
</tr>
<tr>
<td>(3:1)</td>
<td>12.813</td>
<td>0.109</td>
<td>2.645</td>
</tr>
<tr>
<td>(2:2)</td>
<td>12.228</td>
<td>0.117</td>
<td>2.881</td>
</tr>
<tr>
<td>(3:3)</td>
<td>14.530</td>
<td>0.123</td>
<td>3.009</td>
</tr>
<tr>
<td>(2:2)(3:1)</td>
<td>11.816</td>
<td>0.110</td>
<td>2.673</td>
</tr>
<tr>
<td>(3:1)</td>
<td>13.589</td>
<td>0.122</td>
<td>2.976</td>
</tr>
<tr>
<td>(2:2)(3:3)</td>
<td>12.201</td>
<td>0.108</td>
<td>2.644</td>
</tr>
<tr>
<td>(2:2)(3:3,1)</td>
<td>11.883</td>
<td>0.107</td>
<td>2.631</td>
</tr>
<tr>
<td>2</td>
<td>Full</td>
<td>690.778</td>
<td>1.000</td>
</tr>
<tr>
<td>(1:2)</td>
<td>11.913</td>
<td>0.113</td>
<td>2.686</td>
</tr>
<tr>
<td>(2:2)</td>
<td>14.620</td>
<td>0.130</td>
<td>3.097</td>
</tr>
<tr>
<td>(1:2)</td>
<td>15.045</td>
<td>0.129</td>
<td>3.101</td>
</tr>
<tr>
<td>(4:2)</td>
<td>11.913</td>
<td>0.112</td>
<td>2.704</td>
</tr>
<tr>
<td>(4:3)</td>
<td>16.452</td>
<td>0.130</td>
<td>3.196</td>
</tr>
<tr>
<td>(2:3,4:2)</td>
<td>12.272</td>
<td>0.119</td>
<td>2.638</td>
</tr>
<tr>
<td>(1,3,4:2)</td>
<td>11.255</td>
<td>0.107</td>
<td>2.603</td>
</tr>
</tbody>
</table>

Note: model(k:k) is disconnect link input k to unit k.

For model 1, best model is model $(2:2)(3:1)$ or disconnect link input 2 to unit 2 and input 3 to unit 1 based on MS criterion. But, based on MAPE and MAD criterion, best model is model $(2:2)(3:1:3)$ or disconnect link input 2 to unit 2 and input 3 to units 1 and 3. For model 2, minimum of MS, MAPE, MAD are model $(1,3,4:2)$, so best model is model $(1,3,4:2)$ or disconnect link inputs 1,3 and 4 to unit 2. Based on these performance model 1 and model 2, show that model 2 better than model 1. Model 2 is a model that seasonal factor included as input variable.
IV. Conclusion

Among all existing forecasting methods, RBF networks have been widely used, because they are capable of deducting hidden input-output relationship in data. Units in RBF network use the same input to approximate various data. However, there are unimportant input factors existing in same units, that will generate unnecessary adjustable parameter. To identify and eliminate insignificant input in each unit is used analysis inverse of variance.

There are two main model rate of non star hotel occupation data in Yogyakarta in this study. For model 1, best model is model with disconnect link input 2 to unit 2 and input 3 to unit 1 based on MS criterion. But, based on MAPE and MAD criterion, best model is model with disconnect link input 2 to unit 2 and input 3 to units 1 and 3. For model 2, minimum of MS, MAPE MAD are model with disconnect link inputs 1,3 and 4 to unit 2. Based on these performance model 1 and model 2, show that model 2 better than model 1. Model 2 is a model that seasonal factor include as input variable.

V. References


