

Data Mining based Prediction of Unemployment Rate Data by using Radial Basis Function Neural Network (RBFNN)

¹Younis Kadthem Hameed, ²Wahab S. Mohammed and ¹Fadhel Kadhem Zaidan

¹Department of Statistic, College of Administration and Economics,

²Electronic Computing Center, University of Diyala, Baqubah, Iraq
mscyounis@gmail.com

Abstract: Data mining is one of the knowledge discovery steps in database in which modeling techniques are applied. In this study, the unemployment rate in the Iraqi database prediction model is determined according to the RBF neural network where the principle and algorithm of the RBF is used. Our study also predicts the future unemployment rate in Iraq. To increase the efficiency of mining process, some pre-processing needs to be done to the data. Experimental results showed the good accuracy when applied to the adjust data.

Key words: Radial basis function, predict, data mining, database, predicts, mining process

INTRODUCTION

The forecasting methods which are all around embraced have the accompanying sorts: basic relapse examination strategy (SRA), Exponential Smoothing strategy (ES), complete autoregressive versatile investigation strategy (CAMA) and so on. These strategies are great connected in a few expectations, however, regardless they exist numerous deformities. These techniques have a decent impact of straight forecast, yet are not extremely exact of nonlinear expectation. These strategies don't matter to multi-factor conjecture. The foundation of the forecast show is intensely reliant on the information level of the anticipate individuals. Manufactured neural system (ANN) displaying it's one of successful technique to investigation strategies for gauging which can well uncover the connection of nonlinear time arrangement in defer state space. Then, the ANN can accomplish the motivation behind expectation. The Kolmogorov continuation hypothesis in neural system hypothesis guarantees the expectation possibility of the neural system which is utilized for time arrangement from the perspective of arithmetic. The joblessness rate is chosen by numerous goal factors. The conjecture of the joblessness rate has not great ways now. Because the joblessness rate has a decent nonlinear trademark, the best use of the RBF neural system to deal with nonlinear issues, then the RBF neural system can be applied to estimate the joblessness rate. The study sets up forecast demonstrate and predicts the joblessness rate. The after

effect of trial demonstrates that the forecast display has great expectation impact. Joblessness in Iraq has turned into a financial and social measurement of the issue of plainly communicating deficiency in monetary structures and deformity social at the national level where joblessness is a genuine social scourge of human limit crashes and scattered development openings monetary prosperity. In order to monitor the level of unemployment trends in Iraq it must provide periodic statistical indicators reflect the reality of the size of this phenomenon. It has been taking the unemployment rate in Iraq from 1990-2013 (Wei, 2005).

Radial Basis Function (RBF) networks: RBF systems are bolster forward systems prepared utilizing a managed preparing calculation. They are sometimes arranged with a solitary shrouded layers of the units whose enactment work is chosen from class of capacities called premise capacities. While like back spread in numerous regards, spiral premise work systems have a few favorable circumstances. They as a rule prepare substantially quicker than back proliferation systems. They are less powerless to issues with non-stationary information sources due to the conduct of the spiral premise work concealed units.

Popularized by Moody and Darken RBF networks have proven to be useful neural network architecture. The real distinction between RBF arranges and back proliferation organizes (that is multi-layer perceptron prepared by back spread calculation) is the conduct of the single concealed layer. Instead of utilizing the sigmoidal

or S-molded actuation study as in back engendering, the concealed units in RBF systems utilize a Gaussian or some different premise piece work. Each concealed unit goes about as a privately tuned processor that registers a score for the match between the information vector and its association weights or focuses. In actuality, the premise units are exceptionally specific example finders. The weights interfacing the premise units to the yields are utilized to take straight mixes of the shrouded units to item the last characterization or yield (Sun *et al.*, 2002).

MATERIALS AND METHODS

The RBF networks structure: Radial basis function network has 3-layers feed-forward neural network which incorporates an info layer, a shrouded layer with outspread premise work neurons and a yield layer with direct neurons as shown in Fig. 1. Concealed layer is generally utilizing spiral premise work as excitation work and the outspread premise excitation work is normally Gaussian capacity which is typically communicated as:

$$R(x^q - c_i) = \exp[-(\|w_{1i} - x^q\| \times b_{1i})^2] \quad (1)$$

where, $\|w_{1i} - x^q\|$ is the Euclidean distance, c is the center of Gaussian function. $X^q = \{x_1^q, x_2^q, \dots, x_n^q\}$ is the q th to input data. w_{1i} it is the distance between weight vector, associated with the contributing layer and in each neuron in the shrouded layer, the info vector is x^q increased by the limit b_{1i} which is considered as its own particular information as shown in Fig. 2 (Sun *et al.*, 2002). So, we get the i th input of hidden layer neuron which can be expressed as k_i^q :

$$k_i^q = \sqrt{\sum_j (w_{1ji} - x_j^q)^2} \times b_{1i} \quad (2)$$

i th output of hidden layer neuron can be expressed as r_i^q :

$$r_i^q = \exp \left[-\sqrt{\sum_j (w_{1ji} - x_j^q)^2} \times b_{1i} \right] \quad (3)$$

The RBF neural network output is the weighted summation of each concealed layer neuron's yield and the excitation work is utilizing unadulterated straight capacity, so the q th yield layer neuron's yield which is comparing to the q th input can be expressed as y^q :

$$y^q = \sum_{i=1}^n r_i^q \times w_{2ih} \quad (4)$$

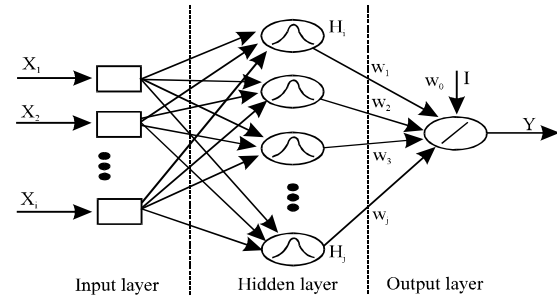


Fig. 1: Note how the inscription is focused in the segment

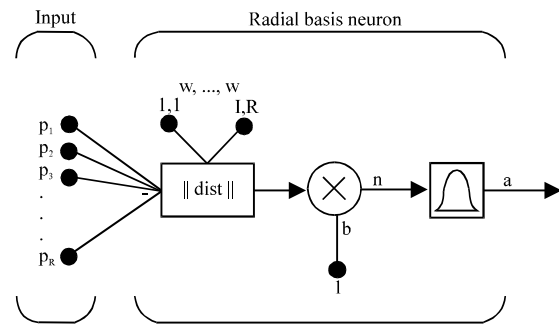


Fig. 2: RBF neural system input and output of shrouded layer neurons

Radial basis function network algorithm: RBF neural network learning procedure can be partitioned into $h(x)$ is the Gaussian initiation work with the parameters r (the span or standard deviation) and c (the middle or normal taken from the information space) characterized independently at each RBF unit. The learning procedure depend on modifying the parameters of the system to duplicate an arrangement of information yield designs. There are three sorts of parameters the weight w between the shrouded hubs and the yield hubs, the middle c of every neuron of the concealed layer and the unit width r (Sun *et al.*, 2002).

Unit center (c): Any bunching calculation can be utilized to decide the RBF unit centers (e.g., k-means clustering). An arrangement of groups each with r -dimensional focuses is dictated by the quantity of info factors or hubs of the information layer. The group focuses turn into the focuses of the RBF units. The quantity of bunches, H is a plan parameter and decides the quantity of hubs in the shrouded layer. The k -implies bunching calculation continues as follows: initialize the focal point of each group to an alternate arbitrarily chose preparing design. Assign each preparation example to the closest group. This can be expert by computing the Euclidean

Table 1: The unemployment rate in Iraq on 1990-2013

Years	Unemployment rate
1990	5.5
1991	6.5
1992	7.5
1993	8.5
1994	10.5
1995	12.9
1996	13.9
1997	15.4
1998	17.4
1999	20.2
2000	22.4
2001	24.6
2002	26.7
2003	28.1
2004	26.8
2005	17.9
2006	17.5
2007	11.7
2008	15.3
2009	14
2010	12
2011	11
2012	11.9
2013	12.1

Data from the unemployment rate in Iraq of central statistical organization

separations between the preparation designs and the bunch focuses. When all preparation designs are appointed, figure the normal position for each group focus. They at that point turn out to be new bunch focuses. Repeat stages 2 and 3, until the point that the group focuses don't change amid the consequent cycles (Table 1).

Unit width (r): At the point when the RBF focuses have been built up, the width of each RBF unit can be ascertained utilizing the k-closest neighbor's calculation. A number k is picked and for each middle, the k closest focuses is found. The root-mean squared separation between the present bunch focus and its k closest neighbors is ascertained and this is the esteem decided for the unit width (r). So, if the current cluster center is c_j , the r value is:

$$r_j = \sqrt{\frac{\sum_{i=1}^k (c_j - c_i)^2}{k}} \quad (5)$$

Variance (σ): The basis function of RBF neural network is Gaussian function, therefore, variance can be given as:

$$\sigma_i = \frac{c_{max}}{\sqrt{2h}} \quad i = 1, 2, 3, \dots, h \quad (6)$$

c_{max} is maximum distance among the selected centers.

Weights (w): The weights associating the shrouded layer to produce a layer of neurons can be obtained specifically through the lowest square technique. Its computation recipe is as follows:

$$w = \exp \left[\frac{h}{c_{max}^2} \|x_q - a\|^2 \right] \quad (7)$$

Input samples pretreatment: Since, the verifiable capacity of RBF neural system is Gaussian capacity in general requires for input an incentive in the vicinity of 0 and 1, do standardize on the joblessness rate in Iraq from 1990-2013. Standardization is fundamentally a similar method to factual information standardization, for the most part utilizing the accompanying structure:

$$\bar{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (8)$$

where, X is the genuine estimation of test, X_{max} takes an extensive esteem, guaranteeing conjecture year is not as much as the esteem; X_{min} takes an example of information is not as much as the base an incentive to guarantee standardized esteem isn't near 0. After the pretreatment of information finishes the preparation, due process information (backwards change) to get the real esteem (Liu *et al.*, 2009).

RESULTS AND DISCUSSION

The quantity of info layer neurons comparing to the measurement of information vectors, the quantity of information layer hubs is excessively, making the system take in the quantity of moderately expansive input hubs is nearly nothings, follow-up values can't mirror the estimation of the relationship between the forerunner. That leads to decide, the quantity of information neuron of RBF neural system is 5 and the quantity of yield neurons is 1. Treating the samples as follows: Input neuron $P = [p(t-5), p(t-4), p(t-3), p(t-2), p(t-1)]$ output neurons $T = [p'(t)]$. Where, $t = 1990, 1991, \dots, 2013$, $P(t)$ denote the normalized number of unemployment rate at t year. We can obtain in this method the training samples and test samples as shown in Table 2. To test the accuracy and efficiency of the network, select group 1-18 group data as the study samples, the groups (19-24) as the test samples and using the trained RBF neural networks to predict (Zhang and Li, 2012). Creating a precise neural network by new be function, this function

Table 2: The train and test group actual unemployment rate in Iraq on 1990-2013

Groups	Input neurons P = [p (t-5), p (t-4), p (t-3), p (t-2), p (t-1)]					Output neurons T
	p(t - 5)	p(t - 4)	p(t - 3)	p(t - 2)	p(t - 1)	p'(t)
Train group						
1	0.194567	0.987435	0.950681	0.892438	0.808902	0.208798
2	0.987435	0.809456	0.987435	0.950681	0.878696	0.246352
3	0.950681	0.987435	0.342334	0.987435	0.934967	0.287871
4	0.892438	0.950681	0.987435	0.242325	0.974468	0.333158
5	0.771664	0.853805	0.921099	0.968884	0.883489	0.414091
6	0.603957	0.697601	0.785644	0.862703	0.965637	0.516276
7	0.509827	0.603957	0.697601	0.785644	0.922668	0.573778
8	0.394785	0.482695	0.575443	0.668881	0.843033	0.649555
9	0.272562	0.346139	0.428602	0.517458	0.71962	0.740945
10	0.152859	0.203679	0.264616	0.335199	0.540095	0.847642
11	0.087347	0.121191	0.163951	0.216259	0.399033	0.918281
12	0.046359	0.066978	0.09435	0.129591	0.277529	0.967183
13	0.023476	0.035273	0.051674	0.073812	0.184786	0.990379
14	0.013602	0.021067	0.031815	0.046845	0.132488	0.993535
15	0.01645	0.025382	0.038185	0.056012	0.150833	0.016358
16	0.091213	0.127358	0.173386	0.230154	0.402869	0.824165
17	0.079751	0.112202	0.153917	0.205866	0.368498	0.803989
18	0.126172	0.167059	0.215672	0.271479	0.41411	0.56472
Test group						
19	0.070261	0.098601	0.134917	0.179998	0.322061	0.699474
20	0.064631	0.090357	0.123169	0.163703	0.29311	0.634265
21	0.061057	0.084287	0.113451	0.148892	0.262939	0.540698
22	0.050829	0.070168	0.094446	0.12395	0.223983	0.486746
23	0.036194	0.05118	0.070564	0.094859	0.185246	0.501795
24	0.026788	0.038459	0.053835	0.073477	0.152504	0.488568

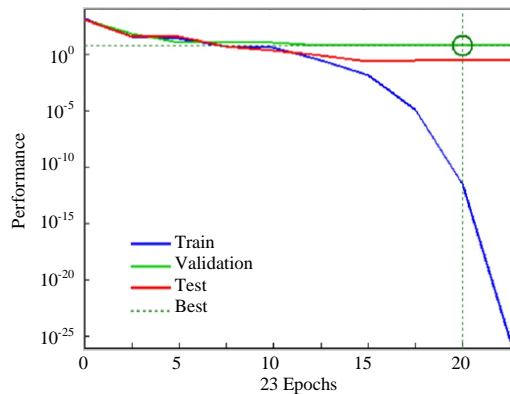


Fig. 3: The result of RBF neural network training in 23 epochs; Performance is 1.02234e-025, Goal is 1e- 005

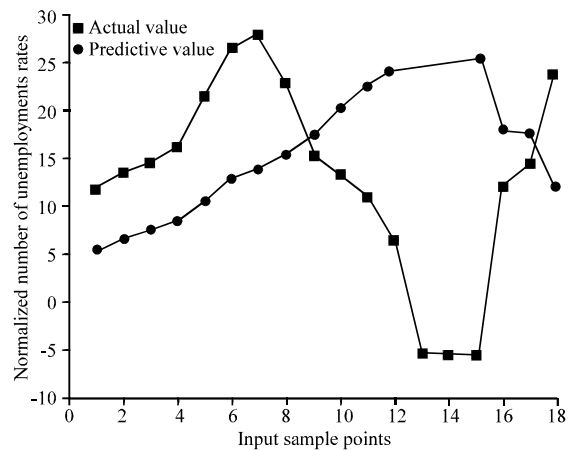


Fig. 4: RBF neural network learning and training curve

creates RBF network, automatically select the number of hidden layer and make the error to 0, training sample, shown in Fig. 3. Using MATLAB (Defeng, 2009a, b). In this Fig. 3, the percentage of training is 60%, validation is 20% and test is 20% in 23 epochs. Then we can get the RBF neural network learning and training curve as shown in Fig. 4. Now, we can be testing the neural network and verifying the prediction performance, shown in Fig. 5. This gives an exact premise to foreseeing joblessness rate in Iraq later on. Enter the genuine incentive from 2005-2013 can get the standardized

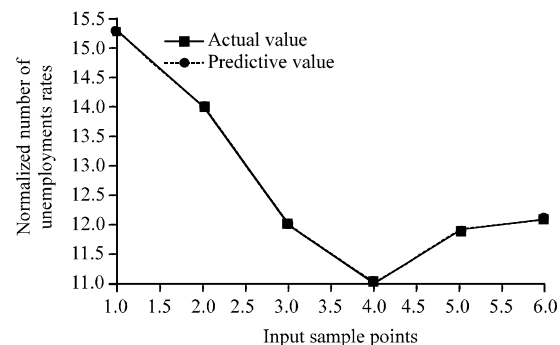


Fig. 5: Output prediction curve to RBF network model

Table 3: The predicted value of unemployment rate in Iraq from 2014-2023

Years	Predictive values
2014	12.7
2015	14.5
2016	23.5
2017	23.9
2018	28.1
2019	24.6
2020	22.4
2021	24.3
2022	24.8
2023	26.5

anticipated an incentive in 2014. Similarly, after a multi-step iterative can get the standardized anticipated an incentive from 2014-2023, at that point after backwards change, the anticipated esteem can be gotten from 2014-2023 as shown in Table 3 (Defeng, 2009a, b; Sing *et al.*, 2003).

CONCLUSION

This study has introduced an information mining of joblessness rate in Iraq forecast strategy in view of RBF neural system. Through RBF neural system versatile, self-association and self-learning capacity, influencing joblessness to rate like clockwork as the RBF neural system's information, we can anticipate joblessness rate in the ten year. After MATLAB preparing, estimating, reenactment, this technique accomplishes a decent

expectation impact. The utilization of this technique gives another state of mind for mimicking and predicting unemployment rate Iraq.

REFERENCES

- Defeng, Z., 2009a. MATLAB Neural Network Application Design. Mechanical Industry Press, Beijing, China.
- Defeng, Z., 2009b. MATLAB Neural Network Simulation and Application. Electronic Industry Press, Beijing, China.
- Liu, X.Q., X.Y. Wang and R.D. Yu, 2009. Study on traffic accidents prediction model based on RBF neural network. *Comput. Eng. Appl.*, 45: 188-190.
- Sing, J.K., D.K. Basu, M. Nasipuri and M. Kundu, 2003. Improved k-means algorithm in the design of RBF neural networks. *Proceedings of the International Conference on Convergent Technologies for Asia-Pacific Region (TENCON 2003) Vol. 2*, October 15-17, 2003, IEEE, Bangalore, India, pp: 841-845.
- Sun, Y.P., L. Zhang and R. Lu, 2002. Tourist quantity forecast by using neural network. *Hum. Geogr.*, 17: 50-52.
- Wei, H., 2005. Neural network structure design theory and method. National Defence Industry Press, Beijing, China.
- Zhang, H. and J. Li, 2012. Prediction of tourist quantity based on RBF neural network. *J. Comput.*, 7: 965-970.