Radial Basis Function Network for Predicting Impact of Trip Reduction Strategies

Rafael A Perez
Department of Computer Science, University of South Florida
Tampa, FL 33620, Email: perez@csee.usf.edu

Abstract:- This paper describes the implementation of Radial Basis Function (RBF) networks in predicting the impact of trip reduction strategies to increase the average vehicle ridership (AVR) at a worksite. Two types of learning mechanisms were adopted to train the network, fixed selection of centers and supervised selection of centers, the latter yielding better results in terms of the performance measures adopted: the mean squared error (MSE), the correlation between the actual and desired results and the ability of the network to classify the AVR change into a given range. We also compare the performance of the RBF network with that of the Backpropagation network. The RBF network outperforms the Backpropagation network in all the three measures of performance mentioned.
Keywords:- RBF networks, GRBF networks, Backpropagation, Average Vehicle Ridership.

I. Introduction

The leading source of air pollution in most urbanized areas is vehicle emissions. In an effort to curb vehicle emissions, large organizations with 100 or more employees, in areas with pollution problems, have been encouraged to reduce the number of single-occupant vehicles arriving at the worksite. These organizations, in turn, have put in effect a number of different incentive plans to attempt to encourage employees to carpool or to use public transportation to work. Since the number of possible incentive plans is large (more than 50) and since an organization typically implements any combination of these plans, it is very difficult for an organization to predict the effect that its incentive plans will have in actually reducing the number of single occupant vehicles arriving at its worksite. The problem under consideration in this paper, then, involves the creation of a model to predict this effect. The average vehicle ridership or AVR (the ratio of employees to vehicles arriving at the worksite) is used to quantify the effect of incentive plans by an organization. Positive changes in AVR indicate that the incentive plans are reducing the number of single occupant vehicles arriving at the worksite.

As a result of legislation, large employers in some urban areas have been required to submit detailed plans for influencing employee travel behavior. Over a period of time, a large amount of data has been collected from these metropolitan areas. The existence and availability of this data provides an opportunity to develop models that are capable of discovering input-output relationships directly from the data. It is this type of data, collected from three major cities (Los Angeles, Phoenix, and Tucson), that has been used to build the predictive models discussed in this paper.

The large number of variables that may affect the AVR at a given company site as well as the lack of experience in dealing with this relatively new situation reduces the choice of predictive models for this problem. There have been human experts called upon to describe how a specific plan for a company of a given size, at a specific location, with a certain number of employees on each shift, will affect the AVR, but their accuracy has been questionable. Consequently, this lack of human expertise has made it impossible to take an expert system approach to build a software model for this problem. Attempts have already been made to use linear regression analysis and linear discriminant analysis, but reports indicate that these attempts have fallen short of building a successful model. Sigmoidal networks have also been used for this problem and have been found to outperform the statistical methods of linear regression analysis and linear discriminant analysis

[6]. However, there is certainly room for a predictive model to improve on the correlation and the classification capabilities of the sigmoidal networks trained via backpropagation. A possible mechanism for improving this predictive model is that of the Radial Basis Function (RBF) networks [7].

The purpose of this paper is twofold. First, to test the performance of the RBF networks on the AVR prediction problem by employing two means of selection of the centers: fixed selection and supervised selection of centers. The results show that the supervised learning of center locations improves the RBF network performance to a significant extent. Second, to compare the performance of the RBF network on the AVR prediction problem with that of a Backpropagation network. The results confirm that there is a substantial improvement in prediction between RBF network and the Backpropagation network.

In the remainder of the paper, we first review the AVR prediction problem domain and describe our testing methodology. Then, we describe the learning mechanisms adopted in training the RBF networks. Finally, we present the results of our comparison of the two types of RBF networks and their relative performance with that of the Backpropagation network on the same data.

II. Data Description

The factors thought to influence the AVR belonged to the following categories:

- Site characteristics like the number of employees, parking availability, etc.
- Alternative commute incentive plans offered by companies.
- Reported cost levels of employee parking.
- Job classification breakdowns of employees.
- Combination of variables, like the ratio of parking available to number of employees.
- Estimated travel impedances (or times) for employees.

In order to utilize the collected data to build the models, the data underwent a preprocessing procedure. This procedure first included the elimination of redundant and obviously erroneous records. A total of 6982 records remained after this first step. This was followed by eliminating those factors (or independent variables) from the data that were not correlated with the AVR. A total of 45 independent variables remained after this second step. These variables consisted of both quantitative (i.e. number of employees at a worksite)as well as class variables (i.e. whether a particular incentive plan was offered). Quantitative variables were scaled to the range -1 to +1. Class variables were assigned values of 0 or 1.

Each record in the database contained the worksites's AVR, as determined from surveys of employees. By combining sequential (approximately yearly) plan submissions and comparing each sequence's AVR, calculating the change in AVR for each worksite resulting from the implementation of a given combination of plan incentives was possible. This quantitative variable became the dependent variable in all the model-building efforts. The change in AVR, was found to be highly skewed to the middle of the data range available as shown in Table I. The vast majority of the data falls near -0.10 to +0.20 change in AVR. Models built on error minimization criteria may force their predictions to the middle of the range, causing to predict little or no change in AVR. This approach causes the models to have much more accuracy in the middle ranges of AVR change than those with large changes in AVR. In order to get a more comprehensive evaluation of the model, not only must the mean squared error be measured but also the ability of the network to correctly classify the AVR change into a range or category. The ranges were developed by partitioning the data into seven equal sized groups based on the value of the dependent variable. The evaluation procedure will therefore involve the comparison of the correlation values of predicted and actual

change in AVR and the ability of the model to classify an observation into a correct group. The different AVR change range categories used for model evaluation are shown in Table II.

A split of 80 percent-20 percent was maintained for randomly sampling the data into training and testing patterns. Therefore, the number of records used during training is 5682 and the records kept aside for testing is 1300.

TABLE I
DISTRIBUTION OF AVR CHANGE LEVELS.

Change in AVR	Percent of data
-0.2 or less	4.2%
-0.1 to -0.19	6.9%
0.0 to -0.09	30.2%
0.01 to 0.1	40.9%
0.11 to 0.2	12.2%
0.21 to 0.3	3.3%
0.31 or more	2.3%

TABLE II
AVR CHANGE RANGE CATEGORIES FOR MODEL EVALUATION

AVR Change category	AVR Change category range
Large decrease	-0.08 or less
Moderate decrease	-0.03 to -0.079
Small decrease	0 to -0.029
Neutral	0 to 0.029
Small increase	0.03 to 0.059
Moderate increase	0.06 to .119
Large increase	0.12 or more

III. The Basic RBF network structure

Major contributions to the theory, design, and application of RBF networks include [2], [7] and [1]. RBF networks are fully-connected feedforward networks. The network consists of an input layer, a hidden layer and an output layer [4]. The first layer is the input layer which is composed of source nodes whose number is equal to the number of independent variables in the problem. The second layer is the hidden layer, consisting of nonlinear units called the centers, which are fully connected to the input layer. The activation function of each hidden unit is a radial basis function of the form [7]

$$exp\left(-\frac{1}{2\sigma_i^2}\|x - t_i\|^2\right) \tag{1}$$

where x is the input vector of training examples. The radial basis function is a multivariate Gaussian function characterized by the variance σ_i and the centers t_i which are initially a subset

of the data points in the training set. In the fixed selection of centers, the centers t_i remain constant, but in the supervised selection the centers are altered during training. The third layer is the output layer which, in our case, comprises of a single linear unit directly connected to each of the hidden units. The transformation from the input space to the hidden-unit space is nonlinear, whereas the transformation from the hidden-unit space to the output space is linear. The RBF network is shown in Figure 1 .

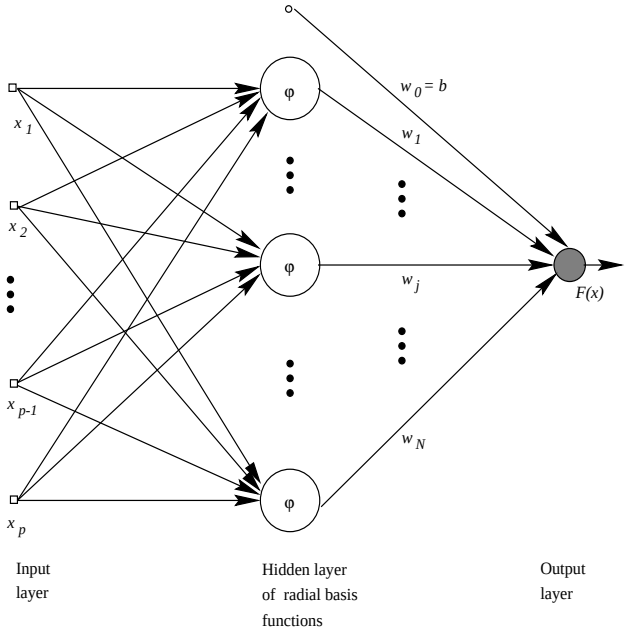


Fig. 1. The Radial Basis Function Network

The following sections discuss in detail the training and testing of the RBF networks and analyze the performance of the RBF model relative to the already existing models.

IV. Training the RBF network

The learning process undertaken by the RBF network can be visualized as follows. The hidden layer's activation functions evolve slowly in accordance with a nonlinear optimization strategy, while the output layer's weights adjust themselves rapidly through a linear optimization strategy. This learning procedure involves the minimization of the cost function

$$\mathcal{E} = \frac{1}{2}\sum_{j=1}^{N} e_j^2 \tag{2}$$

where N is the number of training examples and e_j is the error signal that is calculated as the difference between the desired and actual output values.

Two learning mechanisms were implemented for the RBF network models built: (a) Fixed selection of centers and (b) Supervised selection of centers. Both the mechanisms involve the initialization of the centers, as a subset of the training data set. In the fixed selection of centers, the initial set of chosen centers was not altered during the training process. Whereas, in the case

of the supervised selection, the centers are movable during the training process in an effort to improve the quality of the model. For both the methods, we applied the pseudo-inverse technique of Penrose [5] followed by Gaussian elimination to set the initial values of the output weights. We randomly partitioned our data set into a training set and testing set. As mentioned earlier, the model used 80 percent of the data as the training set and 20 percent as the test set.

A. Fixed Centers

The locations of the centers are chosen randomly from the training data set. If the training data are distributed in a representative manner for the problem, the random choice can produce good results [4]. The normalized radial basis function centered at t_i is defined as

$$G(\|x - t_i\|^2) = exp\left(-\frac{M}{d^2}\|x - t_i\|^2\right) \qquad i = 1, 2, ..., M \tag{3}$$

where M is the number of centers and d is the maximum distance between the chosen centers. In effect, the standard deviation or width of the Gaussian radial basis functions is fixed at

$$\sigma = \frac{d}{\sqrt{2M}} \tag{4}$$

The only parameters that would need to be learned in this approach are the linear weights in the output layer of the network. The initial weights w of the output layer are set using the formula w=G^+d where d is the desired response vector in the training set. The matrix G^+ is the Moore-Penrose pseudoinverse of the matrix G. Once the initial weights of the output layer are determined, the weights are modified using the delta rule [9] to bring the output as close to the desired output as possible.

In the fixed type of learning, the training process primarily involved the changing of the number of centers. We began training with as few as 100 centers and then gradually incremented it till the maximum number (5682) of the number of training examples was reached. With each set of centers, the initial weights of the output layer were determined. This was followed by the training of the output layer using the delta rule, as mentioned earlier.

B. Supervised Selection of Centers

In this approach, the location of the centers, the width of the centers and the linear weights of the output layer undergo a supervised learning process. The RBF network takes its most generalized form and is hence called the Generalized Radial Basis Function (GRBF) network. The gradient descent method is used to learn the location and width of the centers as well as the linear output weights.

The process of minimizing the cost function $\mathcal{E}$, as shown in equation (2), is dependent on varying: (1) the linear weights of the output layer and (2) the location and width of the centers in the hidden layer. In the gradient descent method, the values of w_i, t_i and σ_i that minimize $\mathcal{E}$ are regarded as the coordinates of the stable fixed point of the following dynamical system [7]:
1. Linear output weights

$$\dot{w_i} = \frac{\partial \mathcal{E}(n)}{\partial w_i(n)} = \sum_{j=1}^{N} e_j(n) G(\|x_j - t_i(n)\|^2) \tag{5}$$

2. Positions of centers

$$\dot{t_i} = \frac{\partial \mathcal{E}(n)}{\partial t_i(n)} = 2w_i(n) \sum_{j=1}^{N} e_j(n) G'(\|x_j - t_i(n)\|^2)[x_j - t_i(n)] \tag{6}$$

3. Spread of centers

$$\dot{\sigma_i} = \frac{\partial \mathcal{E}(n)}{\partial \sigma_i(n)} = -w_i(n) \sum_{j=1}^{N} e_j(n) G'(\|x_j - t_i(n)\|^2) Q_{ji}(n) \tag{7}$$

$$Q_{ji}(n) = [x_j - t_i(n)][x_j - t_i(n)]^T \tag{8}$$

In the supervised learning process, the update functions are performed in an iterative manner. n is the number of iterations of the learning process. The change in output layer weights is determined by the error signal e_j for each input x_j and each of the radial basis functions $G\|x_j - t_i(n)\|^2$ centered at t_i. The alteration of the position and the width of the centers is dependent on the error e_j , the first derivative of the Gaussian radial basis functions G' and distance of an input pattern x_j from each of the centers t_i at time instant n. A detailed derivation of the above equations can be found in [7] and [4].

Once the change in the output weights, the location and the spread of the centers are evaluated the corresponding original values are altered. During each iteration, each of the following update equations are implemented.

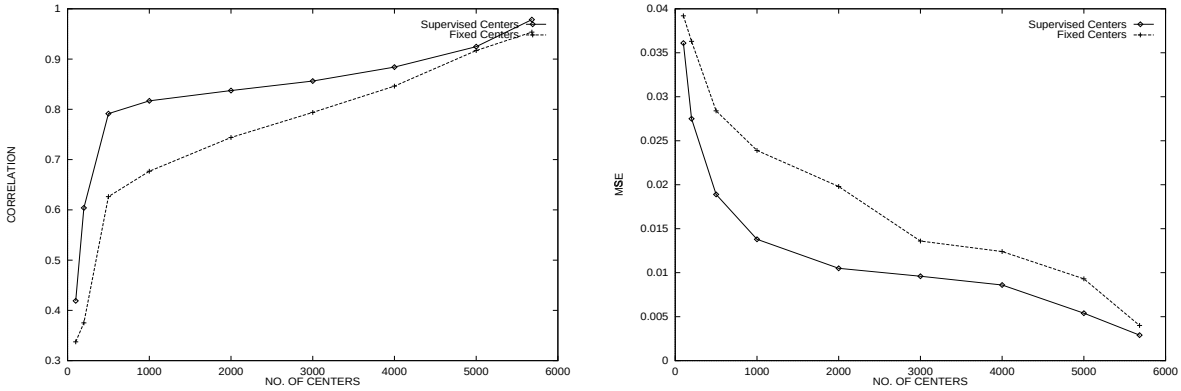
$$w_i(n+1) = w_i(n) - \eta_1 \frac{\partial \mathcal{E}(n)}{\partial w_i(n)} \quad i = 1, 2, ...M \tag{9}$$

$$t_i(n+1) = t_i(n) - \eta_2 \frac{\partial \mathcal{E}(n)}{\partial t_i(n)} \quad i = 1, 2..., M \tag{10}$$

$$\sigma_i(n+1) = \sigma_i(n) - \eta_3 \frac{\partial \mathcal{E}(n)}{\partial \sigma_i(n)} \tag{11}$$

The update equations for w_i (9), t_i (10) and σ_i (11) are assigned different learning parameters η_1, η_2, and η_3, respectively. Each iteration of the learning process involves three modification steps: updating the linear weights of the output layer as shown in equations (5) and (9), followed by changing the positions of the centers in the hidden layer as shown in equations (6) and (10) which is followed by the altering of the width or spread of the centers as shown in equations (7), (8) and (11). After each iteration the error is re-evaluated to check if the desired value has been obtained. If the error has not reached required minimum value, the evaluated values of the output layer weights, the location and width of the centers are now considered the initial values and the learning procedure is repeated.

For both fixed and supervised selection of centers alternate values for the learning rate were tried during the training process and the best values were selected. In addition, the number of centers was initialized to 100 and gradually incremented to 5682 during training.



(a) The Correlation (b) The MSE

Fig. 2. Results obtained for fixed and supervised learning during training

V. Performance of the RBF network on the training set

During training, we first evaluated the performance of the RBF network with fixed centers and then introduced the supervised selection of centers. We compared the results of both the learning mechanisms in terms of the mean squared error (MSE) and the correlation between the actual and desired outputs. The mean squared error is computed by calculating the average of the difference between the squares of the desired and predicted values of the change in AVR.

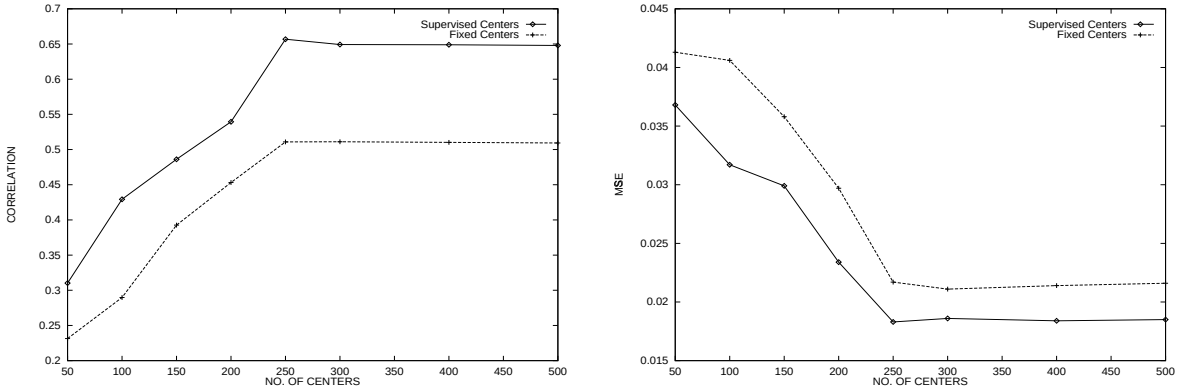
$$MSE = \frac{\Sigma_{i=1}^{N}(\Delta AVR_{desired} - \Delta AVR_{predicted})^2}{N} \tag{12}$$

For both the fixed and supervised selection of centers we used 80 % of the available data for training. We built 5 different RBF networks for each of the learning mechanisms using 5 different 80 % samples for training so that the performance of the networks is not determined by a specific sample set. We used the average of the obtained results to evaluate the performance of the networks. For both the learning mechanisms, we varied the number of centers from 100 to 5682 in increments of 100 initially, up to 500 centers, and then in increments of 500. At each number of centers, we trained the networks using the procedure described in the previous section.

Figure 2.a shows that the correlation improved for the fixed selection of centers from .337 for 100 centers to .953 for 5682 centers. For the supervised selection of centers, the correlation for 100 centers was .418 and increased to .978 for 5682 centers. It is observed that with a fewer number of centers, the supervised mechanism performs much better than the fixed selection. A correlation of .816 is achieved with 1000 centers. An equivalent performance could be obtained by the fixed model with as high as 3500 centers. As the number of centers is increased, the gap between the two correlation curves narrows. This shows that given sufficient number of centers, the performance of the regularization RBF network is equivalent to that of the GRBF network.

Correspondingly, the MSE decreased as the number of centers were increased for both the learning mechanisms. Figure 2.b shows the drop in the MSE for both fixed and supervised learning as the number of centers is increased. The MSE for the GRBF network dropped from .0361 for 100 centers to .0029 for 5682 centers. For the fixed learning, the MSE for 100 was .0392 and it reduced to .0029 for 5682 centers.

For any given number of centers, the supervised selection of centers performed better than the



(a) The Correlation (b) The MSE

Fig. 3. Results obtained for fixed and supervised learning during testing

the fixed mechanism. With the increase in centers, this performance gap reduced until it became almost negligible with the maximum number of centers of 5682.

VI. Performance of the RBF network on the test set

The generalization performance of the trained networks was then measured on the test set. Using this methodology, no information from the test set was used to determine any parameters during training. We tested the 5 different RBF networks built on the corresponding 5 samples of 20 percent of the data (1300 patterns) kept aside for testing.

We compared the correlation values obtained with the two learning mechanisms on the test set. Figure 3.a shows the average correlation over the 5 test sets obtained for both the learning mechanisms during testing. The performance of the RBF network with fixed centers increased till the number of centers was incremented to 300. Beyond 300 centers, the increase in the number of centers did not improve the performance. The correlation curve reached a constant value. For the supervised selection of centers, peak performance was obtained with 250 centers. Increasing the centers further did not improve the correlation. Here again, the supervised method outperformed the fixed one, for a given number of centers. The maximum correlation obtained with the fixed centers was .511 and the maximum correlation obtained for the supervised centers was .649 as shown Figure 3.a.

The MSE obtained for both the mechanisms is shown in Figure 3.b. Corresponding to the increase in correlation, a decrease in MSE was observed up to 300 centers in the case of fixed learning and 250 centers in the case of supervised learning. The minimum MSE obtained for the network with fixed centers was .0211 and that obtained with the network with supervised centers was .0183.

The best values of mean squared error and the correlation obtained for both the fixed and supervised selection of centers is shown in Table III.

As mentioned in section II, another measure of performance used was the ability of the network to correctly classify the change in AVR into a category. This was to measure the quality of performance on the entire range of the change in AVR and eliminate the bias created by the disproportionate number of records in the middle of the range. We measured the percentage of correct classification in each of the seven categories for both the learning mechanisms. We used the best RBF and GRBF networks of the 5 networks during the testing phase, to measure the

TABLE III
RESULTS OF THE RBF NETWORKS.

Model	R	MSE
RBF Network (Fixed centers)	0.51	.0211
GRBF Network	0.65	.0183
Percentage increase	27.4	15.3

classification capability of the RBF networks. The percentage of correct classification obtained for each of the seven categories of the AVR change is shown in Table IV. The RBF network with the supervised centers performed better than the fixed network in all the categories except category 4 where the percentage of classification was the same. The overall correct classification for both the networks are shown in Table IV. The overall classification of the supervised mechanism was better than that of the fixed mechanism.

Three measures were used to test the performance of the two types of RBF networks: the correlation, MSE and the ability of the network to classify the AVR change. In all the performance measures used, the GRBF network performed better than the RBF network. In addition, the best performance of the GRBF network was achieved with fewer number of centers.

TABLE IV
CATEGORY CLASSIFICATION OBTAINED BY THE RBF AND GRBF NETWORKS DURING TESTING

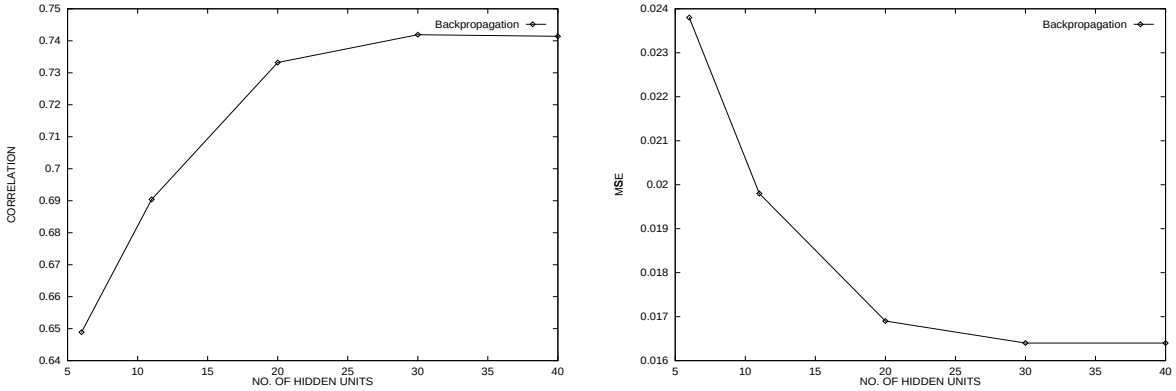
Model	C1	C2	C3	C4	C5	C6	C7	Overall
RBF network (Fixed centers)	53%	42%	50%	75%	76%	62%	31%	62.9%
GRBF network	55%	45%	51%	75%	82%	64%	36%	64.6%

VII. RESULTS WITH A BACKPROPAGATION NETWORK

Multi-layer, fully connected Backpropagation networks have been built to predict the change in AVR [6]. The neural network package used in building the networks in [6] is named PREDICT, a product of NeuralWare, Inc. The package allows the network builder to select from different learning rules and to change learning rates to improve network performance. It has the ability to automatically add new hidden layers to improve network performance. Several types of hidden layer units can be specified including sigmoidal and hyperbolic tangent units. The output units can be either linear or sigmoidal.

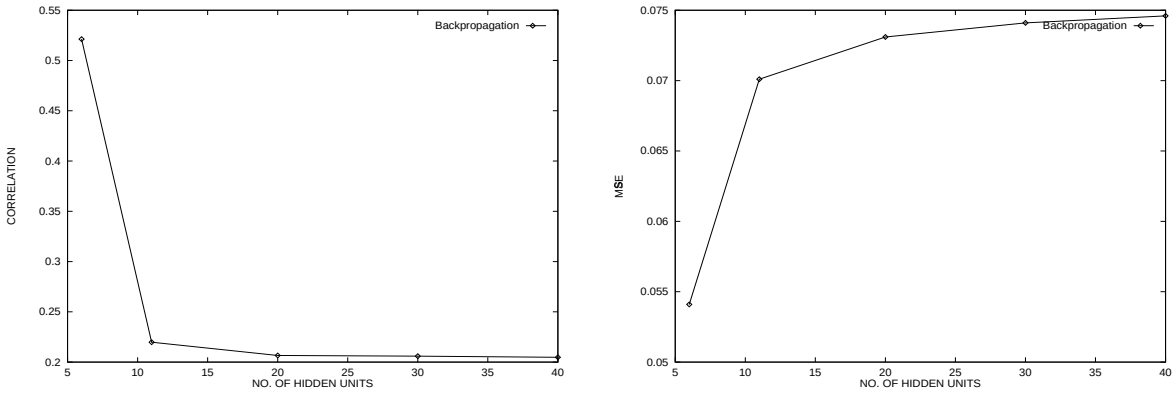
To compare the performance of the RBF networks, described in section VI, to that of Backpropagation, we used PREDICT to build Backpropagation networks to predict the change in AVR. The same 5 different sets of data with a 80%-20% split for training and testing that were used to build and test the RBF networks, were used to train and test the Backpropagation networks. The results reported below are for networks built with hyperbolic tangent units in the hidden layers and a linear unit in the output layer since they showed better performance than other types of units available with PREDICT.

During training, the performance of the Backpropagation network was measured as the number of hidden units was increased until no noticeable improvement was obtained. Figure 4.a shows the average over the 5 training sets of the correlation between the desired and actual values of



(a) The Correlation (b) The MSE

Fig. 4. Results obtained for the Backpropagation network during training



(a) The Correlation (b) The MSE

Fig. 5. Results obtained for the Backpropagation network during testing

the AVR change. As the number of hidden units increases, the correlation also increases up to a maximum of .741 with 30 units. No further improvement was observed on increasing the number of hidden units to more than 30. The MSE measurements show a corresponding decrease with the increase in the number of hidden units. This behavior is similar to that of the RBF networks shown in Figure 2.b. The chart in figure 4.b shows the MSE obtained for different number of hidden units.

We compared the results obtained on the five training sets for both the Backpropagation networks and the RBF networks. During training, the RBF networks outperformed the Backpropagation networks in terms of correlation and the MSE. The best correlation value obtained for the Backpropagation network was .7419 while a maximum correlation of .978 was obtained with the RBF networks. The best MSE obtained for the Backpropagation networks was .0164. For the RBF networks, the best MSE obtained was .0029.

The Backpropagation networks built were then tested on the remaining 20% of the data. Figure 5.a shows the average correlation obtained for the different Backpropagation networks built previously. The best correlation obtained is .5213 with 6 hidden units. Further increase in the number of hidden units did not improve the performance of the networks on the test set. Correspondingly,

the least MSE was observed with 6 hidden units. Figure 5.b shows the average MSE obtained by varying the number of hidden units.

We compared the results of the Backpropagation network obtained during testing with that of the RBF networks. The RBF networks perform better than the Backpropagation networks in terms of the correlation as well as the MSE. The best correlation obtained for the Backpropagation network was .5213 while the best correlation obtained for the RBF network was .6568. Correspondingly, the best MSE obtained for the Backpropagation network was .0541 whereas that of the RBF network was .0183.

During the testing phase, just as in the case of the RBF networks, we introduced the classification capability of the network as another performance measure. The Backpropagation network that performed best during testing was used to determine this capability of the network, as was the case with the RBF networks. The percentage of correct classification in each of the seven categories of AVR change was measured. Table V shows the percentage of correct classification obtained for each of the seven categories of AVR change. The Backpropagation network performed best in categorizing the AVR change in the middlemost range, category 4. The network was able to classify 75% of the AVR change correctly. In the extreme categories, 1 and 7, the performance of the network was poor with the percentage of classification being as low as 19% in category 1. Table V shows the percentage of overall correct classification obtained by the Backpropagation network during testing.

TABLE V

CATEGORYWISE CLASSIFICATION OBTAINED BY THE BACKPROPAGATION NETWORK DURING TESTING

Model	C1	C2	C3	C4	C5	C6	C7	Overall
Backpropagation network	19%	26%	30%	75%	46%	38%	25%	41%

We then compared the results obtained for the classification capability of the Backpropagation network with that of the RBF networks. It was found that the RBF networks are capable of classifying the AVR change into the predetermined categories better than the Backpropagation networks. Table VI compares the performance of the RBF networks and the Backpropagation networks on the test set.

The categorization capabilities of the RBF and the Backpropagation networks in the middle range, the category 4 in particular, are almost identical. Both the networks have percentages of correct classification as high as 75%. It is in the extremum categories that RBF network clearly outperforms the Backpropagtion networks. In categories 1 and 7, the percentage of correct classification is 55% and 36%, respectively, whereas in the case of the Backpropagation network it is 19% and 25%, respectively.

TABLE VI

COMPARISON OF RESULTS.

Model	R	Acceptable Classification
RBF Network (Fixed centers)	0.51	62.9%
GRBF Network	0.65	64.6%
BackPropagation Network	0.52	41%

VIII. Conclusions

In this paper, we analyzed the models built for the prediction of AVR change based on data obtained from three major cities. The goal of the models was to predict the effect of a company's incentives on its employees commuting habits. The predictive models used were Radial Basis Function networks and Backpropagation networks. The RBF networks were implemented using two learning mechanisms: fixed selection and supervised selection of centers.

The results show that allowing the alteration of the location of centers in the hidden layer, apart from the alteration of the output layer weights improved the performance of the RBF networks. With the supervised learning of center locations, the RBF network could achieve better results during both training and testing, with fewer number of centers.

The RBF networks outperformed the Backpropagation networks in terms of all the performance measures discussed in this paper, both during training and testing. During training, the best correlation value obtained for the Backpropagation network was .741. In the case of the RBF networks, the best correlation obtained during training was .978.

During testing the RBF networks were able to exceed substantially the generalization of the Backpropagation networks. The maximum correlation obtained by the Backpropagation network was .5213 while that obtained by the RBF networks was .649. The best MSE obtained by the Backpropagation network was .0541 whereas the best MSE obtained by the RBF network was .0183. The percentage of correct classification obtained using the Backpropagation network was 41%, whereas the percentage of correct classification obtained with the RBF networks was 64.6%.

In the case of the RBF network, the centers are initially a subset of the training set. Therefore, the choice of the centers in a manner that they are representative of the entire training set is possible. This flexibility in the selection of hidden layer units is absent in the Backpropagation networks. This feature of RBF networks makes them particularly more effective in the AVR change prediction problem because the data is unevenly distributed.

References

[1] D.S.Broomhead, D.Lowe, "Multivariable funtional interpolation and adaptive networks," *Complex systems 2*, pp. 321-355, 1988.

[2] C.Darken, J.Moody, "Towards faster stochastic gradient search," *Advances in Neural information Processing Systems 4* (J.E. Moody, S.J. Hanson, R.P.Lippmann, eds.), pp. 1009-1016, 1992.

[3] G.H.Golub, C.F.Van Loan, *Matrix Computations*, 2nd ed., Baltimore, MD, John Hopkins University Press, 1989.

[4] S.Haykin *Neural Networks*, New York, Macmillan, 1994.

[5] R.Penrose, "A generalized inverse for matrices," *Proceedings of Cambridge Philosophical Society 51*, pp. 406-413, 1955.

[6] R.A.Perez, M.C. Pietrzyk, "A neural network approach to the development of trip reduction plans for employers", 1997

[7] T.Poggio, F.Girosi, "A theory of networks for approximation and learning," Report Number AI-1140, MIT Artificial Intelligence Laboratory, Cambridge, MA, July 1989.

[8] D.Wettschereck, T.Dietterich, "Improving the performance of radial basis function networks by learning center locations," *Advances in Neural Information Processing Systems 4* (J.E. Moody, S.J. Hanson, R.P.Lippmann, eds.), pp. 1133-1140, 1992.

[9] B.Widrow, M.E.Hoff,Jr., "Adaptive switching circuits," *IRE WESCON Convention Record*, pp. 435-461, 1960.