Online ISSN : 0975-4172
Print ISSN : 0975-4350

Global Journal

of Computer Science and Technology: D

Neural & AI

Neural Network Algorithms

Artificial Neural Networks

Highlights

Multiple Regression Analysis

Software Effort Estimations

Discovering Thoughts, Inventing Future

VOLUME 13 ISSUE 2 VERSION 1.0

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence

Global Journal of Computer Science and Technology: D
Neural & Artificial Intelligence

Volume 13 Issue 2 (Ver. 1.0)

Open Association of Research Society

Incorporation No.: 0423089
License No.: 42125/022010/1186
Registration No.: 430374
Import-Export Code: 1109007027
Employer Identification Number (EIN):
USA Tax ID: 98-0673427

Global Journals Inc.

(A Delaware USA Incorporation with "Good Standing"; Reg. Number: 0423089)

Sponsors: *Open Association of Research Society*
Open Scientific Standards

Publisher's Headquarters office

Global Journals Inc., Headquarters Corporate Office, Cambridge Office Center, II Canal Park, Floor No. 5th, ***Cambridge (Massachusetts)***, Pin: MA 02141 United States
USA Toll Free: +001-888-839-7392
USA Toll Free Fax: +001-888-839-7392

Offset Typesetting

Global Association of Research, Marsh Road, Rainham, Essex, London RM13 8EU United Kingdom.

Packaging & Continental Dispatching

Global Journals, India

Find a correspondence nodal officer near you

To find nodal officer of your country, please email us at *local@globaljournals.org*

eContacts

Press Inquiries: *press@globaljournals.org*
Investor Inquiries: *investers@globaljournals.org*
Technical Support: *technology@globaljournals.org*
Media & Releases: *media@globaljournals.org*

Pricing (Including by Air Parcel Charges):

For Authors:
22 USD (B/W) & 50 USD (Color)
Yearly Subscription (Personal & Institutional):
200 USD (B/W) & 250 USD (Color)

Editorial Board Members (HON.)

Dr. Bart Lambrecht
Director of Research in Accounting and FinanceProfessor of Finance
Lancaster University Management School
BA (Antwerp); MPhil, MA, PhD (Cambridge)

Dr. Carlos García Pont
Associate Professor of Marketing
IESE Business School, University of Navarra
Doctor of Philosophy (Management), Massachusetts Institute of Technology (MIT)
Master in Business Administration, IESE, University of Navarra
Degree in Industrial Engineering, Universitat Politècnica de Catalunya

Dr. Fotini Labropulu
Mathematics - Luther College
University of ReginaPh.D., M.Sc. in Mathematics
B.A. (Honors) in Mathematics
University of Windso

Dr. Lynn Lim
Reader in Business and Marketing
Roehampton University, London
BCom, PGDip, MBA (Distinction), PhD, FHEA

Dr. Mihaly Mezei
ASSOCIATE PROFESSOR
Department of Structural and Chemical Biology, Mount Sinai School of Medical Center
Ph.D., Etvs Lornd University
Postdoctoral Training,
New York University

Dr. Söhnke M. Bartram
Department of Accounting and FinanceLancaster University Management SchoolPh.D. (WHU Koblenz)
MBA/BBA (University of Saarbrücken)

Dr. Miguel Angel Ariño
Professor of Decision Sciences
IESE Business School
Barcelona, Spain (Universidad de Navarra)
CEIBS (China Europe International Business School).
Beijing, Shanghai and Shenzhen
Ph.D. in Mathematics
University of Barcelona
BA in Mathematics (Licenciatura)
University of Barcelona

Philip G. Moscoso
Technology and Operations Management
IESE Business School, University of Navarra
Ph.D in Industrial Engineering and Management, ETH Zurich
M.Sc. in Chemical Engineering, ETH Zurich

Dr. Sanjay Dixit, M.D.
Director, EP Laboratories, Philadelphia VA Medical Center
Cardiovascular Medicine - Cardiac Arrhythmia
Univ of Penn School of Medicine

Dr. Han-Xiang Deng
MD., Ph.D
Associate Professor and Research
Department Division of Neuromuscular Medicine
Davee Department of Neurology and Clinical NeuroscienceNorthwestern University
Feinberg School of Medicine

Dr. Pina C. Sanelli
Associate Professor of Public Health
Weill Cornell Medical College
Associate Attending Radiologist
NewYork-Presbyterian Hospital
MRI, MRA, CT, and CTA
Neuroradiology and Diagnostic Radiology
M.D., State University of New York at Buffalo,School of Medicine and Biomedical Sciences

Dr. Roberto Sanchez
Associate Professor
Department of Structural and Chemical Biology
Mount Sinai School of Medicine
Ph.D., The Rockefeller University

Dr. Wen-Yih Sun
Professor of Earth and Atmospheric SciencesPurdue University Director
National Center for Typhoon and Flooding Research, Taiwan
University Chair Professor
Department of Atmospheric Sciences, National Central University, Chung-Li, TaiwanUniversity Chair Professor
Institute of Environmental Engineering, National Chiao Tung University, Hsin-chu, Taiwan.Ph.D., MS The University of Chicago, Geophysical Sciences
BS National Taiwan University, Atmospheric Sciences
Associate Professor of Radiology

Dr. Michael R. Rudnick
M.D., FACP
Associate Professor of Medicine
Chief, Renal Electrolyte and Hypertension Division (PMC)
Penn Medicine, University of Pennsylvania
Presbyterian Medical Center, Philadelphia
Nephrology and Internal Medicine
Certified by the American Board of Internal Medicine

Dr. Bassey Benjamin Esu
B.Sc. Marketing; MBA Marketing; Ph.D Marketing
Lecturer, Department of Marketing, University of Calabar
Tourism Consultant, Cross River State Tourism Development Department
Co-ordinator , Sustainable Tourism Initiative, Calabar, Nigeria

Dr. Aziz M. Barbar, Ph.D.
IEEE Senior Member
Chairperson, Department of Computer Science
AUST - American University of Science & Technology
Alfred Naccash Avenue – Ashrafieh

Contents of the Volume

Global Journal of Computer Science and Technology
Neural & Artificial Intelligence
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Neural Network Algorithms for using Radon Emanations as an Earthquake Precursor

By Gupta Dhawal & Shahani D. T
Indian Institute of Technology, New-Delhi

Abstract - The investigation throughout the world in past two decades provides evidence which indicates that significance variation of radon and other soil gases may occur in association with major geophysical events such as earthquake events. The traditional statistical algorithm which included regression to remove the effect of the meteorological parameters from the as is measured radon along with additional variation that periodicity in seasonal variations is computed using Fast Fourier Transform has shown to improve reliability of prediction of earthquake The present paper deals with the use of neural network algorithms which can learn the behavior of radon with respect to known meteorological parameters. This method has potential of tracking "changing patterns" in dependence of radon on meteorological parameters and it may adapt to such changes on its own in due course of time. Another neural network algorithm using Probabilistic Neural Networks that requires neither an explicit step of regression nor use of any specific period is also presented.

Keywords : *radon, anomalies, earthquake precursor, neural networks.*

GJCST-D Classification : *F.1.1, F.2.1*

Strictly as per the compliance and regulations of:

Neural Network Algorithms for using Radon Emanations as an Earthquake Precursor

Gupta Dhawal [α] & Shahani D. T [σ]

Abstract - The investigation throughout the world in past two decades provides evidence which indicates that significance variation of radon and other soil gases may occur in association with major geophysical events such as earthquake events. The traditional statistical algorithm which included regression to remove the effect of the meteorological parameters from the as is measured radon along with additional variation that periodicity in seasonal variations is computed using Fast Fourier Transform has shown to improve reliability of prediction of earthquake The present paper deals with the use of neural network algorithms which can learn the behavior of radon with respect to known meteorological parameters. This method has potential of tracking "changing patterns" in dependence of radon on meteorological parameters and it may adapt to such changes on its own in due course of time. Another neural network algorithm using Probabilistic Neural Networks that requires neither an explicit step of regression nor use of any specific period is also presented.

Keywords : radon, anomalies, earthquake precursor, neural networks.

I. Introduction

In India more than 50% of the land area is seismically active. Any earthquake in these areas of Magnitude 5.5 Richer Scale and above can cause severe loss of human life and property. The vulnerability of our civilization to earthquakes is rapidly growing, raising earthquakes to the ranks of major threats faced by humankind. About a million earthquakes of Magnitude 2 or more are registered each year worldwide. About a hundred of them cause serious damage and, once or twice in a decade, a catastrophic earthquake occurs. The vulnerability of our world to earthquakes is rapidly growing due to well-known global trends like proliferation of high-risk construction such as nuclear power plants, high dams, radioactive waste disposals, deterioration of the ground and destabilization of engineering infrastructures in megacities, destabilization of the environment, population growth and other factors, including the escalating socioeconomic volatility of the global village.

a) Earthquake Precursory Studies

Earthquakes constitute a source of severe human disasters all around the world that occurs in a relatively short time span of occurrence of an earthquake, and considerable loss of life can be averted if a warning could be issued prior to its occurrence. Consequently, short-term indicators — through the search for precursory signals — have received great attention in the last several decades. As earthquakes are physical phenomena, most techniques used currently with prediction purposes are based on geophysical approaches, including seismology, magnetism, electricity, and geodesy. So, a wide range of methods have been proposed, using the monitoring of parameters such as *b*-values (i.e. the slope of the Gutenberg–Richter law relating the local number of earthquakes and their magnitude), VP/VS-values (ratio of the propagation velocities of P and S seismic waves), coda *Q*, tilt values, self-potential anomalies and electromagnetic data, that allowed to exhibit case by case precursory signals [Varostos and Alexopoulos, 1984]; [Jin and Aki, 1986]; [Molchan and Dmitrieva, 1990]. The most relevant success in this field is probably the successful prediction of the February 4, 1975 magnitude 7.3 earthquake of Haicheng (China), on the basis of multiple precursory phenomena.

In India, earthquake precursor related research was started about three decades back and studies were mostly confined to seismological parameters investigations/observations. Though, the seismic gap hypothesis which proposes that the probability of a large earthquake in an individual fault segment is greater for those segments that have not slipped in a long time, has already been applied to Himalaya on the basis of energy release, micro-earthquake activity and seismicity patterns and three well known seismic gaps have been identified in the Indian Himalayan region namely; (1) Himachal gap in Himachal Pradesh, (2) Central gap in Central Himalaya and (3) Assam gap in Northeast Himalaya [Srivastava,1973]; [Srivastava and Rao, 1991]; [Khattri and Wyss, 1978]. After successful medium term forecast of 1988 M 7.3 earthquake in NE Himalayan region [Gupta and Singh, 1986], there was a lull period for quite some time. The first short term forecast of August 30, 1986 earthquake of M 5.0 was made by [Gupta et. al., 2005]. This forecast was based on the nucleation pattern. Subsequently, several such forecast were made for Koyna region like 13 November 2005 M 4.0, 26 December 2005 M 4.2 and 17 April 2006 M 4.7 based on the nucleation process [Gupta et.al., 2007].

Author α : Indian Institute of Technology, Hauz-Khas, New-Delhi-110016. E-mail : dhawal.gupta@gmail.com

b) Radon in Soil Gas: Literature Review

The first evidence of a correlation between radon and earthquake came from observation of radon concentration in the mineral water of the Tashkent Basin prior to the destructive earthquake of 1966 [Ulamov and Mavashev, 1967]. Radon observations, both in soil gas and in ground water revealed many precursory changes of radon emission levels [Lomnitz and Lomnitz, 1978[38]; Virk, 1993; Igarashi et al., 1995]. The effect of meteorological parameters was also analyzed by calculation the correlation coefficients and radon anomalies were found using the standard statistical procedures .The differentiation of radon emissions due to earthquakes from those due to effect of meteorological parameters on the measured radon concentrations were studied by [Wattananikorn, 1998].Observations of radon have also been part of the international prediction projects in the Iceland test area. Significant pre-earthquake changes were found and discussed and described in [Stefansson, 2011].

II. Neural Network Algorithm for Radon Emanations Estimate

An artificial neural network is an information processing system that consists of large number of simple processing elements called neurons. Each neuron is connected to other neuron by means of direct connection with an associated weight, which present information being used by the net to solve a problem. A general neural network is characterized by its pattern connections among the neurons, its method of determining weights and its activation function. The main advantages of the neural network method are learning capability for developing new solutions to problems that are not well defined, an ability to deal with computational complexity, a facility of carrying out quick interpolative reasoning, and finding functional relationship between sets of data. The statistical algorithm involves regression of meteorological parameters with measured radon. The regression equations thus obtained are used to find corrected radon time series. In case of neural networks the regression step is avoided. Hence a neural network model can be found which can learn the behavior of radon with respect to meteorological parameter in order that changing emission patterns may be adapted to by the model on its own. The output of this neural model is the estimated radon values. This estimated radon value is used to decide whether anomalous behavior of radon has occurred and a valid precursor may be identified.

There are varieties of neural network architectures available, which can model time series like Multi-layer perceptrons, Probabilistic neural networks, and Radial Basis function networks. Initially different neural network architectures were tested. Fig 1(a-b) shows the multi layer or MLP neural network architectures which were tested for the estimation of radon. The nomenclature followed for naming the neural network in the figure is <Type of NN>< Input> : <L1><L2><L3> :< Output>. The Fig. 1(a) indicates MLP s20 5:100-3-1:1 which indicates that the type of neural network is Multi Layer Perceptron with five inputs, three hidden layers with 100, 3 and 1 hidden neuron and one output. Fig. 1(a-b) also indicates the training performance, selection performance and test performance.

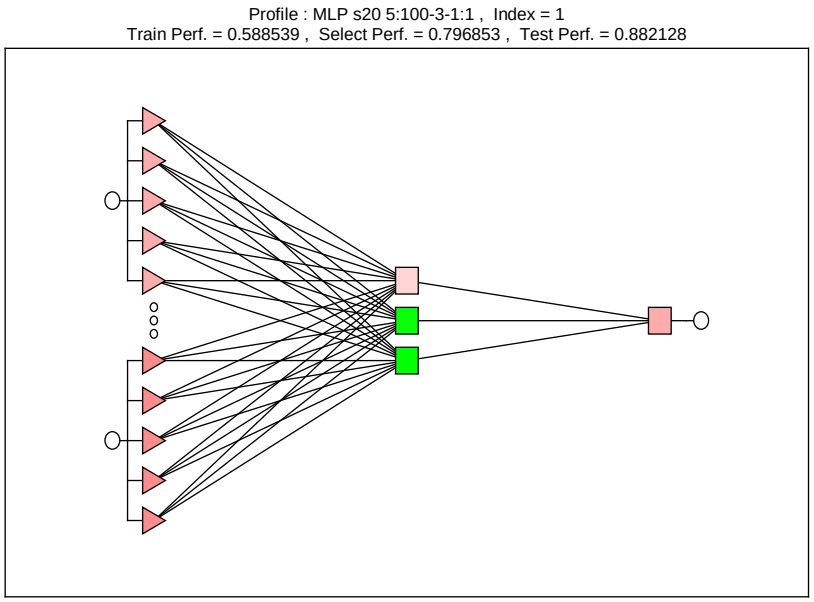


Figure 1 (a)

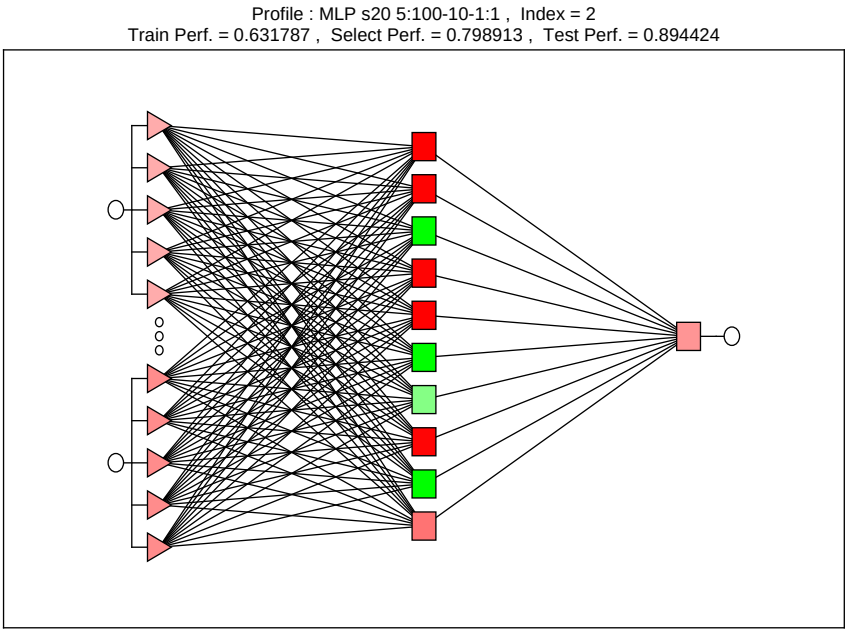


Figure 1 (b)

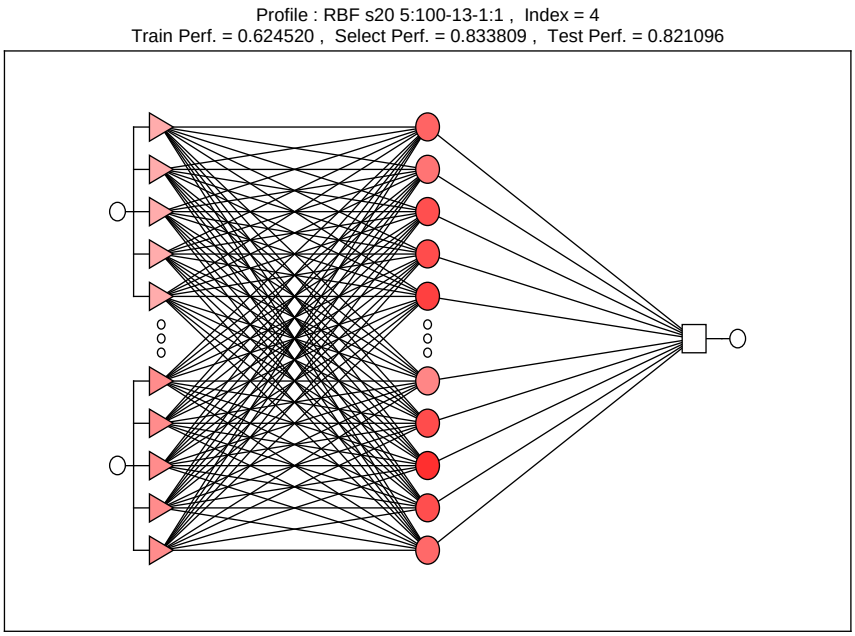


Figure 1 (c)

The inputs to the neural network are Measured Radon, Meteorological parameters Temperature, Rainfall, Relative humidity, and barometric pressure. The selection performance for both the MLP based neural network architectures is not satisfactory. Apart from MLP based neural networks radial basis neural networks are also tried. RBF networks have a number of advantages over MLPs. First, they can model any nonlinear function using a single hidden layer, which removes some design-decisions about numbers of layers. Second, the simple linear transformation in the output layer can be optimized fully using traditional linear modeling techniques, which are fast and do not suffer from problems such as local minima which plague MLP training techniques. RBF networks can therefore be trained extremely quickly. The comparison summary for different networks is presented in table 1.As shown in the table different training methods are employed for these two chosen architectures. For MLPs based neural networks Back Propagation (BP) and Conjugate Gradient descent (CG) are used. For radial basis neural network K-Means (KM) is used for centre assignment, K-Nearest Neighbor(KN) for deviation assignment and

Pseudo Invert (PI) is used for linear least squares optimization.

Index	Model Summary Report (traindata in nnmodeling.stw)							
	Profile	Train Perf	Select Perf	Test Perf	Train Error	Select Error	Test Error	Training
1	MLP s20 5:100-3-1:1	0.588539	0.796853	0.882128	0.125560	0.213866	0.187257	BP100,CG40
2	MLP s20 5:100-10-1:1	0.631787	0.798913	0.894424	0.141368	0.213025	0.188681	BP71
3	RBF s20 5:100-12-1:1	0.617619	0.843557	0.807473	0.064166	0.113283	0.085103	KM,KN,PI
4	RBF s20 5:100-13-1:1	0.624520	0.833809	0.821096	0.064883	0.113147	0.086312	KM,KN,PI
5	RBF s20 5:100-15-1:1	0.631014	0.812851	0.798642	0.065557	0.110153	0.084371	KM,KN,PI

Table 1

The K-means algorithm assigns radial centers to the first hidden layer in the network if it consists of radial units. K-means assigns each training case to one of K clusters (where K is the number of radial units), such that each cluster is represented by the centroids of its cases, and each case is nearer to the centroids of its cluster than to the centroids of any other cluster. It is the centroids that are copied to the radial units. The intention is to discover a set of cluster centers which best represent the natural distribution of the training cases. The radial basis function is indented to be used as a time series approximation wherein the input data represents data samples of certain past times and the network has only one output, which is the estimated value.

The chosen architecture of Radial Basis Function network is shown in Fig. 1 (c). The architecture is chosen based on the selection performance of different networks. The chosen network has five inputs which are Measured Radon, Meteorological parameters like Temperature, Rainfall, Relative humidity, and corrected barometric pressure. The RBF contained three hidden layers with 100, 13 and 1 hidden neurons and single output which is the estimated radon value. The estimation of radon was done for different time periods starting from 10 days going up to 360 days. The radon was predicted for the subsequent day of period selection. If 20 days data is fed to the network then the estimated radon value is for 21st day. The neural network estimated radon value is compared with the measured value to find out the anomaly.

Four cases are presented for the prediction comparison for the above described algorithm:

In the first case the estimation of radon was done over an annual period and the deviations from "raw" radon of the "neural predicted" radon was used to detect the anomaly. The "raw" refers to the actual measured data.

In the second case the estimation of radon was taken over a period corresponding to the seasons. The seasonal period selected offered better results, but it has a problem that the seasonal periods are manually selected for region, can vary from place to place and not amenable to automation.

In the third case the estimation of radon on a period obtained by applying FFT to the measured "raw", "corrected" data [Gupta et. al., 2007], and "neural predicted" data removing human and subjective factor out of the technique. This technique has the advantage that it can be applied automatically to the data of any location and is amenable to computerization and also showed best performance.

In the fourth case the estimation of radon was done on all other randomly varying periods. The results of all the above cases were compared with the results of statistically corrected radon results.

a) Results and Comparison of Proposed Neural Network algorithm

The predicted radon using the Radial Basis Function Network is plotted versus the measured radon for the June 96-May 97 in Fig. 2 and for June 97- May 98 in Fig. 3. It may be observed from the Fig. 2 and Fig. 3 that predicted radon using the neural network algorithm is following the trend of measured radon. This is not observed in case of sudden peaks which signify the precursor for an earthquake.

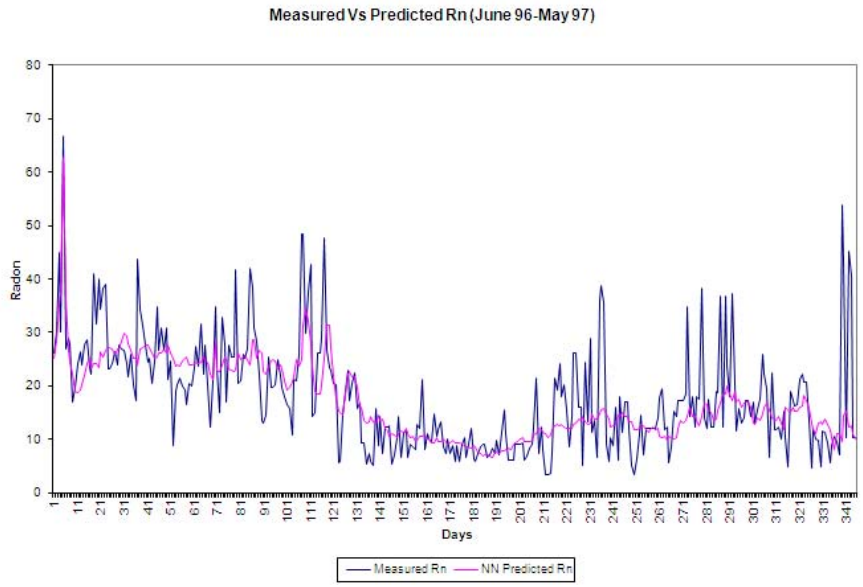


Figure 2

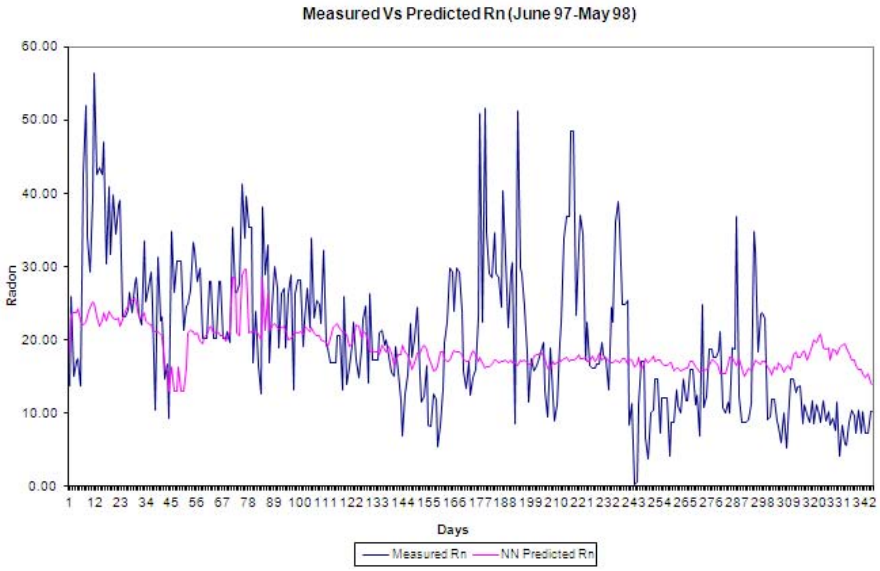


Figure 3

Case 1: In the first case the 360 days of data samples were fed to the neural network. It was observed in this case there were 18,19 and 19 true event predictions in the span of three years for the "raw", "corrected" and "neural predicted" radon out of the total 33 events (Table 2). The false anomalies were 28, 37 and 30 respectively for the "raw", "corrected" and "neural predicted" Radon respectively. The use of neural network for estimating the radon value has not made a significant impact on the prediction rate (Table 3). It was observed that there was no improvement in the event prediction i.e. true anomalies (TA) rate but there was a reduction in the false anomalies (FA). This analysis proved that the neural network was able to learn the meteorological parameter effect of radon, better than regression method used earlier.

Period	Raw TA/33	Corrected TA/33	NN TA/33
Annual	18	19	19
Seasonal	20	26	25
FFT 47	20	**27**	**28**
FFT 32	25	**27**	**29**

Table 2

Period	Raw FA	Corrected FA	NN FA
Annual	28	37	30
Seasonal	35	64	48
FFT 47	32	**23**	**18**
FFT 32	25	**21**	**17**

Table 3

Case 2: The Radon emanation is enhanced in summer months and is somewhat suppressed during winter. The seasons were divided as June-Sep, Oct-Jan, Feb-May, This selection was based on the assumption that June to September is the main rainy season in the area, October to January being the winter season and February to May being the mild summer season in that area. Thus the selected period was 120 days corresponding to the seasons starting from June-1996. It was observed in this case there were 20, 26 and 25 true event predictions in the span of three years for the "raw", "corrected" and "neural predicted" radon out of the total 33 events (Table 2). The false anomalies were 35, 64 and 48 respectively for "raw", "corrected" and "neural predicted" radon respectively (Table 3).

Case 3: In this case periodicity was taken corresponding to the periodicity worked out by FFT. The same has been discussed in detail in chapter 3.It was observed in the case of 47 days there was 20, 27 and 28 true event predictions in the span of three years for the "raw", "corrected" and "neural predicted" radon respectively and the false anomalies were 32, 23 and 15 respectively for the "raw", "corrected" and "neural predicted" radon respectively. However for 32-day period it was observed there were 25, 27 and 29 true event predictions in the span of three years for the "raw", "corrected" and "neural predicted" radon and the false anomalies were 25, 21 and 14 respectively for "raw", "corrected" and "neural predicted" radon respectively. It was observed that there was about 6% improvement in the event prediction rate i.e. true anomalies (TA) as compared to the statistically corrected radon. The false anomalies (FA) were also found to be further reduced [Gupta et. al., 2011].

Case 4: The above analysis represents three specific cases in which specific periods were taken which ranged from annual, seasonal and selection based on Fast Fourier transform technique. It was thought to consider all the time periods starting from 10 days to 360 days. The calculated anomalies were then plotted. Fig 4 shows the three kinds of anomalies for period varying from 10 days to 360 days. The values are calculated as a percentage of each anomaly over the total anomalies observed. It is observed from the graph that the prediction rate of the anomalies is highest in the range of periods defined by FFT also. This proves the using FFT technique to calculate the time period gives most effective results.

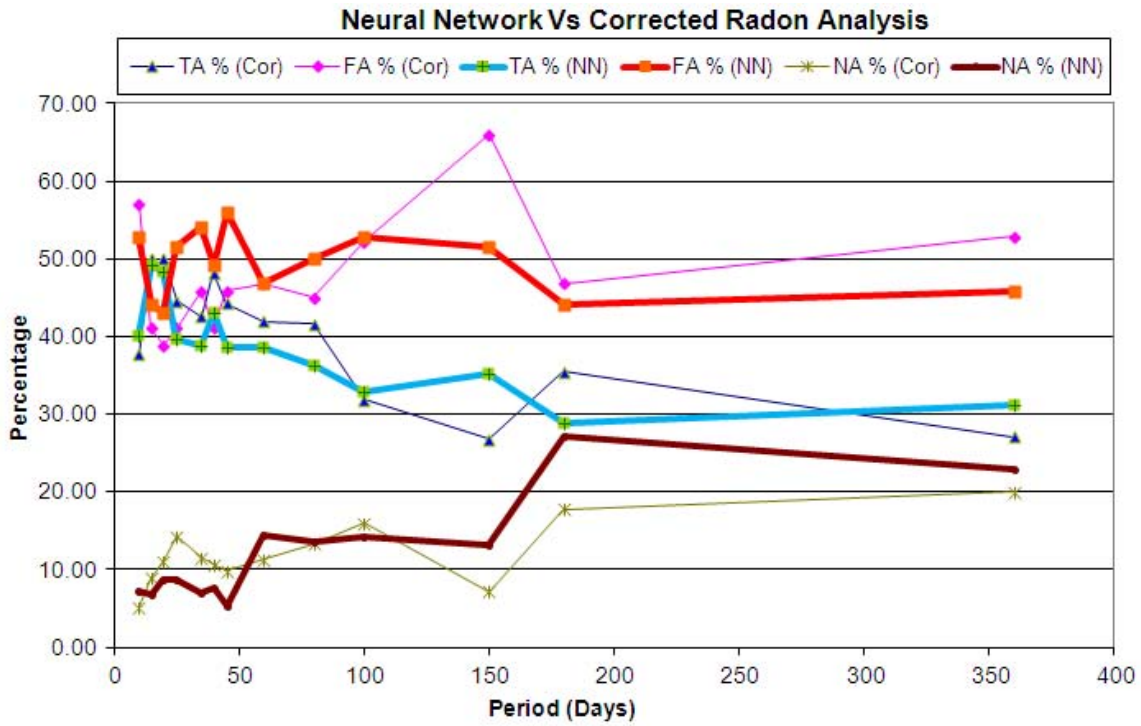


Figure 4

It is observed True Anomaly (TA) is best observed in the case of neural predicted radon as compared with meteorologically "corrected" radon and "raw" measured radon in all cases. Secondly computation of the mean and standard deviation over period given by FFT gives the best result both in high TA and low FA, as compared to the seasonal and annual periods. There is a significant improvement in false anomalies in case of FFT period-defined neural network analysis compared to other methods.

b) Neural Network Algorithm for Probabilistic Event Estimate

The neural network algorithm discussed above gives the radon estimates and further by using these estimates for finding out the anomalies has definitely

given better result as compared to statistical algorithm. There are two basic aspects that need to be improved in this algorithm. Firstly, there are huge numbers of false anomalies which are undesirable. Secondly, radon emanations depend on not only earthquake build up but many other geophysical activities. Also as neural networks have the ability to learn complex non linear patterns inside the data which may not be identified by any statistical approach. Hence, another algorithm of probabilistic estimation of earthquake events is experimented upon. In this algorithm probabilistic neural network architecture is chosen. The probabilistic neural network is predominantly a classifier which maps the input pattern to a number of classifications. As the models involve classification the regression of the data is not done. The measured radon values with meteorological parameters are presented a continuous input. The earthquake event was presented to the network as a categorical output. The duration period for these events was selected to be 10 days before an actual event [Zmazek et. al., 2005]. This not only increased the data set which otherwise is very minimalistic, but it also increased the span of probability output by the network. The chosen network is a probabilistic neural network. The chosen network is shown in the Fig. 5.The result for the above chosen network is presented in Fig. 9. It was observed that although there was not much improvement in the event identification i.e. true anomalies (TA) (Table 4) but the probabilistic neural network reduced the false anomalies (FA) to zero (Table 5).

Secondly, the output of the neural network is event estimation. The inputs presented to the neural network are measured radon and all the meteorological parameters. The primary advantage of this network is that raw measured radon may be presented to the network without any corrections. The neural network takes care of the met corrections on the radon.

Period	Raw TA/33	Corrected TA/33	NN TA/33	PNN TA/33
Annual	18	19	19	19
Seasonal	20	26	25	26
FFT 47	20	**27**	**28**	**28**
FFT 32	25	**27**	**29**	**29**

Table 4

Period	Raw FA	Corrected FA	NN FA	PNN FA
Annual	28	37	30	0
Seasonal	35	64	48	0
FFT 47	32	**23**	**18**	**0**
FFT 32	25	**21**	**17**	**0**

Table 5

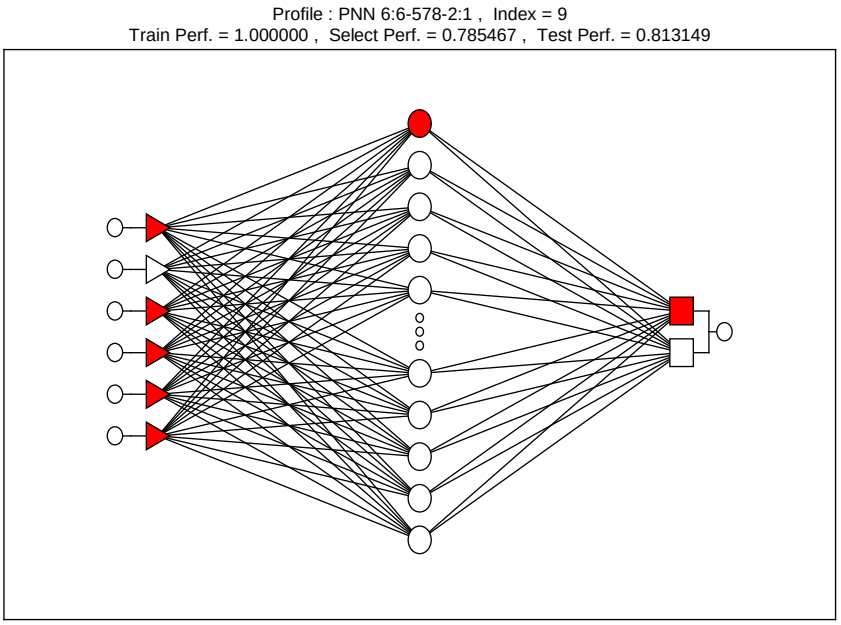


Figure 5

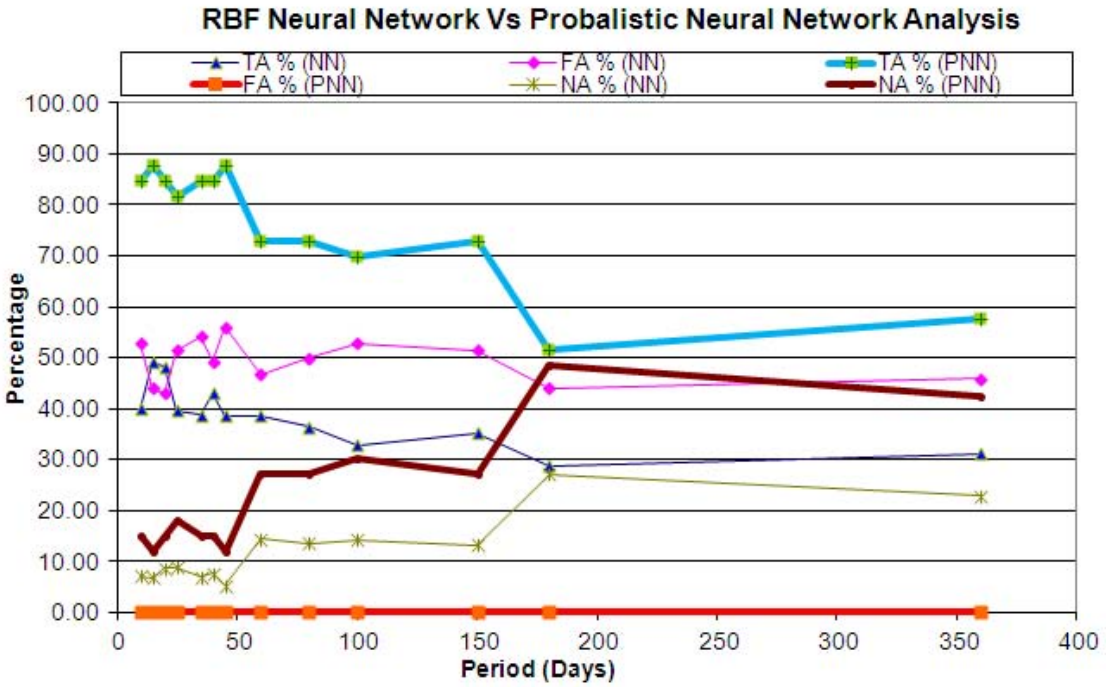


Figure 6

III. Conclusion

Emission of Radon is strongly influenced by day to day meteorological conditions as well as seasonal. Different authors have sought to tackle seasonal variations by normalizing the raw emission values over a local "time period" of observation (varying from days to may be few months or a season) as a way of tackling the periodic variations in the mean emitted value. The day to day meteorological influences on emitted radon have been tackled by some form of regression / corrections on raw data of emission based on measured meteorological parameters like humidity, temperature, pressure etc. There is no uniformity in the methods reported in literature to tackle daily and seasonal influences and no specific method of comparing efficacy of prediction is available.

Neural Network algorithm have been worked out by incorporation of FFT based time period and methods of regression. Additionally, Probabilistic Neural Networks that take all possible measured data (like emitted radon, meteorological conditions) as inputs and focus on event (earthquake) as final output, is also used wherein nonspecific time period or regression is required. The two algorithms are compared by using TA (True anomaly) and FA (False anomaly) on the same basic radon data and the improvement in prediction between the algorithms is clearly brought out.

In this paper it is shown that a period arrived at by applying FFT to annual radon emission data gives improved results. Further the day to day influences of meteorological conditions have been sought to be removed via neural network techniques.

1. It can be concluded that the use of neural networks for characterization and evaluation of radon anomalies gives improved results on account of their known ability to model more complex dependency. The paper has contributed by showing that better dependency modeling reduces FA. It not only shows the extent or scope that is there in improving physical models but also provides better prediction in the interim period as compared to statistical algorithm. The algorithm used automatically models meteorological parameter effects. **The event prediction i.e. true anomalies (TA) in this case showed an improvement of 6% as compared to statistical technique and it further reduces the false anomalies (FA).**
2. It can be concluded that probabilistic neural network (PNN) algorithm which directly gives event as an output from raw data on radon emission gives no false anomalies and event prediction is also at par with earlier neural network technique. Use of probabilistic neural network also shows that the threshold levels used in precursors also have a dependency that is not clearly understood, and hence the PNN by bypassing the simpler regression and threshold models gives lowest FA of all the three algorithms.
3. It can be concluded that algorithms proposed in this paper for earthquake predictive modeling has several advantages namely:

   i. The algorithms are also highly amenable to computerized implementation.

ii. They algorithms offer options of low to nil manual selection and/or specialized perception of the phenomenon.
iii. Due to (b) above they have better potential of being applied at newer locations.
iv. They automatically take into account regional average of the emitted radon and its day to day variations caused by non-tectonic phenomenon.

Acknowledgements

The authors acknowledge the financial assistance by Department of Science and Technology, New-Delhi, under Project No. DST/23 (595)/SU/2006.

References Références Referencias

1. Gupta Dhawal, Shahani, D.T., Khan Shakeb, 2007. Fast Fourier Transform based Earthquake Precursor Analysis of Radon in N-W Himalayas, Computer Measurements and Virtual Instrument Systems, International Conference on Electronic Measurements and Instruments 3, 1-5.
2. Gupta, D., Shahani, D.T., 2011. Estimation of Radon as an Earthquake Precursor: A neural network approach. Journal of Geological Society of India 78, 243-248.
3. Gupta, H.K., Singh, H.N., 1986. Seismicity of the North-Eastern Region. Journal of the Geological Society of India 28, 367-406.
4. Gupta, H.K., Mandal, P., Satyanarayana, H.V.S., Shashidhar, B., Shekhar, M., Singh, A., Devi U.E., Rao, P.C., Dimri, V.P., 2005. An Earthquake of M~5 may occur at Koyna, Current Science 89, 747-748.
5. H.S. Virk, Baljinder, 1993. Radon anomalies in soil gas and ground water as Earthquake Precursor phenomena. Tectonophysics, ELSEVIER Science Publishers 227, 215-217.
6. Igarashi, G., Saeki, S., Takhahata, N., Sumikawa, K., Tasaka, S., Sasaki, Y., Takahashi, M. and Sano, Y., 1995. Ground-Water radon anomaly before the Kobe earthquake in Japan, Science 269, 60-61.
7. J. Planinic, V. Radolic, Vukovic, B., 2004. Radon as an Earthquake Precursor. Nuclear Instruments and Methods in Physics Research, ELSEVIER Science Publishers 530, 568-574.
8. JIN, A., AKI, K., 1986. Temporal change in coda Q before the Tangshan earthquake of 1976 and the Haicheng Earthquake of 1975. Journal of Geophysical Research 91, 665–673.
9. Khattri, K.N., Wyss, M, 1978. Precursory variations of seismicity rate in Assam Area, India. Geology 6, 685-688.
10. Lomnitz, C., Lomnitz, L., Tangshan, 1978. A case history in earthquake prediction. Nature 171, 109-111.
11. Molchan, G.M., Dmitrieva, O.E., Rotwain, I.M., Dewey, J., 1990. Statistical analysis of the results of earthquake prediction based on the bursts of aftershocks. Journal of Physics of the Earth and Planetary Interiors 61, 128 – 138.
12. Srivastava, H.N., 1973. The crustal seismicity and nature of faulting near India, Nepal and Tibet tri-junction. Himalayan Geology 3, 381-393.
13. Srivastava, H.N., Rao, P.C.S., 1990. Spatio-temporal variation of seismicity preceding Bihar- Nepal and Manipur-Burma border earthquake of 1988. Bulletin of Indian Society of Earthquake Technology 28 (3), 13-22.
14. Stefansson, Ragnar., Maurizio, Bonafede., Gunnar, B. Gudmundsson., 2011. Earthquake-Prediction Research and the Earthquake of 2000 in the South Iceland Seismic Zone, Bulletin of the Seismological Society of America 101, 1590-1617.
15. Ulamov, V. I., Mavashev, B. Z., 1967. On fore-runners of strong tectonic earthquakes, Russian Academy of Sciences, Doklady 176, 319-322.
16. Varotsos, P., K. Alexopoulos., 1984. Physical propertied of the variations of the electric field of the earth preceding earthquakes, Tectonophysics, ELSEVIER Science Publishers 110, 73-125.
17. Wattananikorn, K., Kanaree, M., Wiboolsake, S., 1988. Soil gas radon as an earthquake precursor some consideration on data improvement. Radiation Measurements, ELSEVIER Science Publishers 29, 593–598.
18. Zmazek, B., Todorovski L., Dzeroski S., Vaupotic J., Kobal, I., 2003. Application of decision trees to the analysis of soil radon for earthquake precursor. Applied Radiation and Isotopes, ELSEVIER Science Publishers 58, 697-706.
19. Zmazek, B., Zivcic, M., Todorovski, L., Dzeroski, S., Vaupotic, J., Kobal, I., 2005. Radon in soil gas: How to identify anomalies caused by earthquakes. Applied Geochemistry, ELSEVIER Science Publishers 20, 1106-1119.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Neural & Artificial Intelligence
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Self Organization Map to Assess Forest Development and Problems with Multiple Regression Analysis

By Md Saiful Islam & Md Kamor Uddin Sikder
BGC Trust University, Bangladesh

Abstract - Sustainability is a burning factor in the forestry sector around the world. Understanding sustainability is also a critical question. This research examined the status of Bangladesh forestry by indicating the needs of forest products and the efficiency of forest policy and management practices, and relates sustainability with scarcity situations Bangladesh is a country of beauty with forest and river. The natural significant of Bangladesh is mainly depends on its forest. The greenness of our country is the gift of heaven. But we are not aware of this marvelous asset. In this work we have evaluated the status of forest of Bangladesh under the approach of soft computing. The method is Self Organization Map (SOM). We classified the data set by Multiple Regression Analysis (MRA). The study area we have designed the south part of the Bangladesh.

Keywords : *sustainability, multiple regression analysis (MRA), self organization map (SOM), greenness.*

GJCST-D Classification : *I.2.6*

Strictly as per the compliance and regulations of:

Self Organization Map to Assess Forest Development and Problems with Multiple Regression Analysis

Md Saiful Islam [α] & Md Kamor Uddin Sikder [σ]

Abstract - Sustainability is a burning factor in the forestry sector around the world. Understanding sustainability is also a critical question. This research examined the status of Bangladesh forestry by indicating the needs of forest products and the efficiency of forest policy and management practices, and relates sustainability with scarcity situations Bangladesh is a country of beauty with forest and river. The natural significant of Bangladesh is mainly depends on its forest. The greenness of our country is the gift of heaven. But we are not aware of this marvelous asset. In this work we have evaluated the status of forest of Bangladesh under the approach of soft computing. The method is Self Organization Map (SOM). We classified the data set by Multiple Regression Analysis (MRA). The study area we have designed the south part of the Bangladesh.

Keywords : sustainability, multiple regression analysis (MRA), self organization map (SOM), greenness.

I. Introduction

Poverty could occupy many dimensions in space (resource, economic,cultural), in persistence (periodic, fluctuating and lifelong) and in its identity (child, woman and poverty of elders). Peluso, Humphrey, and Fortmann (1994) explained the sustainability issues as a condition of resource poverty,and coined the term natural resource dependent areas (NRDA) to address the unsustainable situation of resource abundant and resource scarce areas together. By NRDA the authors meant the places where natural resources either account for a substantial part of the local economy or attract population. The NRDA concept reiterates that even if resources are there, the nature of the policy may influence the availability of resources to ordinary people, and could influence the poverty. Therefore, sustainability evaluation may need to address different forms of policy measures explaining how they influence resource availability and poverty. The approach of sustainability indicators considered here is compartmentalization of policy evaluation. By compartmentalization of policy this article emphasizes particular issues like the nature of resources, market oriented investment and commitment to sustained supply (e.g. Gaventa, 1980; Marchak, 1983; Peluso et al., 1994). However, natural resource dependence is not a prior cause of poverty and hence sustainability. There are some other causes, such as centralized economic structure (Bunker, 1984), technological inability (Blaikie, 1985; Freudenburg, 1992) and concentration of ownership and control (Marchak, 1983; Freudenburg, 1992) that may bring a sustainability risk to NRDA countries. Thus, policy discourses inevitably become linked with control of resources, such as, resource dependence, resource use, resource waste and nature of capital. Taking the present land use as the end result of past policies, an attempt is made to track the past social indicators of policy discourses. This study presents the resource scarcity situation of Bangladesh as links to those.

a) Issues of sustainability indicators

In Bangladesh about 16% of the land area is legal forest but mostly located in south-west (SW) and south-east (SE) corners of the country (Map 1). Bhuiyan (1994) reported that out of the forest areas, actual managed forest was only 9.2% (1.32 million ha) and Unclassed State Forest (USF) was 6.9% (0.99 million ha) in early 90s. WRI/CIDE (1990) estimated that the forest area would be only 1 million ha or 6.9% of total land area. Moreover, the cover intensity is different in different forest types (Table 2). An early report of Gittins and Akonda (1982) estimated remaining natural forest cover to be only 3.3% of total land, which is less than 0.5 million ha. Other than the mangroves and salt forests the distinct area patterns of forest ecotypes of Bangladesh are not well marked. Therefore, the spatial status of Bangladesh forests is often classified under legal types rather than ecotypes. A legal type may include different ecotypes and may be adjusted with the need of administration. As a result often there is a change in the space status of forests of Bangladesh. The following section presents the situation of changing forested space of Bangladesh. However, the most recent figure is stated by Muhammed, Koike, Sajjaduzzaman, and Sophanarith (2005) from unpublished data of the Forest Management Planning Database Survey 2003. They mentioned that the estimated forest area of the country is about 2.53 million ha and this is about 17.5% of the total land base of Bangladesh.

Author α : BGC Trust University Bangladesh Dept. of Computer Science and Engineering Id no : CSE 1014027.
E-mail : tuhincse507@gmail.com
Author σ : BGC Trust University Bangladesh Dept. of Computer Science and Engineering Id no : CSE 1014009.
E-mail : Kamoruddin00001@gmail.com

II. Data Collection

a) State of Costal Forest in Bangladesh

The coastal part of Bangladesh includes the famous Sundarbans Mangrove Forest. A number of depressed basins are found in the district of greater Mymensingh and Sylhet ,and some are found Chittagong which are inundated by fresh water during the monsoon that gradually dry out during the dry winter season. These depressed basins are known as 'Haor'.

b) Costal Forest in Chittagong

The coastal part of Chittagong includes the famous Sandip Mangrove Forest. A number of depressed basins are found in the district of greate Mirssorai, Gorokghata, Kutubdiya, Shitakundo, Chonuya, Banchkhali, Teknaf and some are found Urirchor. which are inundated by fresh water during the monsoon that gradually dry out during the dry winter season.

c) Costal Forest are three types

It is necessary to examine forest cover under different major categories of land, which are

i. Mangroves
ii. Non Mangroves
iii. Strip

i. *Mangroves*

Mangroves are a group of trees and shrubs that live in the coastal intertidal zone

Year of Garden Create	Mangrove (Hector)	Strip	Non Manograve
1990-91	80.97	--	40.48
1991-92	688.82	--	40.48
1992-93	688.37	--	60.73
1993-94	1044.86	3.00	139.67
1994-95	794.34	--	47.15
1995-96	310	--	384
1996-97	300	60.00	--
1997-98	600	--	--
1998-99	590	93	10

a. *Mangroves Forming a Wall Along Florida Inlets*

There are about 80 different species of mangrove trees. All of these trees grow in areas with lowoxygen soil, where slow-moving waters allow fine sediments to accumulate. Mangrove forests only grow at tropical and subtropical latitudes near the equator because they cannot withstand freezing temperatures.

Many mangrove forests can be recognized by their dense tangle of prop roots that make the trees appear to be standing on stilts above the water. This tangle of roots allows the trees to handle the daily rise and fall of tides, which means that most mangroves get flooded at least twice per day. The roots also slow the movement of tidal waters, causing sediments to settle out of the water and build up the muddy bottom.

Mangrove forests stabilize the coastline, reducing erosion from storm surges, currents, waves, and tides. The intricate root system of mangroves also makes these forests attractive to fishes and other organisms seeking food and shelter from predators.

b. *Mangrove Plantation in Chittagong*

Mangrove afforestation along the entire southern coastal frontier is an innovation of foresters. During 1960-61, Government undertook afforestation programme along the shore land of coastal districts. This initiative got mementum from 1980-81 with the aid of development partners and afforestation programs are extended over foreshore islands, embankments and along the open coasts. Since 1960-61 upto 1990-91, 142,835 hectare of mangrove plantations have been raised under a number of coastal afforestation projects. The present 2012 net area of mangrove plantation is 48466.97 hectare after losing some area due to natural calamities.

ii. *Non-Mangrove*

Table 1 : Forest Area in Bangladesh

Country / region	Reference year	Total area ('000 ha)						
		Land area						Inland water
		Forest			Other wooded land		Other land	
		Closed	Open	Plantation	Shrubs/Trees	Forest fallow		
Bangladesh	1996	720	0	232	105	17	11,943	1,383
% of Bangladesh		5.0	0	1.6	0.7	0.1	82.9	9.6

iii. *Strip*

Table 2 : Total Eco-type of the costal forest in Chittagong south

Total Eco-type at last-2012	Area in hectares
Mangrove	48466.97
Non- Mangrove	1211.95
Strip	4163.93

Source: Chittagong Coastal Forest Division of Bangladesh (2011-12)

d) Man-Made Mangroves, Coastal Afforestation

These mangrove plantations were established on newly accreted lands (mud flats) prior to formal declaration as Reserved Forest. The species used were mostly Keora (Sonneratia apeta) and Gewa (Excoecarai agallocha). Since these lands were not declared as "Reserved Forest" the Forest Act was not strictly applicable on them. Consequently the FD in mostcases failed to provide the required protection because of the land litigations and poor legal back up from other government agencies such as district administrations, police, etc. Many of the good coastal plantations established in Chittagong and Noakhali were lost to shrimp farms with the direct and indirect indulgence of other government officers such as DCs and land administration agencies. Ultimately the wish of the DCs prevailed since they are the most powerful actors and highly favored by ministers, members of parliament, etc. Thus many of the coastal plantations were devastated. Revilla (1998) during the FRMP inventory reported the following growing stock in the mangrove a forestation areas.

Table 3 : FRMP inventory results of coastal afforestation divisions

Description	Chittagong Coastal Afforestation Division
Area in hectares	20042
Sample size	408
Number of trees with DBH 15 cm and above	10
Basal area in m2 per hectare for trees with DBH 15 cm and above	0.29
Volume in m3 per hectare contributed by trees having DBH 15 cm and above	1.02
Poles per hectare	15228
Saplings per hectare	2202
Seedlings per hectare	373

Source: FD, Government of Bangladesh

e) Self Organization Map

A **self-organizing map** (**SOM**) or **self-organizing feature map** (**SOFM**) is a training Map that is trained using unsupervised learning to produce a uncertain result discredited representation of the input space of the training samples, called a **map.** Self-organizing maps are different from other artificial neural networks in the sense that they use a neighborhood function to preserve the topological properties of the input space. The figure 1 shows the impact of SOM implementation in this work.

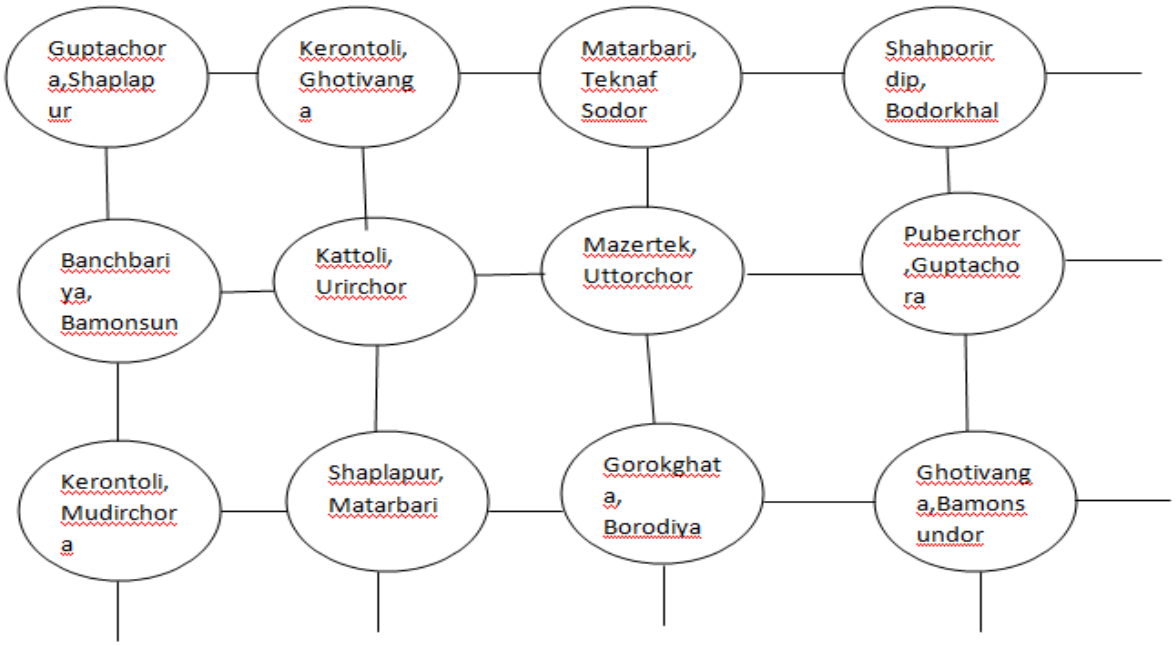
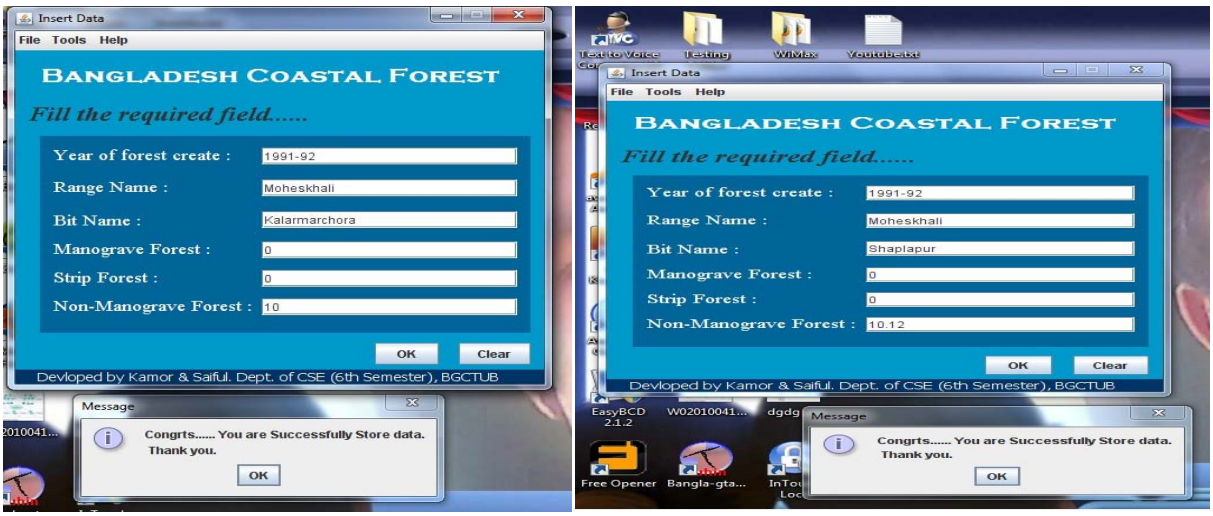
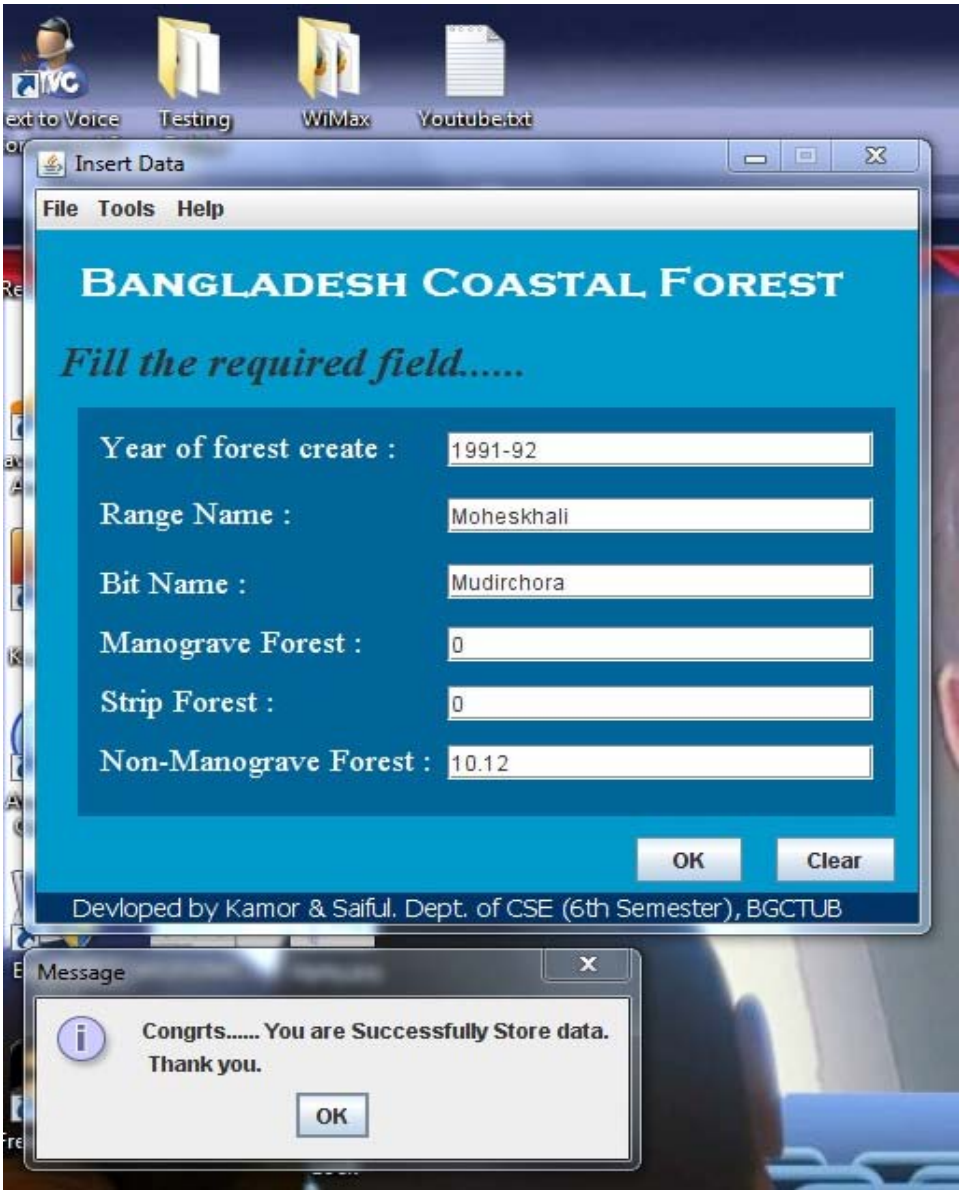


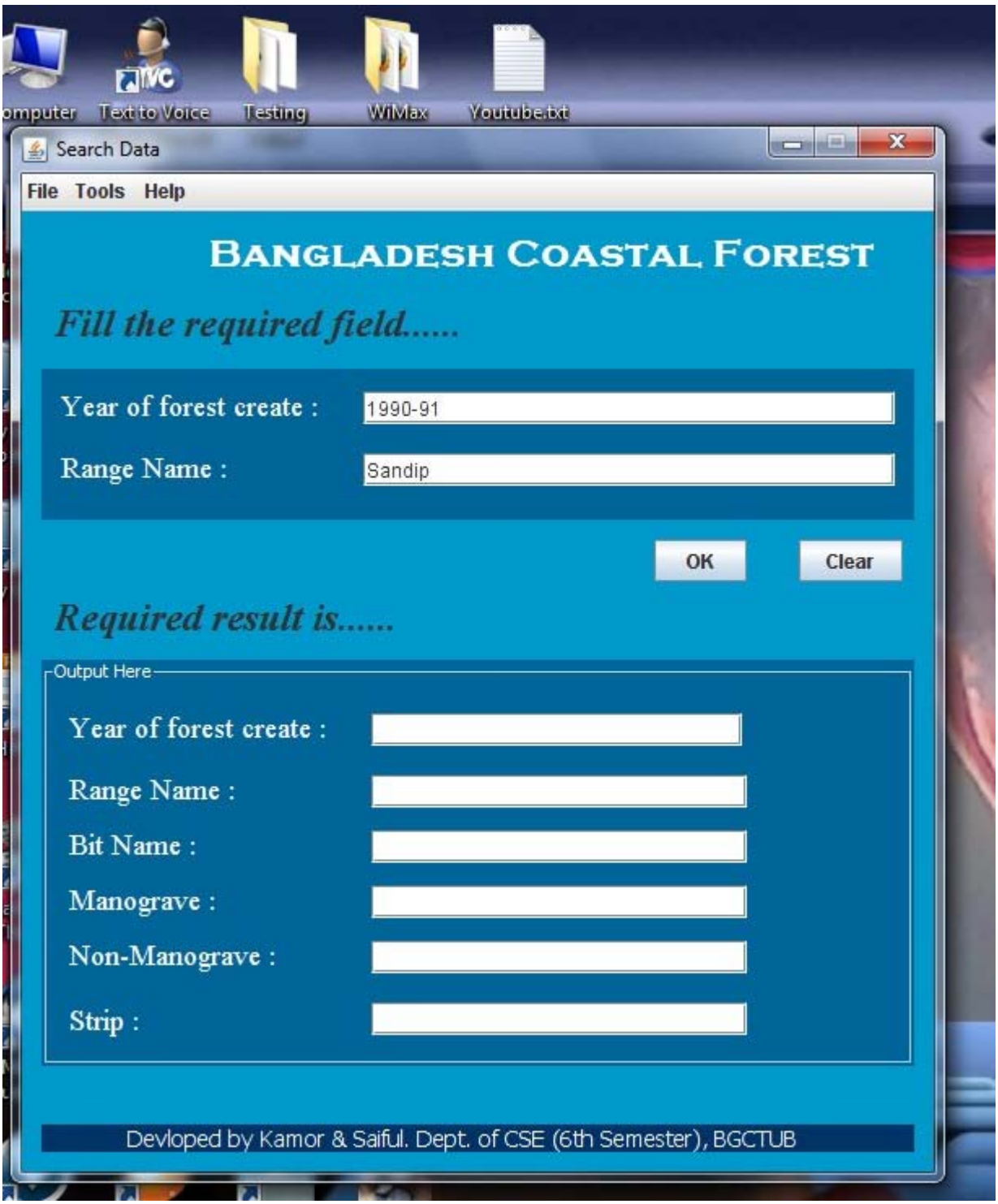
Figure 1 : The Self Organization Map

III. Result

In this work we have implemented the system based on the SOM and got the following result.







IV. Conclusion

This research in general reflected that the "Forests and Forestry in Bangladesh" (Forestry Sector) is experiencing problems. The current trends are in no way favorable for the overall development of the sector. By 2020 there may be a big lockage between the demand and supply of wood. Peoples' expectations from the FD will increase many fold, especially for forest based recreation, small wood supply, environmental parameters, peoples' participation, etc. Though most of the forest lands are managed by the FD the major supply of wood comes from homesteads. All possible lands especially the USF should be brought under proper management. The process of degradation should be stopped. The FD should be drastically reorganized on par with other Government administrative set ups. At the same time the

Government should allocate adequate funds for the forestry sector and encourage large scale social forestry programme.

References Références Referencias

1. Williams, M. (1994). Forests and tree cover. In W.B. Meyer & B.L. Turner (Eds), Changes in land use and land cover: a global perspective (pp. 97-124). Cambridge: Cambridge University Press.
2. Blaikie, P. (1985). The Political economy of soil erosion in developing countries. London: Longman.
3. Bryant, R. L. (1998). Power, knowledge and political ecology in the third world: a review. Progress in Physical Geography, 22(1), 79-94.
4. Bunker, S.G. (1984). Extractive economies in the Brazilian Amazon, 1600-1980. American Journal of Sociology, 89, 1017-1064.
5. Chaffey, D.R., Miller, F.R. & Sandom, J.H. (1985). A forest inventory of the Sunderbans, Bangladesh. Main Report, ODA, Land Resources Development Centre, England, Project Report No. 140.
6. Chittagong hill tracts: Soil and land use survey, vol. 5. (1966). Forestal Forestry and Engineering International Limited, Vancouver, Canada.
7. Choudhury, A.M. (1977). Objects of forest management in Bangladesh. In Proceedings of the first Bangladesh National Conference on Forestry,
8. February 11-15 (pp.66-69). Dhaka, Bangladesh.

This page is intentionally left blank

Global Journal of Computer Science and Technology
Neural & Artificial Intelligence
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Tumor Prediction in Mammogram using Neural Network

By Ms. P. Valarmathi & Dr. V. Radhakrishna
Anna University, India

Abstract - Detecting micro calcifications - early breast cancer indicators – is visually tough while recognizing malignant tumors is a highly complicated issue. Digital mammography ensures early breast cancer detection through digital mammograms locating suspicious areas with benign/- malignant micro calcifications. Early detection is vital in treatment and survival of breast cancer as there are no sure ways to prevent it. This paper presents a method of tumor prediction based on extracting features from mammogram using Gabor filter with Discrete cosine transform and classify the features using Neural Network.

Keywords : *mammograms, micro calcifications, gabor filter, discrete cosine transform, artificial neural network (ANN).*

GJCST-D Classification : *F.1.1*

Strictly as per the compliance and regulations of:

Tumor Prediction in Mammogram using Neural Network

Ms. P. Valarmathi [α] & Dr. V. Radhakrishnan [σ]

Abstract - Detecting micro calcifications - early breast cancer indicators – is visually tough while recognizing malignant tumors is a highly complicated issue. Digital mammography ensures early breast cancer detection through digital mammograms locating suspicious areas with benign/-malignant micro calcifications. Early detection is vital in treatment and survival of breast cancer as there are no sure ways to prevent it. This paper presents a method of tumor prediction based on extracting features from mammogram using Gabor filter with Discrete cosine transform and classify the features using Neural Network.

Keywords : mammograms, micro calcifications, gabor filter, discrete cosine transform, artificial neural network (ANN).

I. Introduction

Digital mammography and computer aided diagnostics ensure that physicians can take accurate decisions with regard to breast cancer. There was much progress recently in the development of computer aided systems to classify mammograms. Mammograms are breast region X-ray images revealing points with high intensity density which could potentially be a tumor. Thus early diagnosis and screening is crucial for successful treatment/cure. Usually, masses and calcium deposits are identified visually as such deposits are denser than the surrounding soft tissue. Malign tumors are associated with unusually smaller clustered calcification. Other calcification types that correspond to benign tumors are diffuse, regional, segmental or linear and they are termed micro calcification.

A mammogram is done through compressing the patient's breast between two acrylic plates and passing an X-ray signal through it. It is al gray scale image indicating details inside the breast through contrast. Such details can also be normal tissues, vessels, muscles, varied masses and noise. Every mass type has varied shape, size, distribution, and brightness acting as features to help a radiologist toe diagnose breast tumors effectively.

Author α : Pursuing Ph.D in Anna University, Professor, Department of Computer Science and Engineering, Mookambigai College of Engineering, Pudukkottai, Tamilnadu, India.
E-mail : goodmathi@yahoo.com
Author σ : Dean (Academic), Mookambigai College of Engineering, Pudukkottai, Tamilnadu, India. E-mail : prof.vrk@gmail.com

Mammograms with clustered micro-calcifications, mass lesions, breast architecture distortion and breast asymmetry have shown that they are linked to breast cancer. Micro calcifications are small, bright and arbitrarily shaped regions, whereas mass lesions are dense, have different size and properties and which are described as circumscribed, speculated or ill-defined [1, 2]. Circumscribed masses are usually uniform and smooth shaped like irregular circles. Speculated lesions are segments distributed as a multi armed star in many directions while ill-defined masses lack a specific pattern. Figure 1 shows examples of these features.

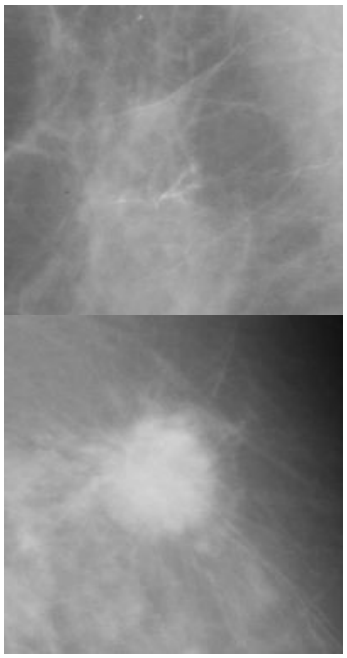


Figure 1 : Abnormal Mammograms

Currently, micro calcification detection is hard due to their fuzzy nature; low contrast and low distinguishing ability from ROS with their sizes ranging between 0.1-1.0 mm with the average being 0.3 mm. Micro calcifications shapes, distribution and size are varied. However it is hard to segment micro calcifications as they are surrounded by tissues [3]. Much research for various types of breast abnormalities was undertaken in the last two decades. Currently, computer aided mammogram detection systems for mass/micro calcification are used clinical routines like Image Checker and Second Look [4].

CAD system's general architecture includes image pre-processing, definition of region(s) of interest, features extraction and selection, and classification. Generally computer aided mammography techniques cover image enhancement, segmentation, detection and classification [4].

Various features were extracted for mammogram abnormalities. Masses feature extraction, [5] has been split into three categories, intensity features, shape features and texture features. The wavelet, fractal, statistical, and vision-models-based features are used for masses detection [1]. Cheng et al[3] summarized micro calcification detection features into individual micro calcification features, statistical texture features, multi-scale texture features and fractal dimension features. Classification methods classify suspicious mammogram areas into benign, malignant or normal tissue. Digital mammograms present classification techniques are common and similar to classification procedures in neural networks, Bayesian belief network, and K-nearest neighbor. Though it was demonstrated that both LDA and ANN (artificial neural network) classify masses well [5].

Image feature extraction is important in signal processing techniques preprocessing. Digital image features can be extracted directly from spatial data or from another space. Using a different space through special data transform like Fourier transform or wavelets transform could separate special data with specific characteristics. Detecting image texture features is difficult as such features are variable and scale-dependent.

An uncorrelated measurement should be investigated to transform the data into a different domain in designing an automated mammogram classifier. Mammogram classification requires a transform that uncorrelated data without losing the main characteristics of the image. Naturally discrete wavelets transform suit mammogram feature extraction. The idea of wavelets is explained by Daubechies (1992) [6] who said that wavelets are functions used to prevent other functions. This is called mother wavelet. A set of functions is generated by mother function translations and dilations.

Wavelet decomposition is through 2D wavelets transform application to an image producing a set of four different coefficients in every decomposition level. Three levels of 2D wavelets decomposition are illustrated in Figure 2 [7]. The produced coefficients are

- Low frequency coefficients (A).
- Vertical high frequency coefficients (V).
- Horizontal high frequency coefficients (H).
- High frequency coefficients in both directions (D).

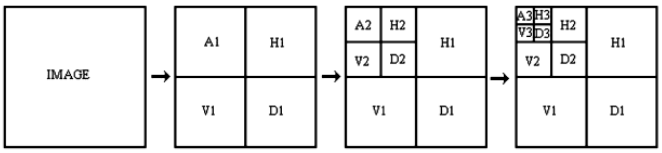


Figure 2 : Wavelets multi resolution decomposition

In this paper, the classification accuracy achieved for mammograms using Multi-Layer Perception Neural Network (MLPNN) is investigated. For predicting tumor, features are extracted from mammogram using Gabor filter with Discrete Cosine Transform (DCT). The rest of the paper is organized as follows: Section 2 reviews some the researches available in the literature; section 3 details the various techniques used in this study, section 4 reports the results and section 5 concludes the paper.

II. Related Works

Buciu et al [8] suggested an approach to deal with digital mammogram classification. Patches around tumors are manually extracted to segment abnormal areas from the rest of the image, considered as background. Gabor wavelets filter mammogram images and directional features extracted at various orientations/frequencies. Principal Component Analysis reduces filtered/unfiltered high-dimensional data dimensions. Proximal Support Vector Machines finally classify data. Superior mammogram image classification performance is attained after Gabor features extraction instead of using original mammogram images. Gabor features robustness for digital mammogram images distorted by Poisson noise of differing intensity levels is also addressed.

Eltoukhy, et al., [9] described a wavelet and curve let transform comparative study for breast cancer diagnosis. Mammogram images are decomposed into various resolution levels sensitive to various frequency bands through the use of multi-resolution analysis. A set of large coefficients is extracted from each decomposition level. Then based on Euclidian distance a supervised classifier system development is undertaken. Classifier performance is evaluated through a 2 X 5-fold cross validation, followed by a statistical analysis. The experiment's results reveal that curve let transform has a higher-throughput than wavelet transform with statistically significant difference.

Suganti et al [10] presented an automated system for breast tumor classification as either malign or

benign. It includes three stages: image enhancement and demising, multiple feature extraction techniques, and final classification stage. Three different classification schemes like ANNs, Support vector machines (SVMs) and Radial Basis Function (RBF) were used. The system was implemented and tested on classifier fusion techniques based on Majority Voting Methods and Behavior-Knowledge Space Method. Also SVMs were used for the first time for cluster characterization. Classifier performance was evaluated by using Receiver Operating Characteristic (ROC) methodology and classification rate. Results obtained results show high classification performance and so this method is quite promising.

Ayer et al. [11] revisited ANN models use in breast cancer risk estimation assessing both discrimination and calibration. Risk prediction was obtained using 10-fold cross-validation on a large data set of 62,219 consecutive mammography findings. ANN model achieved an AZ of 0.965, significantly higher than that of radiologists, 0.939 ($P<0.001$). ANN calibration assessed by Hosmer-Lemeshow (H-L) goodness-of-fit statistic test was 12.46 ($P>0.1$, $df=8$), indicating a good match between risk estimates and malignancy prevalence.

Islam et al [12] presented a computer aided mass classification method in digitized mammograms using Artificial Neural Network (ANN) and performing benign-malignant classification on region of interest (ROI) having mass. A major mass classification mammographic characteristic is texture. ANN exploits this to classify mass as benign or malignant. Statistical textural features in characterizing masses are mean, standard deviation, entropy, sleekness, kurtosis and uniformity. This method aims to increase classification process efficiency objectively to reduce many false-positive of malignancies. Three layers artificial neural network (ANN) with seven features was proposed to classify marked regions into benign or malignant achieving 90.91% sensitivity and 83.87% specificity which is promising compared to a radiologist's 75% sensitivity.

Cede no. [13] suggested improvements in neural network training for pattern classification with the proposed training algorithm being inspired by neuron's biological met plasticity property and Shannon's information theory. During training the Artificial metaplasticity Multilayer Perceptron (AMMLP) algorithm prioritizes updating weights for less frequent activations over those more frequent. This way metaplasticity is modeled artificially. AMMLP achieves better efficient training maintaining MLP performance. Wisconsin Breast Cancer Database (WBCD) is used to test the proposed algorithm. AMMLP performance is tested through classification accuracy, sensitivity and specificity analysis, and confusion matrix. AMMLP's 99.26% classification accuracy is promising compared to back propagation Algorithm (BPA) and recent classification techniques when applied to the same database.

Karabatak et al [14] presented an automatic diagnosis system to detect breast cancer based on association rules (AR) and neural network (NN). AR reduces breast cancer database dimensions in this study with NN being used for intelligent classfication. AR + NN system performance is compared with NN model with input feature dimension being reduced from nine to four through the use of AR. A 3-fold cross validation method was applied to Wisconsin breast cancer database to evaluate system performance in test stage. The proposed system's correct classification rate is 95.6% proving that AR could reduce feature space dimensions and that the AR + NN model can provide quick automatic diagnosis for other diseases

III. Material and Methods

a) Mammogram Database

Mammogram images used in experiments were from the Mammographic Image Analysis Society (MIAS) [15] and the 322 samples database was labeled as one of the three categories: normal, benign and malign. There are 208 normal images, 63 benign and 51 malign. Each 1024×1024 pixels image is centered. Abnormal cases are divided into six categories: micro calcification, circumscribed masses, speculated masses, ill-defined masses, architectural distortion and asymmetry. Coordinates of abnormality center are provided along with approximate radius (in pixels) of a circle enclosing abnormality for every abnormal case. The widest identified abnormality has a radius of 197 pixels, while tightest abnormality has a 3 pixel radius.

b) Gabor Wavelets

2D Gabor wavelets were much used in computer vision applications to model biological-like vision systems. Studies reveal that Gabor elementary functions suit modeling simple cells in visual cortex [16]. Other property is provided by optimal joint resolution in both space and frequency, suggesting simultaneous analysis in both domains. Gabor wavelet orientation property suits it for several applications, including image texture analysis or image retrieval [17]. A complex Gabor wavelet is a product of a Gaussian kernel with a complex sinusoid described as:

$$\psi_k(z)=\frac{k^T k}{\sigma^2}\exp\left(\frac{k^T k}{2\sigma^2}z^T z\right)\left(\exp\left(ik^T z\right)-\exp\left(-\frac{\sigma^2}{2}\right)\right)$$

where k and F_{PCA}^{kc} are characteristic wave vector:

$$k_v=2^{-\frac{v+2}{2}}\pi,\quad 1ex\varphi_\mu=\mu\frac{\pi}{8}$$

The parameters *v* and *μ*define a filter's frequency and orientation. Given an image I (z), a 2D Gabor wavelet transform is a convolution of this image I (z) with a family of Gabor filters and many orientation and frequency values:

$$I_k(z) = \int\int I(z')\psi_k(z - z')dz'$$

c) Discrete Cosine Transform

Orthogonal transforms are used in pattern recognition as it enables a noninvertible transformation from the pattern space to a reduced dimensionality feature space [18]. Thus, classification procedures are carried out with fewer features albeit with a small increase in classification error. Discrete Cosine Transform (DCT) converts time series signal into basic frequency components. On application of DCT an image is decomposed into a set of cosine basis functions. The DCT [19] of a list of n real numbers s(x), x = 0,..., n-1, is the list of length n given by:

$$S(u) = \sqrt{2/n}C(u)\sum_{x=0}^{n-1} s(x)\cos\frac{(2x+1)u\pi}{2n}$$

where $C(u) = 2^{-1/2}$ for u=0 or otherwise $C(u) = 1.$

The constant factors are chosen so that the basis vectors are orthogonal and normalized.

The inverse cosine transform (IDCT) is computed as follows:

$$S(x) = \sqrt{2/n}\sum_{x=0}^{n-1} C(u)s(u)\cos\frac{(2x+1)u\pi}{2n}$$

Where $C(u) = 2^{-1/2}$ for u=0 or otherwise $C(u) = 1.$

d) Artificial Neural Network (ANN)

Artificial Neural Network (ANN)are a collection of mathematical models imitating properties of biological nervous systems and functions of adaptive biological learning, made up of many processing elements highly interconnected with weighted links being similar to synapses. Unlike linear discriminates, ANNs use non-linear mapping functions as decision boundaries. ANN's advantage is their ability to self-learn, and often solve issues too complex for traditional techniques, or hard to find algorithmic solutions.

It includes input and output layers with one or more hidden layers between them. Depending on weight values of w(j, i) and w(k, j), inputs are r amplified/weakened to get a solution correctly. Determined weights train ANN using known samples. Generally, a known mammogram database with chosen features and desired results trains the ANN. After weights determination ANN can readily classify masses.

ANNs are computer models inspired by biologic neural network structures, consisting of interconnected nodes with their overall ability to predict outcomes being determined by intra neuron connections [20]. ANNs simulate neural processes by summing negative (inhibitory) and positive (excitatory) inputs to produce a single output [21]. Though ANNs differ in how neurons are connected and inputs processed, the focus is on "feedforward" networks, a commonly used ANN model in medical research.

Figure 3 illustrates ANN's generic structure consisting of node series in three layers (input, hidden, and output layers). Each input layer node is called an input node and represents an input variable (eg, an imaging feature like calcification/breast density) used as an outcome predictor. Output layer's single node (output node) represents predicted outcome (eg, malignancy probability). An inputs and output correspond to predictor variables and the outcome variable Y, respectively, in logistic regression models. Hidden layer nodes (hidden nodes) have intermediate values calculated by networks without any physical meaning. Hidden nodes allow ANN to model complex relationships between input variables and outcome.

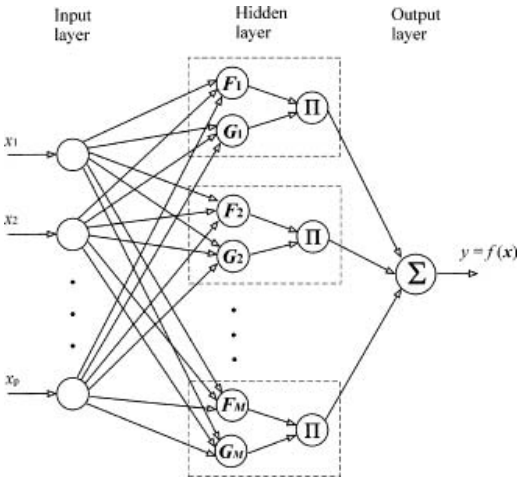


Figure 3 : Generic Structure of Artificial Neural Network

Different layer nodes are connected through connection weights, represented by arcs, containing "knowledge" representing relationships between variables, corresponding to coefficients in a logistic regression model. ANNs "learn" relationships between input variables and effects they have on outcome by strengthening (increasing) or weakening (decreasing) connection weight values through known cases basis. The optimal weight estimation process generating reliable outcomes is called learning/ training [22]. Many algorithms can train ANNs and the most popular is back propagation which in turn is based on the idea of adjusting connection weights to minimize discrepancy between real and predicted outcomes by propagating discrepancy in a backward direction (ie, from output node to input nodes). Table 1 gives the parameters of the ANN used in this study.

Table 1 : Parameters of the ANN

Number of input nodes	25
Number of outputs	2
Number of hidden layer	1
Number of neurons in hidden layer	10
Learning Algorithm	Back propagation algorithm
Learning rate	0.1
Momentum	0.5
Activation function	sigmoid /tanh/gaussian

IV. Results and Discussion

The performance efficiency of the ANN for different activation function for classifying the mammograms is investigated. The mammograms were classified as micro calcified and non-micro calcified. Features are extracted from the mammograms using Gabor filter with DCT. Mini MIAS containing 61 mammograms was used for evaluation. The following Table 2 shows the summary.

Table 2

	Naïve Bayes	Neural Network		
		Sigmoid Activation	Tanh Activation	Gaussian Activation
Correctly Classified Instances	38	55	56	58
Incorrectly Classified Instances	23	6	5	3
Kappa statistic	0.2549	0.8034	0.8359	0.9019
Mean absolute error	0.3692	0.14	0.1135	0.0874
Root mean squared error	0.5903	0.3098	0.277	0.2233
Relative absolute error	73.93%	28.03%	22.72%	17.50%
Root relative squared error	118.06%	61.96%	55.40%	44.66%
Coverage of cases (0.95 level)	70.49%	91.80%	95.08%	95.08%
Mean rel. region size	56.56%	68.85%	67.21%	68.85%
Total Number of Instances	61	61	61	61

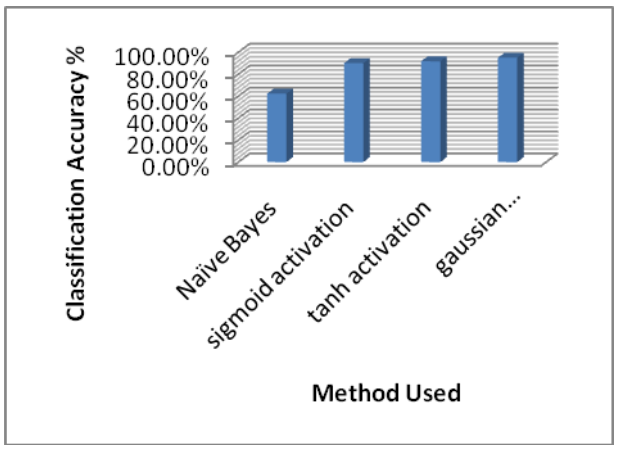


Figure 4 : Classification Accuracy Obtained by Various Activation Function

It is observed from Figure 4 that the ANN with Gaussian function achieves the maximum classification accuracy of 95.08%. Similarly, the RMSE is also the lowest for Gaussian function. Table 3 tabulates the precision, recall and f-measure of various methods. Figure 5 shows the graph of precision and recall.

Table 3 : Precision, Recall and F-Measure

	Precision	Recall	F-Measure
Naïve Bayes	0.64	0.623	0.617
Sigmoid Activation	0.904	0.902	0.902
Tanh Activation	0.919	0.918	0.918
Gaussian Activation	0.955	0.951	0.951

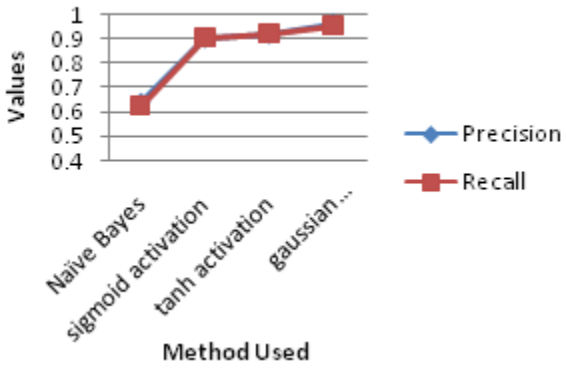


Figure 5 : Precision and Recall

The best precision and recall was achieved for ANN with Gaussian function.

V. Conclusion

Computer aided mammography was extensively studied. This research is mainly to detect and classify masses and micro calcifications. Techniques in computer-aided mammography include pre-processing, segmenting suspicious areas, extracting features, and classifying into benign, malignant or normal tissue. Different techniques-/algorithms were proposed or extended for digital mammograms, but reliable masses or micro calcification detection continues to be a challenge. This paper presents a method of tumor prediction based on extracting features from mammogram using Gabor filter

with Discrete cosine transform and classify the features using Neural Network. The efficiency of various activation functions for ANN is also investigated. Experimental results show that the Gaussian function achieves the best performance for classification.

References Références Referencias

1. Christoyianni, I., et al., Computer aided diagnosis of breast cancer in digitized mammograms. Computerized Medical Imaging and Graphics, 2002. 26(5): p. 309-319.
2. Wei, C.-H., C.-T. Li, and R. Wilson. A General Framework for Content-Based Medical Image Retrieval with its Application to Mammograms. in Proc. SPIE Int'l Symposium on Medical Imaging. 2005.
3. Cheng, H.D., et al., Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition, 2003. 36(12): p. 2967-2991.
4. Rangayyan, R.M., F.J. Ayres, and J.E. Leo Desautels, A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs. Journal of the Franklin Institute, 2007. 344: p. 312-348.
5. Cheng, H.D., et al., Approaches for automated detection and classification of masses in mammograms. Pattern Recognition, 2006. 39: p. 646-668.
6. Daubechies, I. (1992). Ten lectures on wavelets (Vol. 61). Philadelphia, PA: Society for industrial and applied mathematics.
7. Rashed, E. A., Ismail, I. A., &Zaki, S. I. (2007). Multiresolution mammogram analysis in multilevel decomposition. Pattern Recognition Letters, 28(2), 286-292.
8. Buciu, I., & Gacsadi, A. (2011). Directional features for automatic tumor classification of mammogram images. Biomedical Signal Processing and Control, 6(4), 370-378.
9. Meselhy Eltoukhy, M., Faye, I., &Belhaouari Samir, B. (2010). A comparison of wavelet and curvelet for breast cancer diagnosis in digital mammogram. Computers in Biology and Medicine, 40(4), 384-391.
10. Suganthi, M., &Madheswaran, M. (2010, February). An enhanced decision support system for breast tumor identification in screening mammograms using combined classifier. In Proceedings of the International Conference and Workshop on Emerging Trends in Technology (pp. 786-791). ACM.
11. Ayer, T., et al. (2010). Breast cancer risk estimation with artificial neural networks revisited. Cancer, 3310-3321.
12. Islam, M. J., Ahmadi, M., & Sid-Ahmed, M. A. (2010). An efficient automatic mass classification method in digitized mammograms using artificial neural network. arXiv preprint arXiv: 1007.5129.
13. Marcano-Cedeño, A., Quintanilla-Domínguez, J., &Andina, D. (2011). WBCD breast cancer database classification applying artificial metaplasticity neural network. Expert Systems with Applications, 38(8), 9573-9579.
14. Karabatak, M., &Ince, M. C. (2009). An expert system for detection of breast cancer based on association rules and neural network. Expert Systems with Applications, 36(2), 3465-3469.
15. Mammographic Image Analysis Society, http://www.wiau.man.ac.uk/services/MIAS/MIASweb.html.
16. S. E. Grigorescu, N. Petkov, P. Kruizinga, Comparison of Texture Features Based on Gabor Filters, IEEE Trans. on Image Processing 11 (10)(2002)1160–1167.
17. J.Han, K.-K.Ma, Rotation-invariant and Scale-invariant Gabor Features for Texture Image Retrieval, Image and Vision Computing 25 (9) (2007) 1474–1481.
18. Ahmed, N., Natarajan, T., &Rao, K. R. (1974). Discrete cosine transform. Computers, IEEE Transactions on, 100(1), 90-93.
19. N. Ahmed, T. Natarajan, Discrete-Time Signals and Systems, Reston Publishing Company, 1983.
20. Hagan MT, Demuth HB, Beale M. Neural network design. Boston, Mass: PWS, 1997.
21. Guerriere MR, Detsky AS. Neural networks: what are they? Ann Intern Med 1991; 115: 906–907.
22. Haykin S. Neural networks: a comprehensive foundation. Upper Saddle River, NJ: Prentice Hall, 1998.

Global Journal of Computer Science and Technology
Neural & Artificial Intelligence
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Normalized Vector Codes for Object Recognition using Artificial Neural Networks in the Framework of Picture Description Languages

By G.D. Jasmin & E.G. Rajan
Mysore University, Hyderabad

Abstract - Your Understanding how biological visual systems recognize objects is one of the ultimate goals in computational neuroscience. People are able to recognize different types of objects despite the fact that the objects may vary in view, points, sizes, scale, texture or even when they are translated or rotated. In this paper we focus on syntactic approach for the description of objects as Normalized Vector Codes using which objects are recognized based on their shapes.

Keywords : *pattern recognition, formal representation of images, object recognition.*

GJCST-D Classification : *H.5.0, C.1.2*

NORMALIZED VECTOR CODES FOR OBJECT RECOGNITION USING ARTIFICIAL NEURAL NETWORKS IN THE FRAMEWORK OF PICTURE DESCRIPTION LANGUAGES

Strictly as per the compliance and regulations of:

Normalized Vector Codes for Object Recognition using Artificial Neural Networks in the Framework of Picture Description Languages

G.D. Jasmin [α] & E.G. Rajan [σ]

Abstract - Your Understanding how biological visual systems recognize objects is one of the ultimate goals in computational neuroscience. People are able to recognize different types of objects despite the fact that the objects may vary in view, points, sizes, scale, texture or even when they are translated or rotated. In this paper we focus on syntactic approach for the description of objects as Normalized Vector Codes using which objects are recognized based on their shapes.

Keywords : pattern recognition, formal representation of images, object recognition.

I. Introduction

Pattern recognition could be formally defined as categorization of input data into identifiable classes via extraction of significant features or attributes of the data from the background of irrelevant detail. A pattern class is a category, determined by some given common attributes. When a set of patterns falling into disjoint classes are available, it is desired to categorize these patterns into their respective classes through the use of some automatic device. It is important to note that learning or training takes place only during the design (or updating) phase of a pattern recognition system. Once acceptable results have been obtained, the system is engaged in the task of actually performing recognition on samples drawn from the environment in which it is expected to operate. The main objective of this research is to investigate and develop a general approach formally from a new theory based on Structural or Syntactic Pattern Recognition.

The problem of pattern recognition is divided into the following sub problems:

1. Image pre processing
2. Object Description
3. Classification

The pre-processing also known as low level image processing is performed on the input image to improve the quality of the image and to simplify the image for further processing. The pre-processing steps involve noise removal, conversion of gray scale and colour images into binary images and the extraction of contour from the binary image. The object description module takes an input contour image and gives a vector of direction and length called a knowledge vector as output. This plays an important role in the whole process as the knowledge vector gives more information about the objects present in the image which is used to characterize the pattern. We use a syntactic approach for the description of objects. The knowledge vector obtained here gives information about the direction and the length of lines in the contour of objects which can then be given as input to a classifier module. The normalization of this vector plays an important role in the classification process. The vector has to be analyzed and normalized in such a way that it better suits for more variance of objects. The classification module takes the normalized vector as input and identifies them as a member of one of the predefined classes depending on the set of attributes that they hold. The design of a classifier involves the selection of the classifier and the estimate of parameters for the classifier. The feed-forward artificial neural networks with back-propagation learning algorithm could be used for classification. Details of such neural networks could be obtained from standard literature.

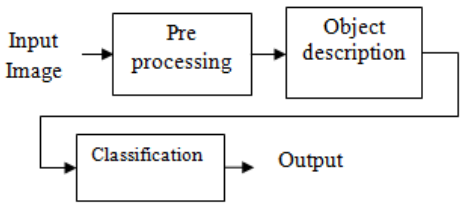


Figure 1 : Description Object Recognition System

a) Pre-Processing

To improve the quality of the noisy images we get from the real time environment image enhancement and restoration techniques can be used. Pixel neighbour processing techniques such as low pass filtering, high pass filtering or median filtering are used for the removal

Author α : Research Scholar in the Department of Computer Science, MG-NIRSA, Mysore University, Hyderabad.
E-mail : jassanth@gmail.com
Author σ : Director, ISIT, MG-NIRSA, Hyderabad, Affiliated to Mysore University. E-mail : rajaneg@yahoo.co.in

of noise in the image. The pixel point processing techniques such as binary contrast enhancement, stretching or shrinking and histogram equalization are used to improve the contrast of the image.

The input image to the system may be a colour image, a gray scale image or a binary image. These images have to be converted to binary images for further processing. Segmentation is an important process in our object recognition system as it highlights the objects present in the image from its background. The function im2bw() in matlab is used to convert the images to binary. The graythresh() function computes a global threshold (level) that can be used to convert an intensity image to a binary image with im2bw. Level is a normalized intensity value that lies in the range [0, 1].

The contours are extracted as they give the outlines of shapes of the objects present in the image. The function bwtraceboundary() in matlab is used for finding the contours. The initial point of the object is identified, and the outline tracing algorithm is used. bwtraceboundary() traces the outline pixels of an object in binary image bw. Nonzero pixels belong to an object and zero pixels constitute the background. Figure 2 shows a sample binary image and its contour.

(a) (b)

Figure 2 : a. Binary Image & b. Contour

b) Object Description

The notion of 'Picture Description Languages' refers to the language of digital images defined over a regular array of vertices. Having extracted the contour of an image this process involves running through the contour pixel by pixel and performing some form of calculations using every pixel and its surrounding pixels. This involves finding the length of the contour in every possible direction. The eight possible directions are two horizontal directions, two vertical directions and four diagonal directions. As the lines other than straight lines have more variations in their directions, the length of the descriptor becomes large. It is then reduced to a single occurrence of eight directions and normalized to fixed number of pixels for making it invariant. Following is the knowledge vector of a hexagon.

<69,98>/R64*D1*DR1*D2*DR1*D1*DR3*D1*DR1*D1*DR1*D2*DR1*D1*DR1*D1*DR1*D1*DR1*D1*DR3*D2*DR1*D1*DR1*D1*DR1*D1*DR1*D1*DR1*D1*DR1*D2*R1*DR1*D1*DR1*D1*DR1*D1*DR1*D1*DR1*D2*DR1*D1*DR3*D1*DR1*D1*DR1*DL1*D1*DL1*D2*DL3*D1*DL1*D1*DL1*D1*DL1*D1*DL1*D2*DL1*D1*DL3*D1*DL1*D1*DL1*D1*DL1*D2*DL1*D1*DL1*D1*DL3*D1*DL1*D2*DL1*D1*DL1*D1*DL1*D1*DL1*D1*DL1*D1*DL1*D1*DL2*D1*DL1*D1*DL1*D1*DL2*L63*U1*UL1*U1*UL1*U1*UL1*U2*UL1*U1*UL3*U1*UL1*U1*UL1*U1*UL1*U2*UL1*U1*UL1*U1*UL3*U1*UL1*U1*UL1*U2*UL1*U1*UL1*U1*UL1*U1*UL3*U1*UL1*U2*UL1*U1*UL1*U1*UL1*U1*UL2*U1*UL1*U1*UL1*UR1*U1*UR3*U1*UR1*U2*UR1*U1*UR1*U1*UR1*U1*UR1*U1*UR3*U2*UR1*U1*UR1*U1*UR1*U1*UR1*U1*UR1*U1*UR1*U2*UR3*U1*UR1*U1*UR1*U1*UR1*U1*UR1*U2*UR1*U1*UR3*U1*UR1*U1*UR1*U1*UR1*U2*UR1*/<70,98>#

This can then be reduced to single occurrence of eight directions as

<69,98>/R64+1*DR31*D58*DL33*L63*UL32*U57*UR32*/<70,98># and then normalized to □R17*DR16*DL17*L17*UL16*UR17

So, the final vector is a normalised vector consisting of a single occurrence of eight directions with their length, which can then be fed into an artificial neural network for further process of classification.

II. Generation of Direction Length Code (DLC)

The connectivity between the pixels is an important concept used in the field of object recognition. Two pixels are said to be connected, if they are adjacent in some sense and their intensity levels satisfy a specified criterion of similarity. If pixel p with coordinates (x,y) and pixel q with coordinates (s,t) are pixels of object (image subset) S, a path from p to q is a sequence of distinct pixels with coordinates

(x0, y0), (x1, y1),……., (xn, yn)

where (x0, y0) = (x,y) and (xn, yn) = (s,t), (xi, yi) is adjacent to (xi-1, yi-1), $1 \leq i \leq n$, and n is the length of the path. If there a path exists from p to q consisting of pixels in S, then p is said to be connected to q. The gap of one or two pixels marks the presence of the next component in the same object or the beginning of the next object which can be then found by analysing the relationship among the components.

With this background, we shall see now as to how to apply Look Ahead Tracer (LAT) algorithm to various image types and extract knowledge from them. Let (x,y) be the initial point. The algorithm continues to find the immediate pixel connected to (x,y) in one of the 8 directions. In this technique, a previously recognized direction during a scan is given the first priority status when compared to the other preferential ones. If R is the identified direction, then the other preferential directions are DR, D, DL, L, UL, U, UR.

The trace continues until there is no pixel connectivity to the current point (x,y), after removing the

pixels which are already traced, thus producing a knowledge string component.

By 'knowledge', we mean the direction code and length code of each contour component in the image.

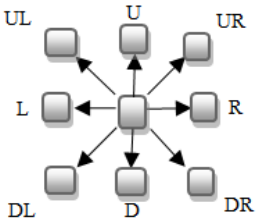


Figure 3 : Direction Codes

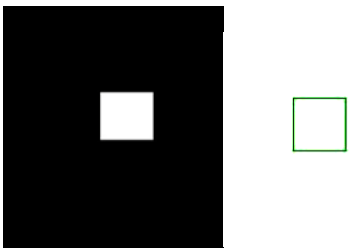
The vector data consists of both the Direction Code (DC) and the Length Code (LC). The direction data is a string of the elements from the set V_N, that is the set of non-terminals. The string elements are separated by a delimiter, which is denoted by the symbol □. An arbitrary square can be represented by the vector code in the following manner:

R□D□L□U*n□n□n□n

In a vector code, the part preceding the symbol * is called direction code and the part following the symbol * is called the length code. The symbol n represents the length of the sides of the square. For utility point of view, here we represent a knowledge about a square as <x1,y1>/Rn*Dn*Ln*Un*/<x2,y2># instead of R□D□L□U*n□n□n□n. <x1,y1> refers to the coordinates of the starting point; the delimiter * denotes change in direction; <x2,y2> refers to the coordinates of the end point of the contour component.

For example, let us consider a sample image of a square and its contour shown in figure 4. Now, the knowledge vector of a square is

<104,109>/R81*D81*L81*U80*/<105,109>#



(a) Image of a Square (b) Contour Map

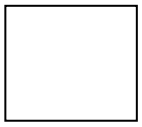
(c) <104,109>/R81*D81*L81*U80*/<105,109>#

Figure 4 : The generation of knowledge vector of a square

One can easily say that the above knowledge vector represents a square from the fact that the four directions R,D,L and U has the same length except the direction U that has 1 pixel less than other 3 sides. This is because of the immediate removal of the scanned pixels. The starting pixel at position <104,109> is removed after its scan.

III. Knowledge Vector Analysis

The direction codes in the knowledge vector obtained by tracing the contour of an object are R, DR, D, DL, L, UL, U and UR. For better processing, let us categorize these directions into basic directions R,D,L and U and diagonal directions DR,DL,UL and UR. For example, consider the knowledge vector of a square in figure 5.



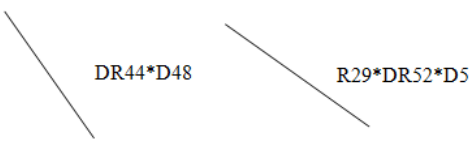
R100*D100*L100*U99

Figure 5 : The knowledge vector of a square

This shows the square consists of four basic directions with all lengths equal. Shapes with all the basic directions but with equal alternate length can be identified as a rectangle. Shapes with four sides consisting of only diagonal directions can be considered as rhombus.

The task of identifying regular shaped objects is simple. Objects in a real time environment may not be regular always. For the effective identification of those objects, the shapes of the objects can be approximated to some regular polygons to which it is closer to. This is possible by predefining the classes of regular shapes with some pre conditions set and approximating the other shapes to one of these shapes by identifying them as a member of the predefined class.

It is important to note that, the diagonal sides present in the shape of an object do not necessarily appear alone as the connection of only diagonal pixels. The occurrence of other pixels depends on the angle at which the diagonal lines appear in the image. Possibly D and R pixels appear more with DR, D and L pixels appear more with DL, U and L pixels appear with UL and U and R pixels appear with UR. For example consider a line with the direction code DR. As shown in fig: if the line bends more towards right the more R pixels appear, in the same way if it bends more towards down the more D pixels present with DR as shown in figure 6. In this respect, when the lineis rotated with a small change in the slope, they can be approximated to one of these shapes.

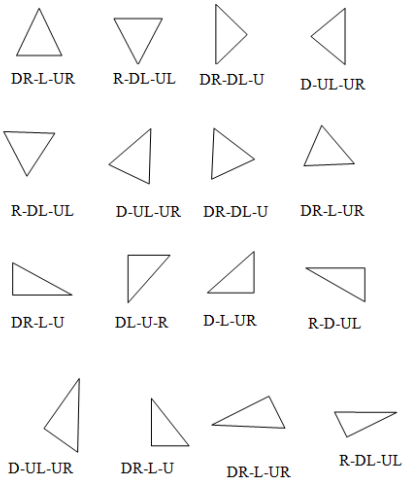


(a) Line with DR (b) Line with direction code

Figure 6 : Slant lines with small changes in the slope

a) Three Sided Shapes

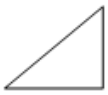
Listed below are the three sided shapes rotated in different angles with their direction codes.



Three sides with one basic side and two diagonal directions or two basic and one diagonal direction or three sided shapes that do not satisfy the above said conditions can be approximated to one of these shapes. For instance, let us consider the rotated triangle with no basic sides.

DL-UL-UR

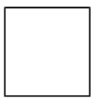
Could be approximated to



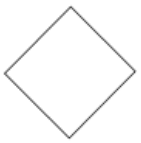
D-L-UR

b) Four Sided Shapes

- Square



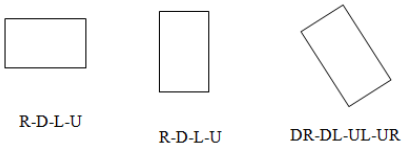
R-D-L-U



DR-DL-UL-UR

Figure 10 : Slant lines with small changes in the slope

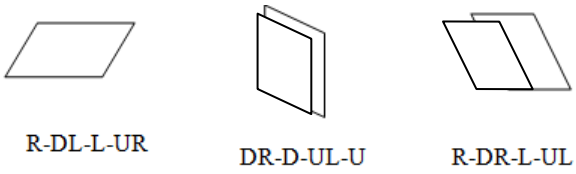
Four basic sides or four diagonal sides with equal lengths.



Four basic sides or four diagonal sides with equal alternate lengths.

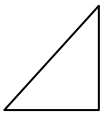
In the same manner, shapes with more sides can be predefined. When the shape of an object appears it can be then classified by artificial neural networks as a member of one of these classes.

- Parallelogram



c) Vector Normalization

This section shows the approximation and the normalization of the Knowledge vector of different shapes to produce the input vector for aneural network.



The knowledge vector obtained is

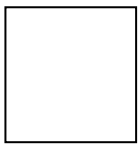
<71,167>/D109*L96*UR3*R1*UR2*U1*UR2*U1*UR10*U1*UR4*U1*UR6*R1*UR1*U2*UR6*U1*UR4*R1*UR2*U2*UR7*U1*UR3*R1*UR1*U1*UR2*U1*UR3*UR8*U1*UR4*U1*UR5*R1*UR2*U2*UR6*U1*UR4*R1*UR1*U1*UR1*U1*UR2*/<72,166>#

Vector limited to single occurrence of eight directions

<71,167>/R6*D109*L96*U19*UR89*/<72,166>#

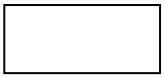
Vector approximated and normalized to 100 pixels

➔ D34*L30*UR36



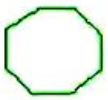
<104,109>/R81*D81*L81*U80*/<105,109>#
<104,109>/R81*D81*L81*U80*/<105,109>#

➔ R25*D25*L25*U25



<92,72>/R104*D63*L104*U62*/<93,72>#/<273,254>#
<92,72>/R104*D63*L104*U62*/<93,72>#/<273,254>#

➔ R31*D19*L31*U19



<86,102>/R25*D1*DR7*D1*DR1*R1*DR6*D1*DR2*D24*DL1*L1*DL5*L2*D1*DL4*L1*DL3*D1*DL1*L2*DL1

*L25*UL2*U1*UL7*U1*UL3*L1*UL4*U28*R1*UR2*R1*UR1*R2*UR5*R1*UR2*R1*UR1*R2*UR2*/<87,102>#

Vector limited to single occurrence of eight directions

<86,102>/R25+9*DR16*D24+5*DL15*L25+7*UL16*U28+2*UR13*/<87,102>#

Vector normalized as

➔ R14*DR13*D14*DL12*L14*UL11*U15*UR10

<81,138>/R11*DR1*R4*DR1*R2*DR1*R2*DR4*R1*DR2*R1*DR5*D1*DR3*D1*DR2*D1*DR1*D1*DR1*D2*DR1*D2*DR1*D3*DR2*D3*DR1*D16*DL1*D4*L1*DL1*D3*DL1*D2*DL1*D2*DL1*D1*DL1*D1*DL1*D1*DL4*D1*DL5*L1*DL2*L1*DL4*L3*DL1*L1*DL1*L5*DL1*L10*UL1*L4*UL1*L2*UL1*L1*UL3*L1*UL1*L1*UL3*L1*UL4*U1*UL1*U1*UL1*U1*UL1*L1*UL1*U1*UL1*U2*UL1*U1*UL1*U3*UL1*U3*UL1*U3*UL1*U15*UR1*U4*UR1*U3*UR1*U3*UR1*U1*UR1*U2*UR1*U1*UR1*R1*UR1*U1*UR1*U1*UR1*U1*UR4*R1*UR3*R1*UR2*R2*UR3*R2*UR1*R4*/<82,137>#

Vector limited to single occurrence of eight directions

<81,138>/R32*DR26*D45*DL25*L33*UL24*U48*UR23*/<82,137>#

Vector normalized as

➔ R4*DR20*D6*DL20*L4*UL21*U6*UR20

IV. The Short Comings of Traditional Pattern Recognition Systems and the Need for Adaptive Processing

Objects in the real world from an image or image sequence of the world are detected with the advanced software tools the imaging applications like Automatic Object Recognition (AOR) that uses different object models which are known a-priori. Humans can perform this task of object recognition effortlessly and instantaneously. However, such a task of implementation based on machines is performed using algorithms and has been very difficult in networks for real time imaging problems which are non-linear in nature. In some applications, the patterns/objects to be recognized are so fuzzy that they cannot be modelled with conventional tools. In order to solve these neural network processors can be used as the best tool because of the fact that the processor can build a recognition engine from simple image annotations made by the programmer. It then extracts the characteristics or feature vectors from the annotated objects and sends them to the neural network. Neural networks having adaptive processing elements are capable of performing generalization and can consequently classify situations never seen before by associating them to similar learned situations.

a) Artificial Neural Networks

Learning from a set of examples is an important attribute needed for most pattern recognition systems. Artificial neural network is an adaptive system being widely used in pattern recognition systems that changes its structure based on external or internal information that flows through the network. Artificial neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The neural network design part consists of two processes, training and application. The training of the neural network continues until the mean squared error reduces to a certain threshold or until the maximum number of iterations is reached. Once training is completed, the network can be applied for the actual classification of the data. The classification technique used may be one of the following:

1. Supervised classification - in which the input pattern fall as a member of a predefined class.
2. Unsupervised classification - in which the pattern falls into an unknown class as there are no predefined classes.

The learning or training takes place only during the design phase of a pattern recognition system. Once the results obtained are satisfactory, the system is ready to perform the task of recognition on samples drawn from the environment in which it is expected to operate.

b) Feed-forward neural networks

Feed-forward networks are commonly used for pattern recognition. A three-layer feed forward neural network is typically composed of one input layer, one output layer and one hidden layer. In the input layer, each neuron corresponds to a given input pattern while in the output layer each neuron corresponds to a predefined pattern. Once a certain sample is input into the network, he best situation is that the output will be a vector with all elements as zero only except the one corresponding to the pattern that the sample belongs to. Of course, it is very complex to construct such types of neural networks. The commonly used networks for minimizing the cost are multi-layer-feed-forward neural networks, which uses the back-propagation learning algorithm for training neural networks.

i. *Back-Propagation Algorithm*

Multi-layer feed-forward networks have been used as powerful classifiers. The training of back-propagation algorithm involves 4 stages

1. Initialization of weights
2. Feed forward stage
3. Back propagation of errors
4. Updating of weights and bases

During the feed-forward stage each layer in the network calculates its activation value and passes it to

the layers in the next level. The neurons in the output layer produce the output of the network and compare it with the target output to determine the error. During the back-propagation stage the error is back propagated from the output layer to each layer in the previous level for the correction of the adjustable parameters. This process repeats until the error reaches a minimum threshold or until the maximum number of iterations performed. The Back propagation algorithm is discussed below.

Parameters used

$x = (x_1, x_2, \ldots\ldots, x_i, \ldots, x_n)$ - Input training vector
$t = (t_1, t_2, \ldots\ldots\ldots, t_k, \ldots, t_m)$ - Output target vector
δ_k = error at output unit y_k
δ_j = error at hidden unit z_j
α = learning rate
v_{ij} = weights of input layer
v_{oj} = bios on hidden unit j
z_j = activation of hidden unit j
w_{jk} = weights of hidden layer
w_{ok} = bios on output unit k
y_k = activation of output unit k

1. Initialize the weights to small random values
2. While the stopping condition is false, do steps 3 to 10
3. For each training pair do steps 4 to 10
4. Each input receives the input signal x_i (i=1,…..,n) and transmits it to all units in the hidden layer z_j (j=1,….,p).
5. Each hidden unit z_j sums its weighted input signals

$$z_{-inj} = v_{oj} + \sum_{i=1}^{n} x_i v_{ij}$$

and applies its activation function

$$z_j = f(z_{-inj})$$

and sends this signal to all units in the output layer y_k (k=1,….,m).

6. Each output unit y_k (k=1,…,m) sums its weighted input signals

$$y_{-ink} = w_{ok} + \sum_{j=1}^{p} z_j w_{jk}$$

and applies its activation function to calculate the output signals

$$y_k = f(y_{-ink})$$

7. Each output unit y_k receives a target pattern t_k corresponding to an input pattern and calculates the error term as

$$\delta_k = (t_k - y_k) f'(y_{-ink})$$

8. Each hidden unit z_j sums its delta inputs from units in the layer above. The error information term is calculated as,

$$\delta_j = \sum_{k=1}^{m} \delta_k w_{jk} f'(y_{-ink})$$

9. Each output unit y_k updates its bias and weights. The change in weight is given by

$$\Delta w_{jk} = \alpha \delta_k z_j$$

And the bias correction term is given by

$$\Delta w_{ok} = \alpha \delta_k$$

Therefore,

$$w_{jk}(new) = w_{jk}(old) + \Delta w_{jk}$$

and

$$w_{ok}(new) = w_{ok}(old) + \Delta w_{ok}$$

10. The hidden unit z_j updates its bias and weights. The weight correction term is given by

$$\Delta v_{ij} = \alpha \delta_j x_i$$

and the bias correction term is

$$\Delta v_{oj} = \alpha \delta_j$$

Therefore,

$$v_{ij}(new) = v_{ij}(old) + \Delta v_{ij}$$

and

$$v_{oj}(new) = v_{oj}(old) + \Delta w_{oj}$$

11. Test the stopping condition.

V. Results

Neural network with 8 neurons in the input layer and 9 neurons in the output layer each representing 9 different classes of shapes is trained with a set of input patterns. The fastest back-propagation algorithm known as Levenberg-Marquardt back propagation (trainlm) present in matlab toolbox is used to train the neural network. *trainlm* is a network training function that updates weight and bias values according to Levenberg-Marquardt optimization. Fig. 5.7. shows the neural network created for our case of problem. Fig. 5.8 shows the performance of the neural network for the given set of training and test patterns.

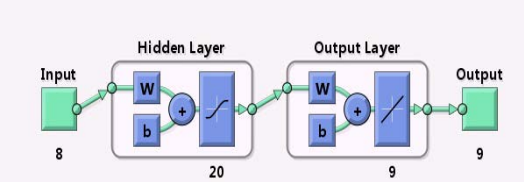


Figure 7 : Neural Network with 8 neurons in the input layer and 9 neurons in the output layer

Table 1 shows nine object classes based on regular geometric polygons such as triangle, rectangle, square, parallelogram, pentagon, hexagon, heptagon, octagon and nonagon which is approximated as circle.

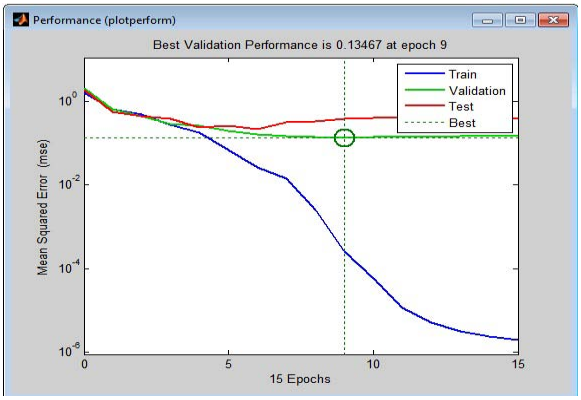


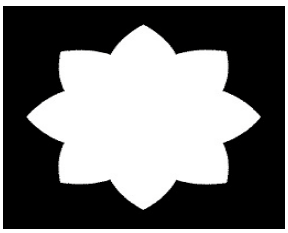
Figure 8 : Neural Network Performance

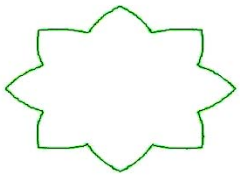
Table 1 : Nine basic classes recognized by the neural network

	R	DR	D	DL	L	UL	U	UR	Class
1.	0.0	0.0	34.0	0.0	30.0	0.0	0.0	36.0	1
2.	33.0	0.0	31.0	0.0	0.0	36.0	0.0	0.0	1
3.	28.0	0.0	25.0	0.0	25.0	0.0	25.0	0.0	2
4.	25.0	0.0	25.0	0.0	25.0	0.0	25.0	0.0	2
5.	31.0	0.0	19.0	0.0	31.0	0.0	19.0	0.0	3
6.	30.0	0.0	0.0	21.0	28.0	0.0	0.0	20.0	4
7.	0.0	20.0	32.0	0.0	0.0	20.0	29.0	0.0	4
8.	0.0	25.0	0.0	13.0	23.0	14.0	0.0	24.0	5
9.	0.0	25.0	0.0	25.0	0.0	13.0	23.0	14.0	5
10.	0.0	17.0	17.0	16.0	0.0	17.0	17.0	16.0	6
11.	17.0	16.0	0.0	17.0	17.0	16.0	0.0	16.0	6
12.	0.0	17.0	17.0	11.0	11.0	12.0	16.0	16.0	7
13.	17.0	11.0	11.0	12.0	16.0	16.0	0.0	16.0	7
14.	14.0	13.0	13.0	12.0	14.0	11.0	15.0	10.0	8
15.	15.0	11.0	13.0	12.0	13.0	11.0	14.0	11.0	8
16.	6.0	20.0	4.0	21.0	6.0	20.0	4.0	19.0	9
17.	0.0	38.0	0.0	0.0	25.0	0.0	0.0	37.0	1
18.	0.0	25.0	0.0	19.0	12.0	19.0	0.0	25.0	5
19.	0.0	37.0	0.0	0.0	33.0	0.0	31.0	0.0	1
20.	17.0	8.0	18.0	8.0	17.0	8.0	18.0	7.0	8
21.	0.0	25.0	0.0	25.0	0.0	25.0	0.0	25.0	2
22.	23.0	14.0	0.0	25.0	0.0	25.0	0.0	13.0	5
23.	5.0	20.0	2.0	20.0	5.0	20.0	2.0	20.0	9
24.	2.0	20.0	5.0	20.0	2.0	20.0	5.0	20.0	9
25.	0.0	17.0	16.0	17.0	0.0	17.0	16.0	17.0	6
26.	16.0	17.0	0.0	17.0	16.0	17.0	0.0	17.0	6
27.	0.0	17.0	16.0	17.0	0.0	17.0	16.0	17.0	6
28.	15.0	17.0	11.0	12.0	14.0	13.0	10.0	12.0	8
29.	0.0	25.0	0.0	15.0	19.0	15.0	0.0	25.0	5
30.	19.0	16.0	0.0	25.0	0.0	25.0	0.0	15.0	5
31.	16.0	0.0	35.0	0.0	16.0	0.0	34.0	0.0	3
32.	0.0	35.0	0.0	0.0	32.0	0.0	33.0	0.0	1
33.	14.0	0.0	33.0	0.0	0.0	35.0	0.0	0.0	1
34.	0.0	0.0	31.0	0.0	33.0	0.0	0.0	35.0	1

35.	31.0	3.0	12.0	3.0	31.0	3.0	12.0	3.0	8
36.	0.0	28.0	0.0	0.0	44.0	0.0	0.0	28.0	1
37.	0.0	0.0	45.0	0.0	0.0	28.0	0.0	27.0	1
38.	21.0	0.0	29.0	0.0	21.0	0.0	29.0	0.0	3
39.	39.0	0.0	0.0	30.0	0.0	30.0	0.0	0.0	1
40.	27.0	0.0	23.0	0.0	27.0	0.0	23.0	0.0	3
41.	4.0	20.0	6.0	20.0	4.0	21.0	6.0	20.0	9
42.	11.0	13.0	15.0	12.0	11.0	12.0	14.0	12.0	8
43.	0.0	15.0	20.0	15.0	0.0	25.0	0.0	25.0	5
44.	35.0	0.0	16.0	0.0	35.0	0.0	15.0	0.0	3
45.	33.0	0.0	0.0	35.0	0.0	0.0	32.0	0.0	1
46.	12.0	3.0	31.0	3.0	12.0	3.0	31.0	3.0	8
47.	45.0	0.0	0.0	28.0	0.0	0.0	0.0	0.0	1

In the same way the DLC of real time images also can be approximated and normalized to fixed number of pixels. Such code gives us the input vector that can be given to the neural networks for the classification of their shapes as the shape in one of the 9 classes. Later more analysis can be done for further classification of these shapes inside the classes to produce more detailed classes for the actual identification of objects.





<64,184>/R4*DR2*R2*D1*DR1*R2*DR4*R2*DR1*R1*DR1*D1*DR1*R2*DR1*D1*DR4*R1*DR2*D1*DR4*R1*DR1*D2*DR3*D1*DR2*R1*DL1*DR1*D1*DR1*R2*UR1*R3*UR1*R5*UR1*R5*UR1*R4*UR1*R32*DR1*UR1*R2*DR1*R6*D6*DR1*DL1*DR1*D23*DL1*D5*L1*DL1*D3*DL1*D4*DL3*DR4*R2*DR1*R2*DR1*R2*DR1*R2*DR2*R2*DR1*R1*DR1*D1*DR1*R1*DR1*R2*DR2*R1*DR6*R1*DR2*R1*DR1*D1*DR1*R2*D1*DR1*D1*DR1*R1*DR1*D1*DR1*R1*DR1*DL6*D1*DL1*L2*D1*DL4*L1*DL2*L1*DL1*D1*DL2*L1*DL2*L2*DL1*L1*DL4*L2*DL2*L2*DL1*L2*DL2*L2*DL1*L3*D1*DL2*DR1*D2*DR1*D3*DR1*D3*R1*DR1*D4*DR1*D19*R1*DR1*DL2*D10*DL1*D3*L5*DL1*L13*DL1*UL1*L23*UL1*L4*UL1*L5*UL1*L4*UL1*L3*UL1*L1*DL2*D2*L1*DL1*D1*DL3*D1*DL4*D1*DL4*L1*DL1*D1*DL1*L1*DL3*DL3*L1*DL2*L1*DL1*L2*DL1*D2*DL1*L2*DL1*L1*DL1*L1*DL1*L1*DL1*UL2*L2*UL1*L1*UL1*L2*UL1*U2*UL1*L2*UL2*L1*UL2*L1*UL5*L1*UL2*U1*UL1*L1*UL1*U1*UL1*L1*UL1*U1*UL1*L1*UL1*U1*UL2*U1*UL1*U1*UL1*U1*UL4*DL2*L2*DL1*L5*DL1*L5*DL1*L6*DL1*L29*UL1*L1*DL1*UL1*L8*U5*UL1*U1*UR1*UL1*U24*UR1*U4*UR2*D1*U3*UR1*U4*UR3*U2*UL2*L2*UL1*L2*UL1*L2*UL2*L4*UL2*L1*UL1*U1*UL1*L2*UL2*L1*L2*UL1*L1*UL1*U2*UL1*L2*UL1*L1*UL1*U1*UL4*U1*UL1*L2*U1*UL1*L1*UL1*U3*UR5*R1*UR1*U1*UR1*R1*UR1*U1*UR2*R1*UR5*R1*UR1*R2*UR2*R1*UR3*R1*UR2*R1*UR1*R2*UR1*R2*UR2*R2*UR1*R2*UR2*R1*UR1*U1*UL1*U2*UL2*U3*UL1*U3*UL2*U5*UL1*U20*UR1*U8*UR1*R11*UR1*R26*DR1*R7*DR1*R5*DR1*R5*DR1*R3*DR1*R2*UR3*D1*U2*UR2*U1*UR3*U1*UR4*R1*UR1*U2*UR1*R1*UR2*U1*UR1*R1*UR1*U1*UR1*R1*UR2*R1*UR2*R2*UR1*R1*UR1*U1*UR2*R1*UR1*R2*UR1*/<65,183>#<64,184>/R26+160*DR74*D19+91*DL82*L29+153*UL74*U20+94*UR77*/<65,183>#

➔ R3*DR22*D2*DL23*L3*UL22*U2*UR23
Is identified as a member of shape class circle.

VI. Conclusion

In this Paper a syntactic approach is proposed for shape based object recognition. The knowledge vector is reduced to the input vector to a neural network by some vector approximation and normalization processes. The research efforts during the last decade have made significant progresses in both theoretical development and practical applications. The method presented here may offer a promising solution for object recognition problem.

References Références Referencias

1. Belongie, S., Malik, J. and Puzicha, J. "Shape matching and object recognition using shape contexts", IEEE Transactions on Pattern Analysis and Intelligence, Vol. 4(24), pp. 509-522, April 2002.
2. Cohen, F. S. and Wang, J.-Y. "3-D recognition and shape estimation from image contours", in Proc. 1992 IEEE Conf Computer Vision Pan. Recognition., June 1992.
3. Rajan, E. G. "Cellular Logic Array Processing Techniques for High- Throughput Image Processing Systems", Invited paper, SADHANA, Special Issue on Computer Vision, The Indian Academy of Sciences, Vol.18, Part-2, pp 279-300, June 1993.
4. Ping Chen, Zhaohui Fu, Andrew Lim and Brian Rodrigues "Two-Dimensional Packing For Irregular Shaped Objects" Proceedings of the 36th Hawaii International Conference on System Sciences – 2003, 0-7695-1874-5/03 $17.00 (C) 2003 IEEE.
5. Greg Mori, member, IEEE, Serge Belongie, member, IEEE, and Jitendra Malik, senior member, IEEE, "Efficient shape matching using shape

contexts", IEEE Transactions on pattern analysis and machine intelligence, vol. 27, no. 11, November 2005.

6. Hafiz T. Hassan, Muhammad U. Khalid and Kashif Imran "Intelligent Object and Pattern Recognition using Ensembles in Back Propagation Neural Network" International Journal of Electrical & Computer Sciences IJECS-IJENS Vol: 10 No: 06.
7. Rowley H. A Baluja S, and Kanade T, "Neural network-based face detection", IEEE Transactions on Pattern Analysis and Machine intelligence. 20 (1998), 23–38.

Global Journal of Computer Science and Technology
Neural & Artificial Intelligence
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Evaluation of Job Offers using the Evidential Reasoning Approach

By Tanjim Mahmud, Kazi Namirur Rahman & Dr. Mohammad Shahadat Hossain
University of Chittagong, Chittagong

Abstract - The word 'Job' term as a regular activity performed in exchange for payment is considered as one of the most important activities for many families worldwide .Evaluation is necessary when more than one opportunity come to an To fulfill their desired goal, it is the 'evaluation' which assesses among the factors. In addition, it is difficult to measure qualitative factors in a quantitative way, resulting incomplete-ness in data and hence, uncertainty. Besides it is essential to address the subject of uncertainty by using apt methodology; otherwise, the decision to choose a job will become inapt. There exist many methods name as Analytical Hierarchical Process (AHP), Analytical Network Process (ANP) and so on. But the mentioned methods are not suitable to address the subject of uncertainty and hence, resulting inappropriate selection to the expecting job. Therefore, this paper demonstrates the application of a novel method named Evidential Reasoning (ER), which is capable of addressing the uncertainty of multi-criterion problem, where there exist factors of both subjective and objective nature. The ER method handles uncertainties by using a belief structure is aggregating degrees of belief from lower level factors to higher level factors.

Keywords : *multiple criteria decision analysis (MCDA), uncertainty, evidential reasoning (ER) and analytical hierarchy process (AHP).*

GJCST-D Classification : *I.2.6*

Strictly as per the compliance and regulations of:

Evaluation of Job Offers using the Evidential Reasoning Approach

Tanjim Mahmud [α], Kazi Namirur Rahman [σ] & Dr. Mohammad Shahadat Hossain [ρ]

Abstract - The word 'Job' term as a regular activity performed in exchange for payment is considered as one of the most important activities for many families worldwide .Evaluation is necessary when more than one opportunity come to an individual personality. Then it requires the job offer evaluation. To fulfill their desired goal, it is the' evaluation' which assesses them well. This involves many factors to be measured and evaluated. These factors are expressed both in objective and subjective ways where as a hierarchical relationship exists among the factors. In addition, it is difficult to measure qualitative factors in a quantitative way, resulting incompleteness in data and hence, uncertainty. Besides it is essential to address the subject of uncertainty by using apt methodology; otherwise, the decision to choose a job will become inapt. There exist many methods name as Analytical Hierarchical Process (AHP), Analytical Network Process (ANP) and so on. But the mentioned methods are not suitable to address the subject of uncertainty and hence, resulting inappropriate selection to the expecting job. Therefore, this paper demonstrates the application of a novel method named Evidential Reasoning (ER), which is capable of addressing the uncertainty of multi-criterion problem, where there exist factors of both subjective and objective nature. The ER method handles uncertainties by using a belief structure is aggregating degrees of belief from lower level factors to higher level factors.

Keywords : multiple criteria decision analysis (MCDA), uncertainty, evidential reasoning (ER) and analytical hierarchy process (AHP).

I. Introduction

When we attempt to evaluate of job offers, it involves multiple criterions such as, location, salary, job content, long-term prospects, safety, and environment, proximity to hospitals, main road, office, transportation cost and utility cost, which are quantitative and qualitative in nature. Numerical data which uses numbers is considered as quantitative data and can be measured with 100% certainty. [4] Examples of quantitative data utility cost, transportation cost are the examples of quantitative data since they can be measured using number and with 100% certainty. On the contrary, qualitative data is descriptive in nature, which defines some concepts or imprecise characteristics or quality of things [5]. Hence, this data can't describe a thing with certainty since it lacks the precision and inherits ambiguity, ignorance, vagueness. Consequently, it can be argued that qualitative data involves uncertainty since it is difficult to measure concepts or characteristics or quality of a thing with 100% certainty. Examples of qualitative data associated with in choosing a job are quality of location, safety and environment. "Quality of Location" is an example of equivocal term since it is an example of linguistic term. Hence, it is difficult to extract its correct semantics (meaning). However, this can be evaluated using some evaluation grade such as excellent, good, average and bad. Therefore, it can be seen that qualitative criterions which have been considered in selecting a job involves lot of uncertainties and they should be treated with appropriate methodology. There exists a number of techniques to handle multi-criterion problems such as AHP (Analytical hierarchy process), ANP (Analytical network process) and IPV (inner product vector) approach [8][9]. These approaches use a pair wise comparison matrix in order to identify the importance between two attributes or data. For example, whether the quality of location is more important than environment [16][17]. By applying pair wise comparison method we are able to calculate the weight of these two attributes, for example they can be 0.59 for location and 0.41 for safety. It can be seen that both are qualitative data. However, the calculation of such weight of the attributes is unable to address the problem of incompleteness or vagueness. If a belief structure is used taking account of evaluation grade of the attribute this incompleteness may be addressed and hence the uncertainty. Moreover, when we add another attribute, for example environment with location and safety it can be seen that the ranking of the attributes in terms of their importance will be changed. These types of problems associated with AHP [8] and ANP causes serious problems in decision making. The issues as mentioned can be addressed by using Evidential Reasoning Approach (ER), which is a multi-criteria decision analysis (MCDA) method[13][14]. ER deals with problems, consisting of both quantitative and qualitative criteria under various uncertainties such as incomplete information, vagueness, ambiguity [7].The ER approach, developed based on decision theory in particular utility theory [1][21], artificial intelligence in particular the theory of evidence [18][19]. It uses a belief structure to model a judgment with uncertainty. For example, in AHP

Author α σ : Lecturer, Dept. of Computer Science & Engineering, University of Information Technology and Sciences.
E-mails : tanjim_cse@yahoo.com, namirurad_cse@yahoo.com
Author ρ : Professor, Dept. of Computer Science & Engineering, University of Chittagong. E-mail : hossain_ms@hotmail.com

approach the importance of the attribute location and safety has been calculated as 0.51 and 0.49 respectively. However, such calculation of importance of the attributes contains uncertainty. The reason for this is that qualitative attribute such as location or safety needs to be evaluated using some linguistic evaluation grades such as excellent, average, good and bad etc. This requires human judgment for evaluating the attributes based on the mentioned evaluation grades. In this way, the issue of uncertainty can be addressed and more accurate and robust decision can be made.

II. Evidential Reasoning Approach

The evidential reasoning algorithm is considered as the kernel of the ER approach. This algorithm has been developed based on an evaluation analysis model [22][23] and the evidence combination rule of the Dempster-Shafer (D-S) theory [15][18][19], which is well-suited for handling incomplete uncertainty [22]. The ER approach uses a belief structure to model an assessment as a distribution. It differs with other Multi Criteria Decision Making (MCDM) modeling model a judgment with uncertainty. For example, in AHP methods in that it employs evidence-based reasoning process to derive a conclusion [13][14][20]. The main strength of this approach is that it can handle uncertainties associated with quantitative and qualitative data, related to MCDM problems [13][14] [20].

The ER approach has addressed such issue by proposing a belief structure which assigns degree of belief in the various evaluation grades of the attributes, which is not the case in AHP in other multi-criterions decision techniques.

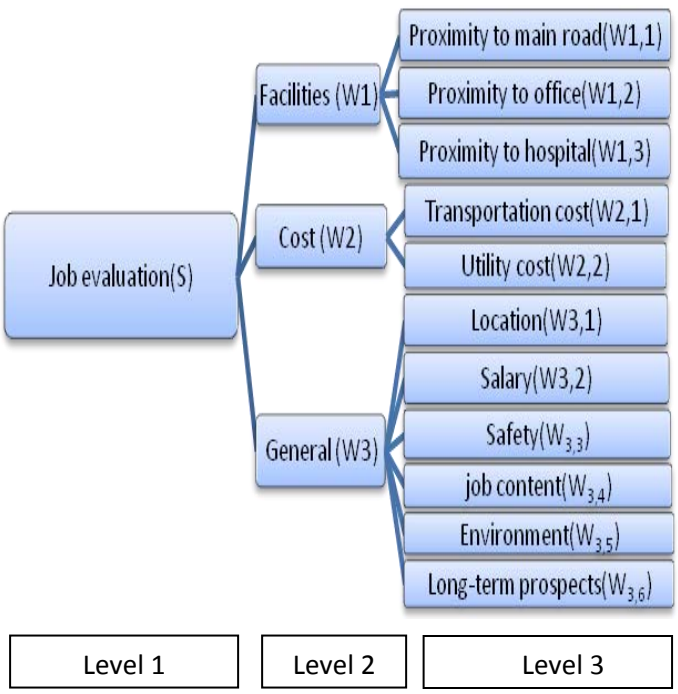


Figure 1 : Evaluation Hierarchy for Operation

In section 2 will briefly represent ER algorithm. Section 3 will demonstrate the application of ER in job evaluation problem. Section 4 will represent the results and achievement. Finally section 5 will conclude the research.

The ER approach consists of five phases[27] including 1) Information acquisition and representation or assessment, 2) weight normalization, 3) basic probability assignment 4) attribute aggregation, 5) Combined degree of belief calculation, 6) utility function 7) ranking.

a) Assessment

One of the critical tasks of developing a decision support system is to acquire information and to represent them in appropriate format so that it will feed into a model. Since ER approach employs belief structure to acquire knowledge, appropriate information should be selected to feed the ER algorithm, which is used to process the information.

Let 'Job evaluation' (S) be an attribute at level 1 as shown in Fig. 1, which is to be assessed for an alternative (A) (i.e. a job at a certain location) and this assessment can be denoted by A(S). This is to be evaluated based on a set of w_i sub-attributes (such as facilities, cost, general) at level 2, denoted by:

$$S = \{w_1, w_2, w_3, \ldots\ldots\ldots w_i, \ldots\ldots w_n\}.$$

Job evaluation (S) can be assessed by using a set of evaluation grades consisting of

Excellent (H_1), Good (H_2), Average (H_3), Bad (H_4)

and this set can be written as $H = \{H_1, H_2, \ldots\ldots\ldots H_n, n = 1,2,\ldots\ldots\ldots, N\}$. These evaluation grades are mutually exclusive and collectively exhaustive and hence, they form a frame of discernment in D-S terminology.

A degree of belief is associated with each evaluation grade, which is denoted by

$$\{(H_n, \beta_n), n = 1,\ldots\ldots, N\}$$

Hence,

$$A(S) = \{(H_n, \beta_n), n = 1,\ldots\ldots, N\}$$

denotes that the top attribute S is assessed to grade H_n with the degree of belief β_n. In this assessment, it is required that $\beta_n \geq 0$ and $\sum_{n=1}^{N} \beta_n \leq 1$. If $\sum_{n=1}^{N} \beta_n = 1$ the assessment is said to be complete and if it is less than one then the assessment is considered as

incomplete. If $\sum_{n=1}^{N}\beta_n = 0$ then the assessment stands for complete ignorance. In the same way, sub-attribute w_i is assessed to grade H_n with the degree of belief $\beta_{n,i}$ and this assessment can be represented as

$$A(w_i) = \{(H_n, \beta_{n,i}), n = 1,.....,N \quad and \quad i = 1,.....,n\}$$

Such that $\beta_{n,i} \geq 0$ and $\sum_{n=1}^{N}\beta_n \leq 1$.

The incompleteness as mentioned occurs due to ignorance, meaning that belief degree has not been assigned to any specific evaluation grade and this can be represented using the equation as given below.

$$\beta_H = 1 - \sum_{n=1}^{N}\beta_n \qquad (1)$$

Where β_H is the belief degree unassigned to any specific grade. If the value of β_H is zero then it can argued that there is an absence of ignorance or incompleteness. If the value of β_H is greater than zero then it can be inferred that there exists ignorance or incompleteness in the assessment. The ER algorithm, as will be discussed, has the procedures to handle such kind of ignorance. It is also necessary to distribute the degree of belief between evaluation grades for certain quantitative input data. For example, sub-attribute 'proximity to hospital', which is at the level 3 of the Fig. 1, consists of four evaluation grades namely Excellent, Good, Average and Bad. When the hospital is located within 1km of the job place, it is considered as excellent, when it is located within 1.5km of the place it is considered as good, when it is located within 2 km of the place it is considered as average and when it is located within 3 km of the place it is considered as bad. However, when a hospital is located 1.3 km of the place, it can be both excellent and average. However, it is important for us to know, with what degree of belief it is excellent and with what degree of belief it is average. This phenomenon can be calculated with the following formula.

$$\beta_{n,i} = \frac{h_{n+1} - h}{h_{n+1}, i - h_{n,i}}, \beta_{n+1,i} = 1 - \beta_{n,i}$$
$$if \qquad h_{n,i} \leq h \leq h_{n+1,i} \qquad (2)$$

Here, the degree of belief $\beta_{n,i}$ is associated with the evaluation grade 'average' while $\beta_{n+1,i}$ is associated with the upper level evaluation grade i.e. excellent. The value of h_{n+1} is the value related to excellent, which is considered as 1km i.e. the location of the hospital. The value of h_{n+1} is related to average, which is 1.5 km. Hence, applying equation (2) the distribution of the degree of belief with respect to 1.3 Km of the location of the hospital from the job place can be assessed by using equation (2) and the result is given below:

{(Excellent, 0.4), (Good, 0.6), (Average, 0), (Bad,0)},

b) Weight Normalization

The identification of the importance of the attributes is very important, since each attribute does not play the same role in decision making process. For example, the sub-attribute of the "Facilities" attribute at level 2 consists of three attributes namely, proximity to main road, hospitals and office. It is important for us to know among three attributes which is the most important in evaluating their parent attribute "Facilities". This can be carried out by employing different weight normalization techniques such as Eigenvector, AHP, Pair wise comparison [8][9][16][17]. In this research Pair wise comparison method has been considered for the normalization of the weights of the attribute by considering the following equations

$$\omega_i = \frac{y_i}{\sum_{i=1}^{j} y_i} \quad ;i= 1.....j \qquad (3)$$

$$\sum_{i=1}^{L}\omega_i = 1 \qquad (4)$$

Equation (3) is used to calculate the importance of an attribute (w_i) .This has been calculated by dividing the importance of an attribute (y_i) (this important of the attribute has been determined from survey data) by the summation $\sum_{i=1}^{j} y_i$ of importance of all the attributes. Equation (4) has been used to check whether the summation of the importance of all the attributes is within one i.e whether they are normalized.

c) Basic Probability Assignment

The degrees of belief as assigned to the evaluation grades of the attributes need to be transformed into basic probability masses. Basic probability mass measures the belief exactly assigned to the n-th evaluation grade of an attribute. It also represents how strongly the evidence supports n-th evaluation grade (H_n) of the attribute. The transformation can be achieved by combining relative

weight (w_i) of the attribute with the degree of belief $(\beta_{n,i})$ associated with n-th evaluation grade of the attribute, which is shown by the following equation.

$$m_{n,i} = m_i(H_n) = w_i \beta_{n,i}(a_l),,,,$$

$$n = 1,.....,N;\ ,,,,,,,,\ i = 1,.....,L, \quad (5)$$

However, in case of hierarchical model, the basic probability mass represents the degree to which the i-th basic attribute supports the hypothesis that the top attribute y is assessed to n-th evaluation grade.

The remaining probability mass unassigned to any individual grade after the ith attribute has been assessed can be given using the following equation.

$$m_{H,i} = m_i(H) = 1 - \sum_{n=1}^{N} m_{n,i} = 1 - w\sum_{n=1}^{N} \beta_{n,i}(a_l),$$

$$i = 1,.....,L, \quad (6)$$

d) Kernel of ER Approach

The purpose of ER algorithm is to obtain the combined degree of belief at the top level attribute of a hierarchy based on its bottom level attributes, also known as basic attributes. This is achieved through an effective process of synthesizing/aggregating of the information. A recursive ER algorithm is used to aggregate basic attributes to obtain the combined degree of belief of the top level attribute of a hierarchy, which can be represented as $A(S) = \{(H_n, \beta_n), n = 1,.....,N\}$. In this recursive ER algorithm, all the basic attributes are aggregated recursively in the following manner as shown in Fig. 2.

In this Fig.2 "Facilities" is considered as the top level attribute, which consists of three sub-attributes. The top level attribute "Facilities" can be denoted by w (i) such that i= 1, 2, 3,..n. This means at this level there could be other attributes. For example, in our case, this level consists of three attributes and the level is considered as second level as shown in Fig. 1. It is interesting to note that top level of Fig.1 contains only one attribute and that can be denoted by So (Job evaluation) and has three sub-attributes at second level. For the top level attribute (S) the combined degree of belief needs to be calculated based on the second level attributes.

From Fig.2 it can be observed that w(1), [considering the value of i as 1] consists of three sub-attributes and hence

$$w_1 = \{w_{11}, w_{12}, w_{13},.....w_{18}\}$$

Or

$$w(i) = \{w_{i,j}, w_{i,j+1}, w_{i,J+2},.....w_{i,J+n}\}$$

such that i=1…..n and j = 1…..L. Taking account of the basic probability assignment and remaining unassigned probability mass of three sub-attributes mass of w_1 matrix (1) has been developed as shown below. These bpa (such as m_{11}, m_{21},,etc and reaming unassigned bpa such M_{H1}) have been calculated by using equations 5 and 6.

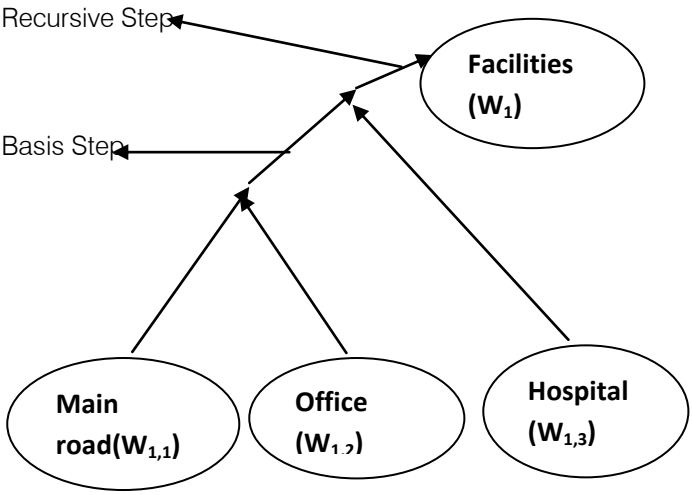


Figure 2 : Recursive Manner of Assessment [27]

$$M = \begin{bmatrix} m_{11} & m_{21} & m_{31} & m_{41} & m_{H1} \\ m_{12} & m_{22} & m_{32} & m_{42} & m_{H2} \\ m_{13} & m_{23} & m_{33} & m_{43} & m_{H3} \\ m_{14} & m_{24} & m_{34} & m_{44} & m_{H4} \\ .... & .... & .... & .... & .... \\ .... & .... & .... & .... & .... \\ .... & .... & .... & .... & .... \\ m_{n,I} & m_{n+1,I} & m_{n+2,I} & m_{n+3,I} & m_{H,I} \end{bmatrix} \quad (1)$$

From matrix (1), it can be seen that each sub-attribute is associated with five basic probability assignment(bpa), where four first four bpa $(m_{11}, m_{21}, m_{31}, m_{41})$ are associated with four evaluation grades (H_1, H_2, H_3, H_4) and final bpa i.e. $m_{H,i}$ is showing the remaining probability mass unassigned to any individual grades after the assessments on sub-attribute have been considered. Each row in this matrix represents bpa related to one basic attribute or sub-attribute.

Now it is necessary to aggregate the bpa of different sub-attributes. The aggregation is carried out in a recursive way. For example, the bpa of first sub-attribute attribute (which is shown in the first row of the matrix 1) is aggregated with the bpa of second sub-attribute. The result of this aggregation is illustrated in the first row of the matrix (2) and this can be considered as the base case of this recursive procedure since this will be used in the latter aggregation of the sub-attributes. This aggregation can be achieved by using the following equation, which will yield combined bpa (such as $m_{1I(2)}, \ldots\ldots\ldots\ldots m_{4I(2)}$) as shown in the first row of the second matrix.

$$m_{1I(2)} = K_{I(2)}(m_{11}m_{12} + m_{H1}m_{12} + m_{H2}m_{11}) \quad (7)$$

$$M = \begin{bmatrix} m_{1I(2)} & m_{2I(2)} & m_{3I(2)} & m_{4I(2)} & m_{HI(2)} \\ m_{13} & m_{23} & m_{33} & m_{43} & m_{H3} \\ m_{14} & m_{24} & m_{34} & m_{44} & m_{H4} \\ m_{15} & m_{25} & m_{35} & m_{45} & m_{H5} \\ .... & .... & .... & .... & .... \\ .... & .... & .... & .... & .... \\ m_{n,I} & m_{n+1,I} & m_{n+2,I} & m_{n+3,I} & m_{H,I} \end{bmatrix} \quad (2)$$

Similarly $m_{2I(2)}, m_{3I(2)}, m_{4I(2)}$ can be calculated.

Where $K_{I(2)}$ is a normalization factor used to resolve the conflict and this can be calculated using the equation (8).

$$K_{I(i+1)} = \left[1 - \sum_{n=1}^{N} \sum_{\substack{t=1 \\ t \neq n}}^{N} m_{n,I(i)} m_{t,i+1} \right]^{-1}, i = 1, \ldots\ldots, L-1 \ldots\ldots \quad (8)$$

The aggregation of the third attribute is carried out with the resultant of the aggregation of the bpa of the first two attributes. In this way, the aggregation of the other attributes is carried out and finally, the combined aggregations of all the attributes are obtained. This phenomenon has been depicted in Figure 2, where the combined aggregation is obtained, which will be used to obtain the combined degree of belief for the second level attribute "facilities". Equation (9) represents the more generalized version of equation (7)

$\{H_n\}$:

$$m_{n,I(i+1)} = K_{I(i+1)} \left[m_{n,I(i)} m_{n,i+1} + m_{n,I(i)} m_{H,i+1} + m_{H,I(i)} m_{n,i+1} \right] \quad (9)$$

$$m_{H,I(i)} = \overline{m_{H,I(i)}} + \tilde{m}_{H,I(i)}, \quad n = 1, \ldots\ldots\ldots\ldots, N, \quad (10)$$

$\{H\}$:

$$\tilde{m}_{H,I(i+1)} = K_{I(i+1)} \left[\tilde{m}_{H,I(i)} \tilde{m}_{H,i+1} + \tilde{m}_{H,I(i)} \overline{m}_{H,i+1} + \overline{m}_{H,I(i)} \tilde{m}_{H,i+1} \right] \quad (11)$$

$$\{H\}: \overline{m}_{H,I(i+1)} = K_{I(i+1)} \left[\overline{m}_{H,I(i)} \overline{m}_{H,i+1} \right], \quad (12)$$

Equation 13 is used to calculate the combined degree of belief by using final combined basic probability assignment, say in this case "facilities".

$$\{H_n\}: \beta_n = \frac{m_{n,I(L)}}{1-\overline{m}_{H,I(L)}}, n = 1,.....,.N, \quad (13)$$

$$\{H\}: \beta_H = \frac{\widetilde{m}_{H,I(L)}}{1-\overline{m}_{H,I(L)}}, \text{ Where}$$

$$m_{n,I(1)} = m_{n,1}(n = 1,.....N) \quad (14)$$

β_n and β_H represent the belief degrees of the aggregated assessment, to which the general factor (such as "facilities") is assessed to the grade H_n and H, respectively. The combined assessment can be denoted by $S(y(a_l)) = \{(H_n, \beta_n(a_l)), n = 1,.....,N\}$,. It has been proved that $\sum_{n=1}^{N} \beta_n + \beta_H = 1$.

The recursive ER algorithm combines various piece of evidence on a one-by-one basis.

e) The Utility Function (Ranking Job)

Utility function is used to determine the ranking of the different alternatives. In this research different job sector have been considered as the alternatives. Therefore, the determination of ranking of the alternatives will help to take a decision to decide the suitable job. There are three different types of utility functions considered in the ER approach namely: minimum utility, maximum utility and average utility. In this function, a number is assigned to an evaluation or assessment grade. The number is assigned by taking account of the preference of the decision maker to a certain evaluation grade. Suppose the utility of an evaluation grade H_n is $u(H_n)$, then the expected utility of the aggregated assessment $S(y(a_l))$ is defined as follows:

$$u(S(y(a_l))) = \sum_{n=1}^{N} u(H_n)\beta_n(a_l)$$

The belief degree $\beta_n(a_l)$ represents the lower bound of the likelihood that a_l is assessed to H_n, whilst the corresponding upper bound of the likelihood is given by $(\beta_n(a_l) + \beta_H(a_l))$ The maximum, minimum and average utilities of a_l can be calculated by:

$$u_{\max}(a_l) = \sum_{n=1}^{N-1} \beta_n(a_l)u(H_n) + (\beta_N(a_l) + \beta_H(a_l))u(H_N)),$$

$$u_{\min}(a_l) = (\beta_1(a_l) + \beta_H(a_l))u(H_1) + \sum_{n=2}^{N} \beta_n(a_l)u(H_n),$$

$$u_{average}(a_l) = \frac{u_{\max}(a_l) + u_{\min}(a_l)}{2}.$$

It is important that if $u(H_1) = 0$, then $u(S(y(a_l))) = u_{\min}(a_l)$ if all the original assessments $S(e_i(a_l))$ in the belief matrix are complete, then

$$\beta_H(a_l) = 0$$

and

$$u(S(y(a_l))) = u_{\min}(a_l) = u_{\min}(a_l) = u_{average}(a_l).$$

It has to be made clear that the above utilities are only used for characterizing a distributed assessment but not for the aggregation of factors.

III. Results and Discussion

In the previous section, we have discussed about the ER method and how to implement it. Therefore, in this section we will look at the results from using this method on the different types of job. The ER approach for job evaluation consists mainly of four key parts, which are the identification of factors, the ER distributed modeling framework for the identified factors, the recursive ER algorithms for aggregating multiple identified factors, and the utility function [3] based ER ranking method which is designed to compare and rank alternatives/options systematically. Each part will be described in detail in above section. Job evaluation, can be described in two broad categories: the Objective attribute, and subjective attribute as shown in Fig. 1 and each attribute weights are

W1=0.20,w2=0.20,w3=0.60,w11=0.33,w12=0.33,w13=0.33,w21=0.70,w22=0.30,w31=0.05,w32=0.15,w33=0.05,w34=0.2,w35=0.05,w36=0.5

Figure 3 shows the assessment grades defined by the decision maker for Level 3(Fig. 1). Shows the assessment distribution which must be done first by employing the transformation equation. Any measurements of quality can be translated to the same set of grades as the top attribute which make it easy for further analysis.

The assessments given by the Decision Maker (DM) in Figure 1 are fed into Decision support system (DSS) [25][26] and the aggregated results are yielded at the main criteria level (Fig. 1). The assessment grades for each main criterion are abbreviated in Figure 3. The numbers in brackets show the degrees of belief of the DM that are aggregated from the assessments of the sub-criteria. One can rank the job for each criterion in

order of preference by comparing the distributed assessments shown in Figure 4.

The results in Figure 3. are also useful in that they indicate the weak and strong points of each alternative regarding the decision criteria applied. The DSS [25][26] provides a graphical display of the results presented in Figure 6. The assessments in Figure 3 need to be propagated to the top level. The numbers under each grade indicate the aggregated assessments (or degrees of belief) of the DM. For instance, the results for job **Acme Manufacturing (A)** can be interpreted as follows: job **Acme Manufacturing (A)** is assessed to be 15% bad, 10% average, 23% good, and 52% excellent. The total degree of belief does not add up to one (or 100%) as a result of incomplete and/or missing assessments. The results in Figure 5. are supported by decision support system(DSS). The job could be ranked in order of preference by comparing them with each other as in Fig.3. However, a comparison may not be possible when job have very similar degrees of belief assigned to each grade. One way to solve this problem is to quantify the grades. There are several ways of quantifying grades. One of them is to assign a utility for each grade and then obtain an expected utility for each job. Then, jobs are ranked based on their expected utility [3]. In this research, the former approach is used. A number of hypothetical lottery type questions were presented to the DM in order to establish preference among grades. The following utilities are assigned to each grade:
(Bad, 0.4), (Average, 0.7), (Good, 0.85) and (Excellent, 1).

Attributes	Assessment Grades			
Location	Excellent	Good	Average	Bad
Job Content	Excellent	Good	Average	Bad
Safety	Excellent	Good	Average	Bad
Environment	Excellent	Good	Average	Bad
Cost	Excellent	Good	Average	Bad
General	Excellent	Good	Average	Bad
Facilities	Excellent	Good	Average	Bad
Long-term Prospects	Excellent	Good	Average	Bad
Job Evaluation	Excellent	Good	Average	Bad
Transportation Cost	Quantitative			
Proximity to Hospital	Quantitative			
Proximity to Office	Quantitative			
Utility Cost	Quantitative			
Salary	Quantitative			
Proximity to Main Road	Quantitative			

Figure 3 : Assessment Grades Defined by the Decision Maker for the 3rd Level

Attributes	Acme Manufacturing (A)	Bankers Bank (B)	Creative Consulting (C)	Dynamic Decision Making (D)
Location	B(0.2)A(0.8)	G(0.4)E(0.6)	G(0.4)E(0.6)	E(1.0)
Job Content	G(0.4)E(0.6)	B(0.2)A(0.8)	B(0.2)A(0.8)	G(0.4)E(0.6)
Safety	B(0.2)E(0.8)	A(1.0)	G(1.0)	A(1.0)
Environment	E(1.0)	G(1.0)	G(0.4)E(0.6)	G(1.0)
Long-term Prospects	G(1.0)	B(0.2)E(0.8)	E(1.0)	B(0.2)A(0.8)
Proximity to Hospitals(Km)	2.3	2.6	2.4	2.0
Proximity to Office(Km)	2.0	1.6	1.0	2.0
Proximity to Main Road(Km)	1.4	1.0	2.1	2.5
Salary(Thousand)	1.0	1.6	1.6	2.0
Transportation Cost(Thousand)	2.3	1.0	1.1	1.4
Utility Cost(Thousand)	2.0	2.3	2.0	2.0

Figure 4 : Assessment Scores of Job Sector Based on Sub Criteria
(E-Excellent, G-Good, A-Average, B-Bad)

Attributes	Acme Manufacturing (A)	Second level assessment	Third level assessment
Location	B(0.2)A(0.8)	General{(E,0.15),(G,0.78),(A,0.05),(B,0.02)}	Job Evaluation {(E,0.14),(G,0.8),(A,0.04),(B,0.01)}
Job Content	G(0.4)E(0.6)		
Safety	B(0.2)E(0.8)		
Environment	E(1.0)		
Salary(Thousand)	40000		
Long-term Prospects	G(1.0)		
Proximity to Hospitals(Km)	2.3	Facilities{(E,0.2),(G,0.6),(A,0.15),(B,0.05)	
Proximity to Office(Km)	2.0		
Proximity to Main Road(Km)	1.4		
Transportation Cost (Thousand)	2.3	Cost{(E,0.22),(G,0.68),(A,0.05),(B,0.05)	
Utility Cost (Thousand)	2.0		

Figure 5 : Overall Assessment For Acme Manufacturing (A)

Alternative	Excellent	Good	Average	Bad	Total DoB	Unassigned DoB
Acme Manufacturing (A)	0.14	0.8	0.04	0.01	0.99	0.01
Bankers Bank (B)	0.16	0.23	0.48	0.13	1.00	0.00
Creative Consulting (C)	0.17	0.70	0.10	0.03	1.00	0.00
Dynamic Decision Making (D)	0.18	0.40	0.40	0.02	1.00	0.00

Figure 6 : The Overall Assessment (Alternatives)
(Dob-Degree of Belief)

Alternative	Minimum Utility	Maximum Utility	Average Utility	Rank
Acme Manufacturing (A)	0.850	0.855	0.853	1
Bankers Bank (B)	0.743	0.743	0.743	4
Creative Consulting (C)	0.847	0.847	0.847	2
Dynamic Decision Making (D)	0.808	0.808	0.808	3

Figure 7 : The Expected Utilities of Alternative Job

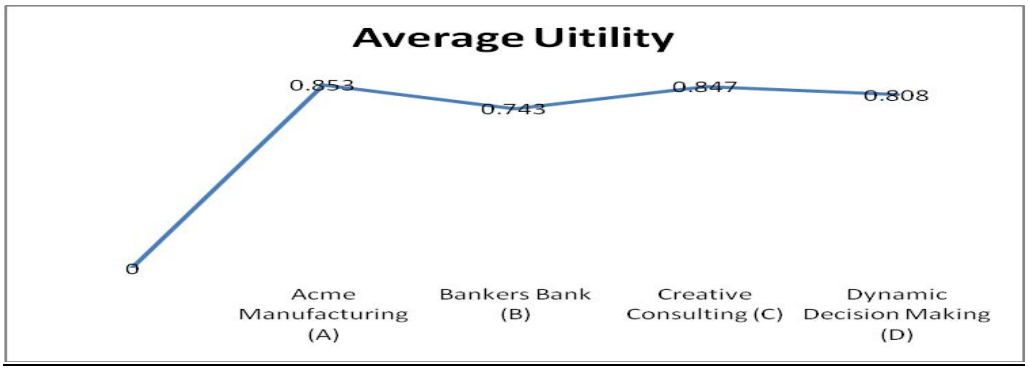


Figure 8 : Expected Average Utility of Alternative

The total Degree of belief for each job in Figure 6 does not add up to one, because some of the assessments were incomplete and missing. For example, the total Degree of belief assigned to job alternative is 97%. That is, there is a 3% unassigned degree of belief. The DSS uses the concept of utility interval to characterize the unassigned Degree of belief (or ignorance) which can actually fall into any grade. The ER algorithm generates a utility interval enclosed by two extreme cases where the unassigned Degree of belief goes either to the least preferred grade (minimum utility) or goes to the most preferred grade (maximum utility). The minimum and maximum possible utilities of each alternative generated by the DSS [25][26] (based on the given utility values for each grade above) are shown in Figure 7. or Fig. 5. For example, the results for job **Acme Manufacturing (A)** from FIG.6. are as follows: job **Acme Manufacturing (A)** is minimum utility.

{[(Degree of belief assigned under grade bad + unassigned Degree of belief) utility of grade bad] + (Degree of belief assigned under grade average * utility of grade average) + (Degree of belief assigned under

grade good *utility of grade good) + (Degree of belief assigned under grade excellent * utility of grade excellent)}.

Hence, job **Acme Manufacturing (A)** minimum utility. {[(0.1+ 0.01) 0.4] + (0.1 * 0.7) + (0.29 *0.85)+ (0.5 *1.0) =0.860 job **Acme Manufacturing (A)** maximum utility . {(Degree of belief assigned under grade bad* utility of grade bad) + (Degree of belief assigned under grade average* utility of grade average) + (Degree of belief assigned under grade good* utility of grade good) + [(Degree of belief assigned under grade excellent + unassigned Degree of belief) *utility of grade excellent]}.

Hence, job **Acme Manufacturing (A)** maximum utility {(0.1* 0.4) + (0.1 * 0.7) + (0.29 *0.85) + [(0.5 +0.01)*1.0]} =0.866 job **Acme Manufacturing (A)** average utility. (Maximum utility + minimum utility)/2 =0.863.

The job may be ranked based on the average utility but this may be misleading. In order to say that one job theoretically dominates another, the preferred job minimum utility must be equal or greater than the dominated job maximum utility. The ranking of job is as follows:

Acme Manufacturing (A) > Creative Consulting (C)> Dynamic Decision Making (D) > Bankers Bank (B)

IV. Conclusion

This paper established the scheme of the application of this evidential reasoning to solve a multiple criteria job offers evaluation with uncertain, incomplete, imprecise, and/or missing information. From the results shown above, it is reasonable to say that the evidential reasoning method is a mathematically sound approach towards measuring the job quality as it employs a belief structure to represent an assessment as a distribution. This approach is quite different from the other Multi Criteria Decision Making model such as the Saaty 's AHP method which uses a pair wise comparison matrix[8][9][13[14]. Hence, the ER method can handle a new attribute without recalculating the previous assessment because the attribute can be arranged or numbered arbitrarily which means that the final results do not depend on the order in which the basic attributes are aggregated. Furthermore, any number of new job alternative can be added to the assessment as it does not cause a 'rank reversal' as in the Saaty's AHP method[8][9][13[14]. Finally, in a complex assessment as in the job quality appraisal which involved objective and subjective assessments of many basic attributes as shown in Figure 1, it is convenient to have an approach which can tackle the uncertainties or incompleteness in the data gathered. Therefore, the ER is seen as reasonable method for 'quality job' evaluation.

References Références Referencias

1. M Sonmez, G. Graham and **J. B. Yang** and G D Holt, "Applying evidential reasoning to pre-qualifying construction contractors", *Journal of Management in Engineering*, Vol.18, No.3, pp.111-119, 2002.
2. **J. B. Yang**, "Rule and utility based evidential reasoning approach for multiple attribute decision analysis under uncertainty", *European Journal of Operational Research*, Vol. 131, No.1, pp. 31-61, 2001.
3. Y. M. Wang, **J. B. Yang** and D. L. Xu, "Environmental Impact Assessment Using the Evidential Reasoning Approach", *European Journal of Operational Research*, Vol.174, No.3, pp.1885-1913, 2006.
4. Lisa M. (2008). *The Sage encyclopedia of qualitative research methods*. Los Angeles, Calif.: Sage Publications. ISBN 1-4129-4163-6.
5. http://www.pearson.ch/1449/9780273722595/An-Introduction-to-Geographical.aspx
6. Dodge Y. (2003) *The Oxford Dictionary of Statistical Terms,* OUP. ISBN 0-19-920613-9
7. D. L. Xu and **J. B. Yang**, "Introduction to multi-criteria decision making and the evidential reasoning approach", *Working Paper Series*, Paper No.: 0106, ISBN: 1 86115 111 X (http://www.umist.ac.uk/management), Manchester School of Management, UMIST, pp. 1-21, 2001.
8. Saaty, T.L. (1980). *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*. New York: McGraw-Hill.
9. Grandzol, John R. (August 2005). "Improving the Faculty Selection Process in Higher Education: A Case for the Analytic Hierarchy Process". *IR Applications 6.* Retrieved 2007-08-21.
10. Atthirawong, Walailak; Bart McCarthy (September, 2002). "An Application of the Analytical Hierarchy Process to International Location Decision-Making". In Gregory, Mike. *Proceedings of the 7th Annual Cambridge International Manufacturing Symposium: Restructuring Global Manufacturing*. Cambridge, England: University of Cambridge. pp. 1–18.
11. Larson, Charles D.; Ernest H. Forman (January, 2007). "Application of the Analytic Hierarchy Process to Select Project Scope for Videologging and Pavement Condition Data Collection". *86th Annual Meeting Compendium of Papers CD-ROM*. Transportation Research Board of the National Academies.
12. Drake, P.R. (1998). "Using the Analytic Hierarchy Process in Engineering Education". *International Journal of Engineering Education* 14 (3): 191–196. Retrieved 2007-08-20.
13. Köksalan, M., Wallenius, J., and Zionts, S. (2011). Multiple Criteria Decision Making: From Early History to the 21st Century. Singapore: World Scientific.

14. Köksalan, M.M. and Sagala, P.N.S., M. M.; Sagala, P. N. S. (1995). "Interactive Approaches for Discrete Alternative Multiple Criteria Decision Making with Monotone Utility Functions". *Management Science* 41 (7): 1158–1171.
15. A. Taroun and J. B. Yang. "Dempster-Shafer theory of evidence: potential usage for decision making and risk analysis in construction project management." *Journal of the Built and Human Environment Review* 4, no. 1(2011): 155-166.
16. Saaty, Thomas L. (2008). *Decision Making for Leaders: The Analytic Hierarchy Process for Decisions in a Complex World*. Pittsburgh, Pennsylvania: RWS Publications. ISBN 0-9620317-8-X. http://www.amazon.com/dp/096203178X.
17. Dey, Prasanta Kumar (November 2003). "Analytic Hierarchy Process Analyzes Risk of Operating Cross-Country Petroleum Pipelines in India". *Natural Hazards Review* 4 (4): 213–221. DOI:10.1061/(ASCE)1527-6988(2003)4:4(213)
18. L. Zadeh, A simple view of the Dempster-Shafer Theory of Evidence and its implication for the rule of combination, The Al Magazine, Vol. 7, No. 2, pp. 85-90, Summer 1986.
19. Kari Sentz and Scott Ferson (2002); *Combination of Evidence in Dempster–Shafer Theory,* Sandia National Laboratories SAND 2002-0835.
20. Saaty, Thomas L. (1996). Decision Making with Dependence and Feedback: The Analytic Network Process. Pittsburgh, Pennsylvania: RWS Publications. ISBN 0-9620317-9-8.
21. Bragge, J.; Korhonen, P., Wallenius, H. and Wallenius, J. (2010). "Bibliometric Analysis of Multiple Criteria Decision Making/Multiattribute Utility Theory". *IXX International MCDM Conference Proceedings, (Eds.) M. Ehrgott, B. Naujoks, T. Stewart, and J. Wallenius,.* Springer, Berlin 634: 259–268.
22. P. Sen and J. B. Yang, "Design decision making based upon multiple attribute evaluation and minimal preference information", *Mathematical and Computer Modeling*, Vol.20, No.3, pp. 107-124, 1994.
23. B. Yang and P. Sen, "Multiple attribute design evaluation of large engineering products using the evidential reasoning approach", *Journal of Engineering Design*, Vol.8, No.3, pp. 211-230, 1997
24. Henk G. Sol et al. (1987). *Expert systems and artificial intelligence in decision support systems: proceedings of the Second Mini Euroconference, Lunteren, The Netherlands, 17–20 November 1985*. Springer, 1987. ISBN 90-277-2437-7. p. 1-2.
25. Efraim Turban, Jay E. Aronson, Ting-Peng Liang (2008). Decision Support Systems and Intelligent Systems. p. 574
26. Tanjim Mahmud and Mohammad Shahadat Hossain, "An Evidential Reasoning-based Decision Support System to Support House Hunting", *International Journal of Computer Applications* 57(21):51-58, November 2012.Published by Foundation of Computer Science, New York, USA.

Global Journal of Computer Science and Technology
Neural & Artificial Intelligence
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Bayesian Regularization Based Neural Network Tool for Software Effort Estimation

By Harwinder Kaur & Dalwinder Singh Salaria
Lovely Professional University, India

Abstract - Rapid growth of software industry leads to need of new technologies. Software effort estimation is one of the areas that need more concentration. Exact estimation is always a challenging task. Effort Estimation techniques are broadly classified into algorithmic and non-algorithmic techniques. An algorithmic model provides a mathematical equation for estimation which is based upon the analysis of data gathered from previously developed projects and Non-algorithmic techniques are based on new approaches, such as Soft Computing Techniques. Effective handling of cost is a basic need for any Software Organization. The main tasks for Software development estimation are determining the effort, cost and schedule of developing the project under consideration. Underestimation of project done knowingly just to win contract results into loses and also the poor quality project. So, accurate cost estimation leads to effective control of time and budget during software development. This paper presents the performance analysis of different training algorithms of neural network in effort estimation. For sake of ease, we have developed a tool in MATLAB and at last proved that Bayesian Regularization [20] gives more accurate results than other training algorithms.

Keywords : *effort estimation, levenberg-marquardt (trainlm), back propagation, bayesian regularization (trainbr), gradient descent (traingdx), MATLAB.*

GJCST-D Classification : *I.2.5 , I.2.6*

Strictly as per the compliance and regulations of:

Bayesian Regularization Based Neural Network Tool for Software Effort Estimation

Harwinder Kaur [α] & Dalwinder Singh Salaria [σ]

Abstract - Rapid growth of software industry leads to need of new technologies. Software effort estimation is one of the areas that need more concentration. Exact estimation is always a challenging task. Effort Estimation techniques are broadly classified into algorithmic and non-algorithmic techniques. An algorithmic model provides a mathematical equation for estimation which is based upon the analysis of data gathered from previously developed projects and Non-algorithmic techniques are based on new approaches, such as Soft Computing Techniques. Effective handling of cost is a basic need for any Software Organization. The main tasks for Software development estimation are determining the effort, cost and schedule of developing the project under consideration. Underestimation of project done knowingly just to win contract results into loses and also the poor quality project. So, accurate cost estimation leads to effective control of time and budget during software development. This paper presents the performance analysis of different training algorithms of neural network in effort estimation. For sake of ease, we have developed a tool in MATLAB and at last proved that Bayesian Regularization [20] gives more accurate results than other training algorithms.

Keywords : effort estimation, levenberg-marquardt (trainlm), back propagation, bayesian regularization (trainbr), gradient descent (traingdx), MATLAB.

I. Introduction

Software effort estimate is one of the noticeable & mind catching field. But since it was started, it is challenging factor for software industry and Academia to realize the exact estimation of software development. In today's fast changing world, success in managing projects is a critical factor for the success of the entire organization. Estimation that either overestimated or underestimated both is very critical. In case of Overestimating time and effort (or budget), due to a presumed lack of resources or because the projected completion is too late, can convince management not to approve projects that may otherwise contribute to the organization. On the other hand, underestimation may result in approval of projects that will fail to deliver the expected product within the time and budget available. There are many factors that influence the Software estimation, some of them are: uncertainty, level of detail of preparing the project plan, managerial factors, lack of past data, pressure to lower estimation and estimator experience [1]. In spite of the critical role of accuracy, examples of incorrect estimation abound, especially in IT projects, resulting in enormous waste of time and money. Some techniques which were used in the past are not in use during present time, like SLIM [14], checkpoint [2], Seer [2]. In all the way of work time, many of new advance roads have been suggested for effort estimation like Genetic programming [11], Fuzzy logic [10], Neural Network [15], data mining [9], etc.

One cannot state that one model give better accuracy above all. Each and every give different level of accuracy in different Environment. But in recent days, Neural Network gains main attention due to many flavor of algorithm available for it. The main focus of this paper is to investigate the accuracy of estimation using neural network approach based on three different training algorithms: Levenberg-Marquardt (trainlm) [20], Back propagation [20], Bayesian Regularization (trainbr) [20] and this has been done with the help of tool generated by us in MATLAB.

This paper comprises as follow: section II describes the some former effort estimation models and review of related work to Neural Network, section III includes introduction of Neural Network and training algorithms used for this paper, in section IV problem is stated, section V describes methodology used, section VI includes experimental results and comparisons. In last conclusion and future scope is given.

II. Review of Literature

The period of Effort Estimation was started from the expert judgments, which is based on the experiences of experts. But it is only proceed as pillar when current project & pertinent Past projects are similar. Choices of effort estimation techniques footstep from COCOMO [14] to AI approaches [2]. In 1970, Larry Putnam developed the method *SLIM* [14], based on the Rayleigh function and the influence used to Rayleigh curve was Manpower Buildup Index (MBI and Productivity Factor (PF) [2]. Linear programming was key work to drive effort estimation in SLIM [14] and depend upon the source line of code.

In 1981 developer Barry Boehm developed *COCOMO* as constructive cost model [4]. Which is one

Author α : Student, Dept. of CSE Lovely Professional University Phagwara , Punjab (India)-144411.
E-mail : Final destiny 210@gmail.com
Author σ : Assistant Professor, Dept. of CSE Lovely Professional University Phagwara, Punjab (India)-144411.
E-mail : Ds_salaria@yahoo.com

of an easy going & understandable model, could call the effort & time period of project. Due to some problems and some misses found in COCOMO, later on Barry Bohem developed the advance road of this model i.e. COCOMO 2.0 [7]. As growth of software industry rising tremendously and previous version was not up to need.

After that, Howard Rubin proposed the ESTIMAC model to estimate effort at conception stage [4]. Equations used in this model are not available, because it was a proprietary model. ESTIMAC is high level model but doesn't provide accurate solution [3]. Six critical estimation dimensions identified by Rubin for this model are: effort hour, staff size, cost, hardware resource requirement, risk, portfolio impact [2]. But these methods (COCOMO, SLIM, ESTIMAC) are based on Line of code (LOC). The main problems in Line of Code methods are: lack of universally accepted definition for exactly what line code really is? Other side line of code is language dependence.

So, in 1979 at IBM, developer Allan Albrecht developed measurement method called Function point [3] in order to reduce the issues related with LOC methods. Function point defines the complexity of software system in terms of functions that system delivers to user. It comprise linear combination of five basic software components (input, output, master files, interfaces, inquiries) consider to be low, average, high [3]. In 1990, GC Low and DR. Jeffery also concluded in their paper that Function point method is more consistent then the line of code measure [6]. But on the other side, function point method is unable to deal with Uncertain, imprecise and incomplete data.

Many researcher's use different Neural Network with different datasets in order to generate more accurate result for effort estimation. The main advantage of neural network is its ability to handle non-linear data and confidence in decision making. In 1995, Krishna moorthy Srinivasan and Douglas Fisher applied the machine learning approach for Software Effort Estimation [16]. They applied the Back propagation algorithm on COCOMO dataset, along with configuration of 33 neuron of input layer, 10 neurons for hidden layer and 1 output neuron. Actually they had done three experiments on different datasets. They concluded that Back propagation competitive again traditional approaches but quite sensitive.

In one paper written by Ali Idri, et al. [17] in 2002, in which he uses COCOMO-81 dataset and three layered back-propagation ANN, concluded that accuracy provide by back propagation is acceptable.

In 2005, N Tadayon compares the three models COCOMO II, Neural Network and expert judgments to state the strength of different estimation techniques [13]. In 2006, according to Barcelos Tronto et.al Neural Network approach provides better tune result than the linear regression [18]. In his methodology, he used the Back propagation as training algorithm on COCOMO dataset.

In 2010, Iman Attarzadeh, proposed new model of COCOMO II using neural network, and comcluded that neural network approach gives best accuracy than COCOMO II.

Mrinal Kantri, et al. [19] implemented a back-propagation ANN of 3-3-1 architecture on Lopez Martin dataset consist of 41 projects.

Table I : Summary of Datasets and Neural Network used

Author's Name	Year	Dataset	Project	Training Algorithm	ANN Layers	Conference/ Journal
Krishnamoorthy et.al	1995	COCOMO	63	BPA	33-10-1	IEEE
Ali Idri, et al.	2002	COCOMO	63	BPA	13-13-1	IEEE
N Tadayon	2005	-	-	BPA	-	IEEE
Barcelos Tronto et.al	2006	COCOMO	68	BPA	1-9-4-1	IEEE
Attarzadeh	2010	COCOMO, Artifical	100	BPA	24 input neuron	IEEE
Mrinal Kantri	2011	Lopez martin	41	BPA	3-3-1	IEEE

There are many other techniques such as ordinary least square (OLS) [2], Case based reasoning [12], Date mining [9], Bayesian COCOMO II [2], Genetic Programming [5] etc. also used for the effort estimation but not discussed in this paper.

III. Introduction to Neural Network and Training Algorithms

A Neural Network is massively distributed processor made up of simple processing elements called neuron, which model some functionality like human brain [15]. The use of Neural Network offers the

some useful properties and capabilities: - Nonlinearity, Adaptivity, Evidential Response, Confidence in decision made. A primary advantage of learning systems is that they are nonparametric; predictive models can be tailored to, the data at a particular site [8].

Figure 1 : Model of Neural Network

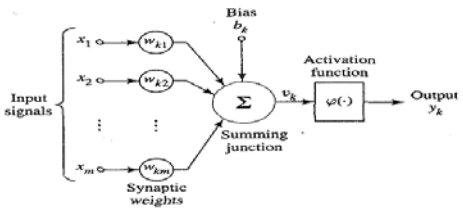


Figure 1 represents the model of neural network. A set of synapses or connecting links is characterized by a weight or strength of its own. A signal x_j at the input of synapse j connected to neuron k is multiplied by the synaptic weight w_{kj}. An adder for summing the input signals. An activation function for limiting the amplitude of output of a neuron.

By adjusting of number of neurons in hidden layer, input layer and choosing of better algorithm for training data and testing phase will give result with high level of accuracy. Training a neural network model refers to choosing one model from the set of permitted models that reduces the cost criterion. Basically, training is of two types Supervised and Unsupervised [15]. In case of Supervised, desired input and output is given to network but in case of unsupervised only desired inputs are given. The network itself makes the decision of output. There are plenty of the training algorithms available in neural network. In this paper three algorithms are used: Leven berg-Marquardt (trainlm) [20], Back propagation algorithm [20], and Bayesian Regularization [20].

Leven berg-Marquardt (trainlm) [20] is a network training function that updates weight and bias values according to Leven berg-Marquardt optimization [20]. This is a simple method for approximating a function. trainlm is highly recommended as a first-choice supervised algorithm. One of the main drawbacks of the Leven berg-Marquardt algorithm is that, for certain problems it needs the large storage of some matrices.

Back propagation [20] learning updates the network weights and biases in the direction in which the performance function decline most quickly, the negative of the gradient [20]. There are too many flavors of Back propagation. For this study, Gradient descent with momentum and adaptive learning rate back propagation (traingdx) is used. The function traingdx combines adaptive learning rate with momentum training. It is a simple method with no specialization needed. But due to low prediction capability, results are not accurate. This has been shown in Experiment section.

One of the problems that occur during above neural network training algorithms is over fitting. Due to this, error in early stage is very small, but, when new data is presented to the network the error is large. The solution to this problem is Bayesian regularization (trainbr) [20]. trainbr updates the weight and bias values according to Levenberg-Marquardt [20] optimization. It minimizes a grouping of squared errors and weights, and generates a network that generalizes well. The process is called Bayesian regularization. It is suitable method for estimation when large number of inputs is used for best output. Till now, Levenberg-Marquardt and Back-propagation algorithm used by many researchers for training phase.

IV. Problem Statement

The main aim of any software development organizations is to finish the project within acceptable or customary schedule and budget. Budget is mainly driven by labor cost and time and together they form a measure called effort. From quality point of view estimating effort is one of the major important factors. Because estimation either it be over estimate or under estimate, produces worst results. In case of over estimation of time and effort project completion is too late due to lack of resources, which refuses the management to approve that favored project. On the other hand, under estimation may result in approval of projects that will fail to deliver the expected product within the time and budget available [1]. So, there is a need of accurate estimation effort technique at early stages of software development. In this research, the main aim is to improve software effort estimation by using different training algorithms of Neural network.

The main reason for using such a learning system for this problem is to keep the estimation process up-to-date by incorporating up-to-date project data. At last Comparison is drawn between training algorithms used in this research to state that Bayesian Regularization gives much accurate estimation. One algorithmic approach, COCOMO is also compared with all three algorithms.

V. Proposed Methodology

Following are the steps used for Effort Estimation:

a) Data Collection

The dataset used in this work is NASA93 (http://promisedata.googlecode.com) a public available data set consisting of a total of 93 projects at the time of this study.

b) Division of Data

Data set is divided into two parts: Training and Testing. For our work we divide the data into 85-15% ratio i.e. 80 rows for training and 13 for testing. These 13 rows are randomly choose by formula (ceil (1+(93-1)*rand(13,1))), available in MATLAB. From this, for

testing row number 15,40,92,74,91,94,5,80,59,64,71,63, 38 are chosen.

c) Cost Drivers

Cost drivers for this work choose from the cost drivers designed for COCOMO II. Table II represents, Cost drivers for COCOMO.

Table II : Cost-drivers of COCOMO model

Attribute	Type	Description
RELY	Product	Required system reliability
CPLX	Product	Complexity of system modules
DOCU	Product	Extent of documentation required
DATA	Product	Size of database used
RUSE	Product	Required percentage of reusable components
TIME	Computer	Execution time constraint
PVOL	Computer	Volatility of development platform
STOR	Computer	Memory constraints
ACAP	Personnel	Capability of project analysts
PCON	Personnel	Personnel continuity
PCAP	Personnel	Programmer capability
PEXP	Personnel	Programmer experience in project domain
AEXP	Personnel	Analyst experience in project domain
LTEX	Personnel	Language and tool experience
TOOL	Project	Use of software tools
SCED	Project	Development schedule compression
SITE	Project	Extent of multisite working and quality of inter-site communications

d) Tool Generation

For the sake of ease, tool is generated with the help of MATLAB. This has been shown in Figure II.

e) Preparation of Neural Network

Depending upon the architecture the Neural Network is of two types: feed-forward and Feed-back. A feed-forward, is the architecture in which the network has no loops. But feed-back neural network is architecture in which loops occurs in the network.

Figure II : Tool developed with the help of MATLAB

For our work, we use feed-forward network with three different training algorithms: LM, BPA, BR. The Neural Network is implemented using 12 neurons for input layer, 12 for hidden layer and 1 for output layer.

f) Performance Criteria

Mean Magnitude Relative Error: MMRE is frequently used to evaluate the performance of any estimation technique. It seems obvious that the purpose of MMRE is to assist us to select the best estimation approach. It measures the percentage of the absolute values of the relative errors, averaged over the N items in the "Test" set and can be written as [18]:

MMRE = {actual effort} - {estimated effort}|}\ {actual effort}

VI. Experimental Results and Comparison

Neural Network trained by three different training algorithms, with same dataset i.e. NASA93. Table III summarizes the result obtained by COCOMO model and three different training algorithms.

Table III : Effort Estimation by using different training algorithms in Neural Network and COCOMO model

Row No.	Expected	COCOMO	LM	BPA	BR
15	48	85.9557	53.7929	1737.61	61.9294
40	114	66.9477	186.747	1702.08	121.206
92	240	85.9557	117.681	1694.9	85.847
74	4178.2	1649.24	1730.38	1843.92	4058.46
91	1772.5	539.26	1400.97	1829.47	2902.12
94	1924.5	393.61	2524.9	1830.02	1201.62
5	25.2	38.2213	260.445	1731.69	83.0016
80	703	904.279	367.178	1836.86	562.929
59	4560	6718.84	1347.35	1945.73	4471.23
64	150	115.445	270.15	1048.19	61.3017
71	72	155.732	85.29	1759.78	106.606
63	160	270.499	294.428	1056.21	61.7749
38	444	463.311	150.555	1706.99	338.233

Figure III : Column chart for effort estimation

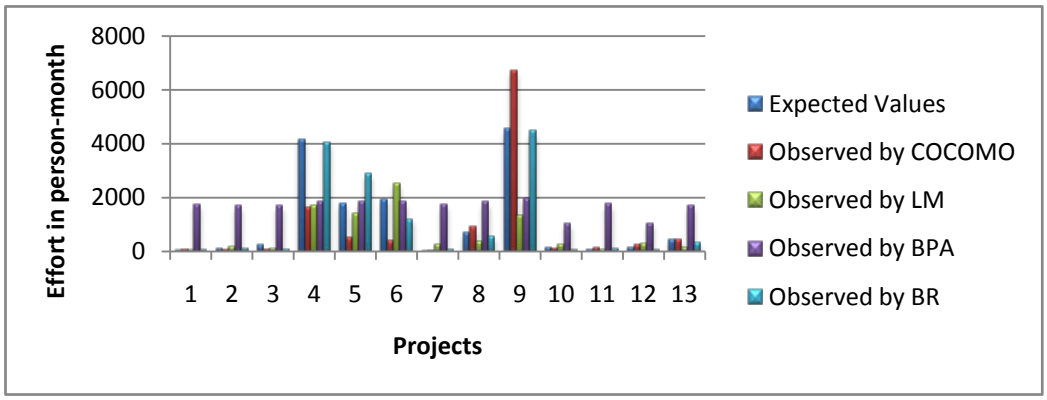


Table IV : Comparison between different training algorithms

Performance Criteria	COCOMO	LM	BPA	BR
MMRE	0.52	1.23	12.18	0.48

In the testing phase the calculated efforts and errors using different training algorithms and COCOMO is shown in table III and table IV respectively. Figure III clearly present Bayesian Regularization is more accurate than others. As evident from the table III, the predicted values of the Bayesian Regularization efforts is very close to the expected or actual values as compare to LM, Back propagation and COCOMO.

VII. Conclusion

Effort Estimation is one of the crucial tasks in software project management. This simulation with NASA93 dataset has been carried out using tool created with the help of MATLAB. Neural Network is trained using “trainlm”, “traingdx” and “trainbr” algorithm. The result from our simulation shows that Bayesian

Regularization gives the best performance, among the other training algorithms. We have experimented with 15 attributes of the COCOMO and further investigation can be done with other attributes and also concentration needed for process maturity.

References Références Referencias

1. Morgenshtern, Ofer, Tzvi Raz, and Dov Dvir. "Factors affecting duration and effort estimation errors in software development projects." Information and Software Technology 49, no. 8 (2007): 827-837.
2. Boehm, Barry, Chris Abts, and Sunita Chulani. "Software development cost estimation approaches—A survey." Annals of Software Engineering 10, no. 1 (2000): 177-205.
3. Matson, Jack E., Bruce E. Barrett, and Joseph M. Mellichamp. "Software development cost estimation using function points." Software Engineering, IEEE Transactions on 20, no. 4 (1994): 275-287.
4. Kemerer, Chris F. "An empirical validation of software cost estimation models."Communications of the ACM 30, no. 5 (1987): 416-429.
5. Burgess, Colin J., and Martin Lefley. "Can genetic programming improve software effort estimation? A comparative evaluation." Information and Software Technology 43, no. 14 (2001): 863-873.
6. Low, Graham C., and D. Ross Jeffery. "Function points in the estimation and evaluation of the software process." Software Engineering, IEEE Transactions on 16, no. 1 (1990): 64-71.
7. Boehm, Barry, Bradford Clark, Ellis Horowitz, Chris Westland, Ray Madachy, and Richard Selby. "Cost models for future software life cycle processes: COCOMO 2.0." Annals of software engineering 1, no. 1 (1995): 57-94.
8. Srinivasan, Krishnamoorthy, and Douglas Fisher. "Machine learning approaches to estimating software development effort." Software Engineering, IEEE Transactions on 21, no. 2 (1995): 126-137.
9. Dejaeger, Karel, Wouter Verbeke, David Martens, and Bart Baesens. "Data mining techniques for software effort estimation: a comparative study." Software Engineering, IEEE Transactions on 38, no. 2 (2012): 375-397.
10. Nisar, M. W., Wang, Y. J., & Elahi, M. (2008, October). Software development effort estimation using fuzzy logic-A survey. In Fuzzy Systems and Knowledge Discovery, 2008. FSKD'08. Fifth International Conference on (Vol. 1, pp. 421-427). IEEE.
11. Shan, Y., McKay, R. I., Lokan, C. J., & Essam, D. L. (2002, July). Software project effort estimation using genetic programming. In Communications, Circuits and Systems and West Sino Expositions, IEEE 2002 International Conference on (Vol. 2, pp. 1108-1112). IEEE.
12. Shepperd, M., Schofield, C., & Kitchenham, B. (1996, May). Effort estimation using analogy. In Proceedings of the 18th international conference on Software engineering (pp. 170-178). IEEE Computer Society.
13. Tadayon, N. (2005, April). Neural network approach for software cost estimation. In Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on (Vol. 2, pp. 815-818). IEEE.
14. Boehm, B.W. Software Engineering Economics. Prentice-Hall, Englewood Cliffs, N.J. 1981.
15. Simon Haykin, "Neural Networks: A Comprehensive Foundation", Second Edition, Prentice Hall, 1998.
16. Srinivasan, K., & Fisher, D. (1995). Machine learning approaches to estimating software development effort. Software Engineering, IEEE Transactions on, 21(2), 126-137.
17. Ali Idri and Taghi M. Khoshgoftaar& Alain Abran,"Can Neural Networks be easily Interpreted in Software Cost Estimation", IEEE Transaction, 2002, page: 1162-1167.
18. I.F. Barcelos Tronto, J.D. Simoes da Silva, N. Sant. Anna, "Comparison of Artificial Neural Network and Regression Models in Software Effort Estimation", INPE ePrint, Vol.1, 2006.
19. Mrinal Kanti Ghose, Roheet Bhatnagar and Vandana Bhattacharjee, "Comparing Some Neural Network Models for Software Development Effort Prediction", IEEE, 2011.
20. http://automatika.etf.bg.ac.rs/files/predmeti/os4nm/Materijali/03_BackPropagation/MATLAB_nnet_BackPropagation.pdf

Global Journal of Computer Science and Technology
Neural & Artificial Intelligence
Volume 13 Issue 2 Version 1.0 Year 2013
Type: Double Blind Peer Reviewed International Research Journal
Publisher: Global Journals Inc. (USA)
Online ISSN: 0975-4172 & Print ISSN: 0975-4350

Bayesian Network Model for Epidemiological Data (Radiation exposure and circulatory disease risk: Hiroshima and Nagasaki atomic bomb survivor data)

By Sagar Baviskar
College of Engineering Pune, India

Abstract - This documentation describes the implementation of Bayesian Network on Hiroshima Nagasaki atomic bomb survivor data, using "R" software. Bayesian networks, a state-of-the art representation of probabilistic knowledge by a graphical diagram, has emerged in recent years as essential for pattern recognition and classification in the healthcare field. Unlike some data mining techniques, Bayesian networks allow investigators to combine domain knowledge with statistical data. This tailored discussion presents the basic concepts of Bayesian networks and its use for building a health risk model on Epidemiological data. The main objectives of our study is to find interdependencies between various attributes of data and to determine the threshold value of radiation dosage under which death counts are negligible.

Keywords : *bayesian network; data mining; epidemiological data, health risk model, implementation of bayesian network in R.*

GJCST-D Classification : *C.2.1*

BAYESIAN NETWORK MODEL FOR EPIDEMIOLOGICAL DATA RADIATION EXPOSURE AND CIRCULATORY DISEASE RISK HIROSHIMA AND NAGASAKI ATOMIC BOMB SURVIVOR DATA

Strictly as per the compliance and regulations of:

Bayesian Network Model for Epidemiological Data

(Radiation exposure and circulatory disease risk: Hiroshima and Nagasaki atomic bomb survivor data)

Sagar Baviskar

Abstract - This documentation describes the implementation of Bayesian Network on Hiroshima Nagasaki atomic bomb survivor data, using "R" software. Bayesian networks, a state-of-the art representation of probabilistic knowledge by a graphical diagram, has emerged in recent years as essential for pattern recognition and classification in the healthcare field. Unlike some data mining techniques, Bayesian networks allow investigators to combine domain knowledge with statistical data. This tailored discussion presents the basic concepts of Bayesian networks and its use for building a health risk model on Epidemiological data. The main objectives of our study is to find interdependencies between various attributes of data and to determine the threshold value of radiation dosage under which death counts are negligible.

keywords : bayesian network; data mining; epidemiological data, health risk model, implementation of bayesian network in R.

I. Introduction

Our focus is on identification of the relationships between radiation exposure and its potential risk factors using Bayesian Network, with the emphasis on integrating medical domain knowledge and statistical data analysis.

A Bayesian network is a graphical model that encodes the joint probability distribution for a set of random variables. Here we consider Bayesian networks with mixed variables, *i.e.* the random variables in a network are both discrete and continuous types.

First, raw data are pre-processed into a format that is acceptable to the learning algorithms of Bayesian networks. Some important considerations are discussed to address the uniqueness of the data and the challenges of the learning.

Second, a Bayesian network is learned from the pre-processed data set by integrating medical domain knowledge and generic learning algorithms. Third, the relationships revealed by the Bayesian network are used for finding the probability of death count.To learn a Bayesian network, the user needs to supply a training data set and represent any prior knowledge available as a Bayesian network. We are implementing The Bayesian Network in "R" software in one section, we will explain detail implementation of Bayesian Network in R.

This report makes use of data obtained from the Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan. RERF is a private, non-profit foundation funded by the Japanese Ministry of Health, Labor and Welfare (MHLW) and the U.S. Department of Energy, the latter through the National Academy of Sciences. The conclusions in this report are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies [3] [7].

Abbreviations, notations and Acronyms: Bayesian Network (BN).

II. What is Bayesian Network?

Bayesian network is a graphical model where nodes represent random variables (the two terms are used interchangeably in this article) and arrows represent probabilistic dependencies between them.

The graphical structure G = (V; A) of a Bayesian network is a directed acyclic graph (DAG), where V is the node (or vertex) set and A is the arc (or edge) set. The DAG defines a factorization of the joint probability distribution of V = {X1; X2;…;Xn}, often called the global probability distribution, into a set of local probability distributions, one for each variable.

The Bayesian network is a state-of-the art representation of probabilistic knowledge. Bayesian networks represent domain knowledge qualitatively by the use of graphical diagrams with nodes and arrows that represent variables and the relationships among the variables. Quantitatively, the degree of dependencyis expressed by probabilistic terms.

III. Advantages of Bayesian Network as Data Mining Tool

First, Bayesian networks allow investigators to use their domain expert knowledge in the discovery process, while other techniques rely primarily on coded data to extract knowledge. Second, Bayesian network models can be more easily understood than many of the other techniques via the use of nodes and arrows.

Author : Computer Engineering College of Engineering Pune, India. E-mail : Sagarbaviskar91@gmail.com

These represent the variables of interest and the relationships of variables, respectively. Researchers can easily encode domain expert knowledge through the use of these graphical diagrams, and thus more easily understand and interpret the output of the Bayesian network. In addition, Bayesian network algorithms capitalize on this encoded knowledge to increase their efficiency in modeling process and accuracy in its predictive performance. Next, Bayesian networks are flexible in regards to missing information. Bayesian network models can produce relatively accurate prediction even in the situation where complete data are not available. Last, because Bayesian networks can incorporate domain knowledge into statistical data, Bayesian networks are less influenced by small sample size.

A more detailed discussion will enhance the understanding of how Bayesian networks operate and why they are particularly well-suited to the epidemiological data such as Radiation exposure and circulatory disease risk: Hiroshima and Nagasaki atomic bomb survivor data.

IV. Basic Probabilistic Concepts

Fundamentally, Bayesian networks are designed to, through the complex application of the well-developed probability theory (Bayes rule), obtain probabilities of unknown variables from known probabilistic relationships. To understand Bayesian networks, basic concepts such as the Bayesian probability approach, prior (or unconditional) probability, posterior (or conditional) probability, joint probability distribution, and Bayes rule, need to be discussed.

V. Bayesian Probability vs Classical Probability

There are differences between Bayesian probability and classical probability. The Bayesian probability of an event is a person degree of belief in that event; the classical probability is the probability that an event will occur. Contrary to classical probability, we do not need repeated trials to measure the Bayesian probability. Thus, Bayesian probability based on personal belief is useful where the probability cannot be measured, even by repeated experiments.

VI. Prior Probability

In a situation when no other information (evidence) I available, the probability of an event occurring is a prioror unconditional probability. The commonly used denotation of prior probability is P (A), where the event of A is occurring. Prior probability, P (A), is used only whenno other information is available. Also, denotation, P (^A), can be used to represent the prior probability of an event not occurring. For example, suppose Ineffective Airway Clearance denotes a binary variable whether or not a particular patient admitted in hospital has anursing diagnosis of Ineffective Airway Clearance. The prior probability of Ineffective Airway Clearance may be expressed (estimated) as P (Ineffective Airway Clearance) = 0.15, meaning that without the presence of any other evidence (information), a nurse may assume that a particular patient has a 15% chance of having an Ineffective Airway Clearance. In this example of P (Ineffective Airway Clearance), we can assume that they can have values such as present or absent. Thus, P (Ineffective Airway Clearance) is viewed as P (Ineffective Airway Clearance=present), and P (^Ineffective Airway Clearance) as P (Ineffective Airway Clearance=absent).

A probability term is also used to express random variables with multi-values in the nursing domain. For example, if we are interested in the random variable Cognition of a patient, this variable may have several possible values, such as very good, good, poor, and very poor. We might estimate them based on experience as:

P(Cognition=verygood)=0.60;
P(Cognition=good)=0.30;P(Cognition=poor)=0.08;and P(Cognition=very poor)=0.02. We can also state all the possible values of the random variable, Cognition, as P (Cognition) = (0.6, 0.3, 0.08, and 0.02), which can be defined as a probability distribution for the random variable Cognition.

VII. Conditional Probability

As discussed earlier, the probability of an event occurring is expressed as a prior or unconditional probability; once the evidence is obtained, it becomes posterior or conditional probability. Once we have new information B, we can use the conditional probability of A given B instead of P(A), which can be denoted as P(A|B).This means ''the probability of A, given B''. Suppose P (Ineffective Airway Clearance| Grunting) is estimated to be 0.60. This proposes that if a patient is observed to have a Grunting breathing sound, and no other information is available, and then the probability of the patient having an Ineffective Airway Clearance will be changed from 0.15 to 0.60. That is, without considering the presence of Grunting, the probability ofIn effective Airway Clearance (prior probability) is 0.15; while considering the presence of Grunting, the probability of Ineffective Airway Clearance (posterior probability) becomes 0.60.

VIII. Joint Probability Distribution

The joint probability distribution expresses all the probabilities of all combinations of different values of random variables. As mentioned in the Cognition example, the probability distribution of Cognition is a one dimension alvector of probability for all possible values of a variable. The joint probability distribution is

expres sed as an n-dimensional table (n> 1), which is called the joint probability table. The joint probability table consists of the probabilities of all possible events occurring. Table 2 illustrates an example of joint probability distribution with a two-dimensional table of the two variables Pain and Satisfaction with Care in the nursing care domain, in which each variable has three values. Because all events are mutually exclusive, the sum of all the cells is '1' in the joint probability table. This distribution can answer any probabilistic statemen of interest. Adding across a row or column gives the prior probability of a variable; for example, P(Pain=Level I)=0.3 + 0.15 + 0.01 = 0.46. P(Pain=Level I ∩Satisfaction with Care=High) can also be obtained which is 0.3.

IX. Bayes' Rule

This section demonstrates the details of updating prior probability to conditional (posterior) probability using Bayes_ rule. Conditional probabilities can be redefined in Eq. (1),

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

This equation can also be written as:

$$P(A \cap B) = P(A, B) = P(A|B)P(B) \quad (2)$$

$$P(A \cap B) = P(A, B) = P(B|A)P(A) \quad (3).$$

Based on two equations (Eq. (2) and (3)), we can induce the equation known as Bayes'rule in Eq. (5) (also Bayes' law or Bayes' theorem), by equating the two right hand sides and dividing by P (B)> 0

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (4)$$

Bayes' rule is useful in practice to estimate unknown P (A|B) from three probability terms (i.e., P (B|A), P (A) and P (B)) that nurses may be able to easily estimate in a domain. In a task estimating the probability of Ineffective Airway Clearance, there can be conditional probabilitie son causal relationships as in *Fig. 2:* Nurses may want to derive a nursing diagnosis given information by Grunting. A nurse knows that Ineffective Airway Clearance may cause a patient to have a Grunting breathing sound (an estimated 40% of the time). The nurse also knows some unconditional facts: suppose the prior probability of a patient having Ineffective Airway Clearance is 0.15, and the prior probability of any patient having Grunting is 0.10.When a nurse would like toestimateP(*Ineffective Airway Clearance|Grunting*) which may not be well-known probability, conditional probabilities can be induced based on Bayes' rule in Eq. (4).

P (Grunting Ineffective Airway Clearance)=0.40
P (Ineffective Airway Clearance)=0.15
P (Grunting)=0.10

According to these three probabilities

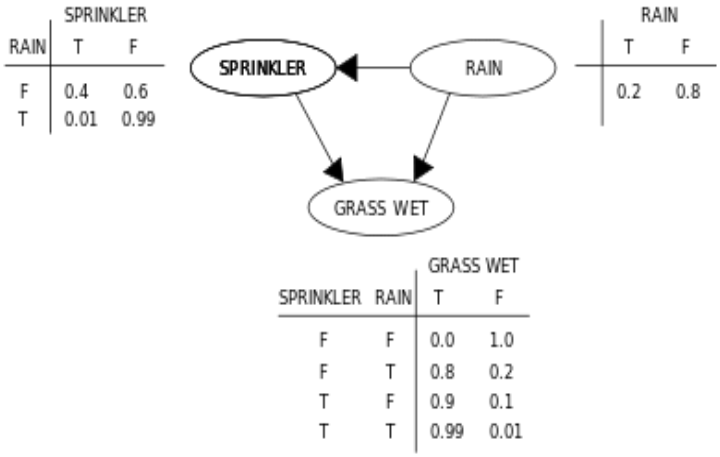
*P(Ineffective Airway Clearance| Grunting)=(P(Grunting|IneffectiveairwayClearance)*P(IneffectiveAirwayclearance))/(P(Grunting))=*

$$\frac{0.40 * 0.15}{0.10} = 0.60$$

This simple example of Bayes' rule demonstrates how unknown probabilities can be computed from the known.

X. A Typical Bayesian Network

Suppose that there are two events which could cause grass to be wet: either the sprinkler is on or it's raining. Also, suppose that the rain has a direct effect on the use of the sprinkler (namely that when it rains, the sprinkler is usually not turned on). Then the situation can be modeled with a Bayesian network (shown). All three variables have two possible values, T (for true) and F (for false).



The joint probability function is:

$$P(G,S,R)=P(G,S|R)P(S|R)P(R),$$

where the names of the variables have been abbreviated to *G = Grass wet (yes/no)*, *S = Sprinkler turned on (yes/no)*, and *R = Raining (yes/no)*.

The model can answer questions like "What is the probability that it is raining, given the grass is wet?" by using the conditional probability formula.

XI. About Data Set Used

The dataset which is used describe circulatory mortality in the Life Span Study of atomic bomb survivors. It is based on Radiation exposure and circulatory disease risk: Hiroshima and Nagasaki atomic bomb survivor (1950-2003). The data set is a detailed tabulation of person-years, case-counts, and summary data constructed from data on individual survivors. The cohort for analysis includes 86,661 survivors. Data on individual survivors are stratified by city, sex, age at exposure, attained age, calendar time, and dose. Cross-classification variables used to define the table are:1)Name 2) City 3) Sex 4) Agexcat 5) Agecat 6) Ctime. Variables that includes the cell-specific numbers of subjects entering the study:1) Dosecat 2) Subjects 3) PYR 4) Agex 5) Age 6) colon10. Disease death counts variables:1) CVD 2)stroke 3)heartd 4)othcvd 5)concvd 6)constroke 7)conheartd 8)conothcvd [3][7].

This report makes use of data obtained from the Radiation Effects Research Foundation (RERF), Hiroshima and Nagasaki, Japan. RERF is a private, non-profit foundation funded by the Japanese Ministry of Health, Labor and Welfare (MHLW) and the U.S. Department of Energy, the latter through the National Academy of Sciences. The conclusions in this report are those of the authors and do not necessarily reflect the scientific judgment of RERF or its funding agencies [3][7].

XII. Implementation of Bayesian Network in "R"

a) Various BN Algorithms

We are using R software to implement the Bayesian Network for the above dataset.

There are two packages named as 1) Deal and 2) BN Learn using which we can implement the BN.

In Deal package, we have Greedy algorithm to implement the BN for given dataset, and in BN learn we have

1. Constraint based algorithm
2. Score based algorithm

In Constraint based algorithm, we can implement the BN by using either Grow-Shrink algorithm or Incremental association markov blanket.

In Score based algorithm package, we have Hill climbing algorithm to build a BN. We will discuss the detailed implementation of BN using Deal package with Greedy algorithm and will just compare the results of rest of the algorithms in the following topics.

b) Implementation of BN in R using Deal Package

The data is in a file named as lsscvd10.csv which is a .csv file. So read the data from the csv file into R.

For the implementation we considered following categorical variables (1) City 2) Sex 3) Agexcat 4) Agecat 5) Ctime 6)Dosecatand continuous variables (CVD, STROKE, HEARTD, PYR, COLON10, SUBJECTS) [7].

Now after reading the data, we need to load the data into a data frame which is an acceptable form in R. Now as the first 6 variables are categorical we need to normalize the data by factorizing those variables.

In deal, a Bayesian network is represented as an object of class network. The network object is a list of properties that are added or changed. By default it is set to the empty network (the network without any arrows) [1].

If the option specify graph is set, a point and click graphical interface allows the user to insert and delete arrows until the requested DAG is obtained.
Note that discrete nodes are grey and continuous nodes are white.

The primary property of a network is the list of nodes. Each entry in the list is an object of class node representing a node in the graph, which includes information associated with the node. Several methods for the network class operate by applying an appropriate method for one or more nodes in the list of nodes. The nodes appear in the node list in the same order as in the data frame used to create the network object.

The parameters of the joint distribution of the variables in the network are then determined by the function joint prior () with the size of the imaginary data base as optional argument. If the size is not specified, deal sets the size to a reasonably small value [1].

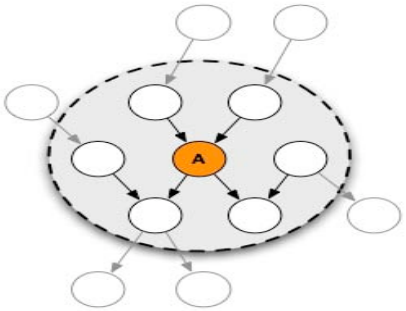
Then comes the most important function which is Learn () function. Using this function, the dataset is learned by R software for finding the relationship between the covariates, i.e, Random variables [1].

To get the best BN the heuristic searching technique is used. The search algorithm is used with restarts which is implemented in the function heuristic (). The initial network is then perturbed according to the parameter degree and the search is performed starting with the perturbed network [1].

After that using fit () function we can compute the probabilities which are desired [1].

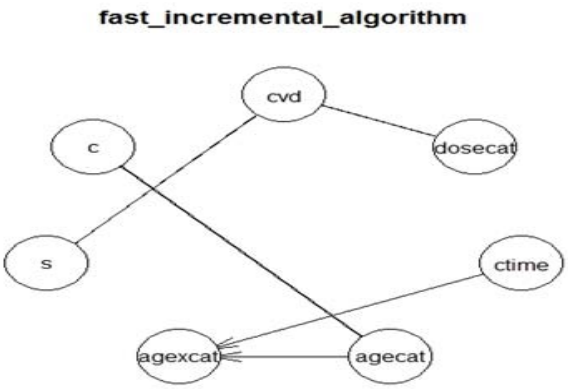
XIII. Outputs of Various BN Algorithms

a) Markov Blanket



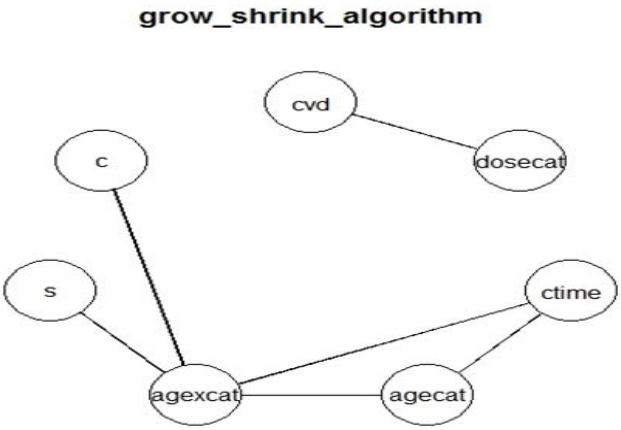
Markov blanket of a variable is set of variables containing its parent, children and children other parents. In figure Markov Blanket of a variable A is set of all variables shown in circle. Those outside circle are not part of Markov blanket of variable A [2].

b) Incremental Algorithm



This algorithm is based on the Markov blanket detection algorithm of the same name, which is based on a two-phase selection scheme (a forward selection followed by an attempt to remove false positives). This algorithm is a variant of Incremental Association which uses speculative stepwise forward selection to reduce the number of conditional independence tests [2].

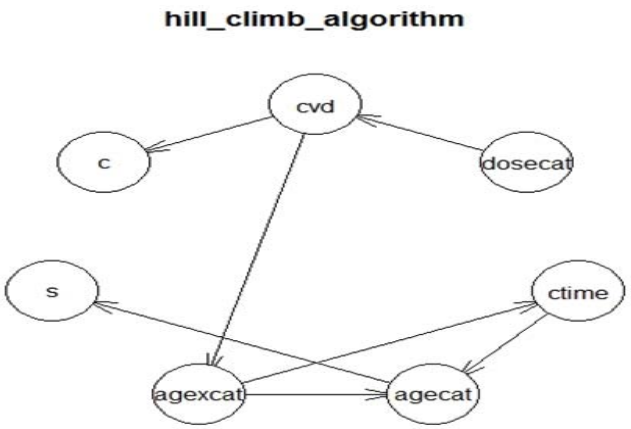
c) Grow Shrink Algorithm



This algorithm is based on the Grow-Shrink Markov Blanket, the first (and simplest)
Markov blanket detection algorithm used in a structure learning algorithm.

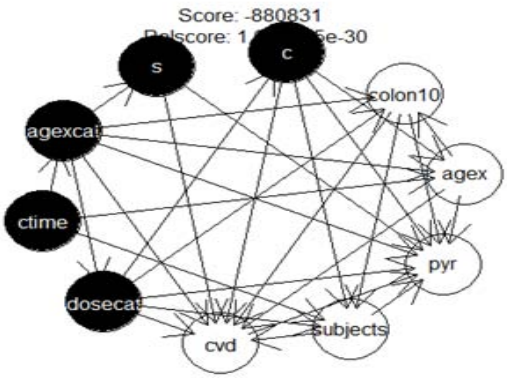
But all the above algorithms are not finding exact relationship between the variables, it means the dependency between variables is unidirectional. So exact inference cannot be drawn using these algorithms. Also these algorithms are not exhaustive.

d) Hill Climbing Algorithm



This algorithm finds the optimal network structure in the restricted space. A hill climbing greedy search on the space of the directed graphs. The optimized implementation uses score caching, score decomposability and score quivalence to reduce the number of duplicated tests. But the network score obtained is much less than that of Greedy algorithm as hill climb algorithm is not an exhaustive algorithm [2].

e) Greedy Algorithm



A greedy algorithm is an algorithm that follows the problem solving heuristic of making the locally optimal choice at each stage with the hope of finding a global optimum. The network score obtained is much higher as greedy algorithm is an exhaustive algorithm. Because of this we got the desired Bayesian Network [2].

XIV. Conclusion

So from the above discussion we can understand that the Greedy algorithm is the best algorithm to implement the Bayesian Network for the given dataset which considers all the possible relationship between the variables and finds a complete Bayesian network(Note: It is not mandatory that the same algorithm is best suitable for all the data.). Also, using the probability distribution obtained from above network, we found that radiation dosage below 0.5 Gy (dosecat 0-13) have negligible effect on death count (CVD, STROKE, HEARTD).

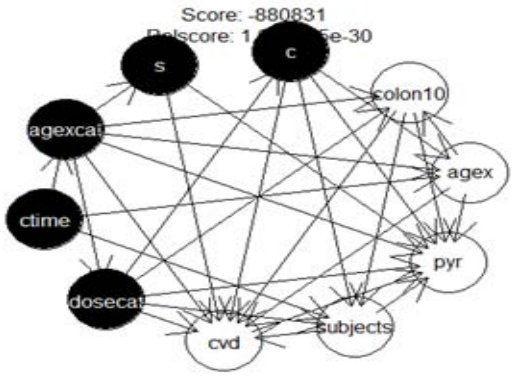


Figure : CVD_BN

Joint probability distribution model for cvdis: [c|dosecat][s|agexcat][agexcat|ctime][ctime][dosecat|agexcat][cvd|dosecat:agexcat:s:c:colon10:agex:subjects]

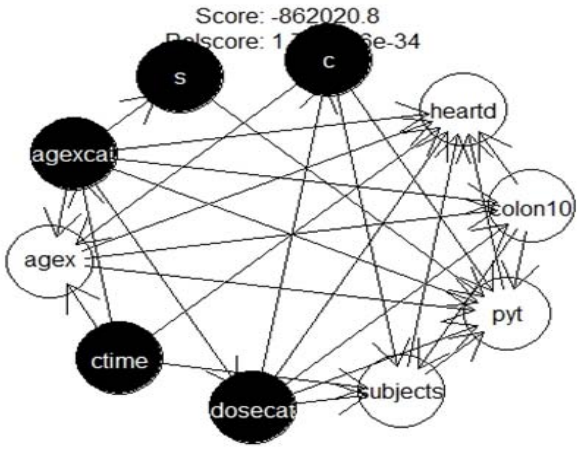


Figure : Heart DBN

Joint probability distribution model for heartd is [c|dosecat][s|agexcat][agexcat|ctime][ctime][dosecat|agexcat] [heartd dosecat gexcat:pyr:ctime:colon10:agex:subject].

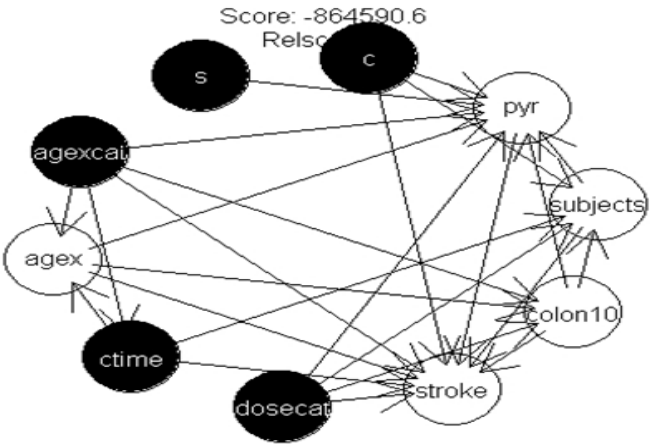


Figure : STROKE_BN

Joint probability distribution model for stroke is:

[c|dosecat][s|agexcat][agexcat|ctime][ctime][dosecat|agexcat][stroke|dosecat:agexcat:s:c:olon10:agex:subject]

The above three are the probabilistic distribution models for CVD, STROKE and Heart D.

References Références Referencias

1. "Susanne G. Bøttcher, ClausDethlefsen", "deal: A Package for Learning Bayesian Networks", "April 2003", "bayesian network with deal".
2. "Marco Scutari",”Learning Bayesian Networks with the bnlearn R Package", "Journal of statistical software", "June 2009","bayesian network with bnlearn".
3. "Yukiko Shimizu", "Radiation exposure and circulatory disease risk: Hiroshima and Nagasaki atomic bomb survivor data, 1950-2003","2010", "Research paper on survivor data".
4. "Jing Li, Jianjun Shi and Devin Satz", "Modeling and Analysis of Disease and Risk Factors through Learning Bayesian Networks from Observational Data", "WileyInter Science", “November 2007", "www.interscience.wiley.com"
5. "Susanne G. Bøttcher, Claus Dethlefsen", Learning Bayesian Networks with R", “3rd International Workshop on Distributed Statistical Computing (DSC 2003)", "March 2003", “http://www.ci.tuwien.ac.at/Conferences/DSC-2003/"
6. "Luis Targo", "Data mining with R", "Chapman and Hall, CRC press", “April 2011", "http://www.crcpress.com"
7. "Shimizu Y", "Life Span Study Circulatory Disease Mortality Data, 1950-2003", “RERF foundation", "2010", "http://www.rerf.or.jp/"
8. “Modeling and Analysis of Disease and Risk Factors through learning Bayesian Network from Observational Data”, Jing Li, Jianjun Shi, Devin Satz.

9. "Radiation exposure and circulatory disease risk: Hiroshima and Nagasaki atomic bomb survivor data", Yukiko Shimizu, Nobuo Nishi, Fumiyoshi Kasagi
10. "Learning Bayesian Networks with R", Susanne G. Bottcher, Claus Deathlefsen
11. "Learning Bayesian networks with bnlearn R package", Marco Scutari
12. "Deal: A Package for Learning Bayesian Networks", Susanne G. Bottcher, Claus Death lefsen.
13. Cios KJ. Medical data mining and knowledge discovery. New York: Physica-Verlag; 2001.
14. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. JClinEpidemiol 1996; 49:1225–32.
15. Heckerman D. Bayesian networks for data mining. Data Min Knowledge Discovery997; 1:79–119.
16. Heckerman DE. A tutorial on learning with Bayesian networks. MSR-TR-95-06. 1996. Redmond, WA, Microsoft Research.
17. Hamilton PW, Montironi R, Abmayr W, et al. Clinical applications of Bayesian belief networks in pathology. Pathologica 1995; 87(3):237–45.

Global Journals Inc. (US) Guidelines Handbook 2013

www.GlobalJournals.org

FELLOW OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (FARSC)

- 'FARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'FARSC" can be added to name in the following manner. eg. **Dr. John E. Hall, Ph.D., FARSC or William Walldroff Ph. D., M.S., FARSC**
- Being FARSC is a respectful honor. It authenticates your research activities. After becoming FARSC, you can use 'FARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 60% Discount will be provided to FARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- FARSC will be given a renowned, secure, free professional email address with 100 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- FARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 15% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- Eg. If we had taken 420 USD from author, we can send 63 USD to your account.
- FARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- After you are FARSC. You can send us scanned copy of all of your documents. We will verify, grade and certify them within a month. It will be based on your academic records, quality of research papers published by you, and 50 more criteria. This is beneficial for your job interviews as recruiting organization need not just rely on you for authenticity and your unknown qualities, you would have authentic ranks of all of your documents. Our scale is unique worldwide.
- FARSC member can proceed to get benefits of free research podcasting in Global Research Radio with their research documents, slides and online movies.
- After your publication anywhere in the world, you can upload you research paper with your recorded voice or you can use our professional RJs to record your paper their voice. We can also stream your conference videos and display your slides online.
- FARSC will be eligible for free application of Standardization of their Researches by Open Scientific Standards. Standardization is next step and level after publishing in a journal. A team of research and professional will work with you to take your research to its next level, which is worldwide open standardization.

- FARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), FARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 80% of its earning by Global Journals Inc. (US) will be transferred to FARSC member's bank account after certain threshold balance. There is no time limit for collection. FARSC member can decide its price and we can help in decision.

MEMBER OF ASSOCIATION OF RESEARCH SOCIETY IN COMPUTING (MARSC)

- 'MARSC' title will be awarded to the person after approval of Editor-in-Chief and Editorial Board. The title 'MARSC" can be added to name in the following manner. eg. Dr. John E. Hall, Ph.D., MARSC or William Walldroff Ph. D., M.S., MARSC
- Being MARSC is a respectful honor. It authenticates your research activities. After becoming MARSC, you can use 'MARSC' title as you use your degree in suffix of your name. This will definitely will enhance and add up your name. You can use it on your Career Counseling Materials/CV/Resume/Visiting Card/Name Plate etc.
- 40% Discount will be provided to MARSC members for publishing research papers in Global Journals Inc., if our Editorial Board and Peer Reviewers accept the paper. For the life time, if you are author/co-author of any paper bill sent to you will automatically be discounted one by 60%
- MARSC will be given a renowned, secure, free professional email address with 30 GB of space eg.johnhall@globaljournals.org. You will be facilitated with Webmail, Spam Assassin, Email Forwarders, Auto-Responders, Email Delivery Route tracing, etc.
- MARSC member is eligible to become paid peer reviewer at Global Journals Inc. to earn up to 10% of realized author charges taken from author of respective paper. After reviewing 5 or more papers you can request to transfer the amount to your bank account or to your PayPal account.
- MARSC member can apply for free approval, grading and certification of some of their Educational and Institutional Degrees from Global Journals Inc. (US) and Open Association of Research, Society U.S.A.
- MARSC is eligible to earn from their researches: While publishing his paper with Global Journals Inc. (US), MARSC can decide whether he/she would like to publish his/her research in closed manner. When readers will buy that individual research paper for reading, 40% of its earning by Global Journals Inc. (US) will be transferred to MARSC member's bank account after certain threshold balance. There is no time limit for collection. MARSC member can decide its price and we can help in decision.

AUXILIARY MEMBERSHIPS

ANNUAL MEMBER

- Annual Member will be authorized to receive e-Journal GJCST for one year (subscription for one year).
- The member will be allotted free 1 GB Web-space along with subDomain to contribute and participate in our activities.
- A professional email address will be allotted free 500 MB email space.

PAPER PUBLICATION

- The members can publish paper once. The paper will be sent to two-peer reviewer. The paper will be published after the acceptance of peer reviewers and Editorial Board.

Process of submission of Research Paper

The Area or field of specialization may or may not be of any category as mentioned in 'Scope of Journal' menu of the GlobalJournals.org website. There are 37 Research Journal categorized with Six parental Journals GJCST, GJMR, GJRE, GJMBR, GJSFR, GJHSS. For Authors should prefer the mentioned categories. There are three widely used systems UDC, DDC and LCC. The details are available as 'Knowledge Abstract' at Home page. The major advantage of this coding is that, the research work will be exposed to and shared with all over the world as we are being abstracted and indexed worldwide.

The paper should be in proper format. The format can be downloaded from first page of 'Author Guideline' Menu. The Author is expected to follow the general rules as mentioned in this menu. The paper should be written in MS-Word Format (*.DOC,*.DOCX).

The Author can submit the paper either online or offline. The authors should prefer online submission.Online Submission: There are three ways to submit your paper:

(A) (I) First, register yourself using top right corner of Home page then Login. If you are already registered, then login using your username and password.

(II) Choose corresponding Journal.

(III) Click 'Submit Manuscript'. Fill required information and Upload the paper.

(B) If you are using Internet Explorer, then Direct Submission through Homepage is also available.

(C) If these two are not convenient, and then email the paper directly to dean@globaljournals.org.

Offline Submission: Author can send the typed form of paper by Post. However, online submission should be preferred.

Preferred Author Guidelines

MANUSCRIPT STYLE INSTRUCTION (Must be strictly followed)

Page Size: 8.27" X 11'"

- Left Margin: 0.65
- Right Margin: 0.65
- Top Margin: 0.75
- Bottom Margin: 0.75
- Font type of all text should be Swis 721 Lt BT.
- Paper Title should be of Font Size 24 with one Column section.
- Author Name in Font Size of 11 with one column as of Title.
- Abstract Font size of 9 Bold, "Abstract" word in Italic Bold.
- Main Text: Font size 10 with justified two columns section
- Two Column with Equal Column with of 3.38 and Gaping of .2
- First Character must be three lines Drop capped.
- Paragraph before Spacing of 1 pt and After of 0 pt.
- Line Spacing of 1 pt
- Large Images must be in One Column
- Numbering of First Main Headings (Heading 1) must be in Roman Letters, Capital Letter, and Font Size of 10.
- Numbering of Second Main Headings (Heading 2) must be in Alphabets, Italic, and Font Size of 10.

You can use your own standard format also.
Author Guidelines:

1. General,

2. Ethical Guidelines,

3. Submission of Manuscripts,

4. Manuscript's Category,

5. Structure and Format of Manuscript,

6. After Acceptance.

1. GENERAL

Before submitting your research paper, one is advised to go through the details as mentioned in following heads. It will be beneficial, while peer reviewer justify your paper for publication.

Scope

The Global Journals Inc. (US) welcome the submission of original paper, review paper, survey article relevant to the all the streams of Philosophy and knowledge. The Global Journals Inc. (US) is parental platform for Global Journal of Computer Science and Technology, Researches in Engineering, Medical Research, Science Frontier Research, Human Social Science, Management, and Business organization. The choice of specific field can be done otherwise as following in Abstracting and Indexing Page on this Website. As the all Global

Journals Inc. (US) are being abstracted and indexed (in process) by most of the reputed organizations. Topics of only narrow interest will not be accepted unless they have wider potential or consequences.

2. ETHICAL GUIDELINES

Authors should follow the ethical guidelines as mentioned below for publication of research paper and research activities.

Papers are accepted on strict understanding that the material in whole or in part has not been, nor is being, considered for publication elsewhere. If the paper once accepted by Global Journals Inc. (US) and Editorial Board, will become the copyright of the Global Journals Inc. (US).

Authorship: The authors and coauthors should have active contribution to conception design, analysis and interpretation of findings. They should critically review the contents and drafting of the paper. All should approve the final version of the paper before submission

The Global Journals Inc. (US) follows the definition of authorship set up by the Global Academy of Research and Development. According to the Global Academy of R&D authorship, criteria must be based on:

1) Substantial contributions to conception and acquisition of data, analysis and interpretation of the findings.

2) Drafting the paper and revising it critically regarding important academic content.

3) Final approval of the version of the paper to be published.

All authors should have been credited according to their appropriate contribution in research activity and preparing paper. Contributors who do not match the criteria as authors may be mentioned under Acknowledgement.

Acknowledgements: Contributors to the research other than authors credited should be mentioned under acknowledgement. The specifications of the source of funding for the research if appropriate can be included. Suppliers of resources may be mentioned along with address.

Appeal of Decision: The Editorial Board's decision on publication of the paper is final and cannot be appealed elsewhere.

Permissions: It is the author's responsibility to have prior permission if all or parts of earlier published illustrations are used in this paper.

Please mention proper reference and appropriate acknowledgements wherever expected.

If all or parts of previously published illustrations are used, permission must be taken from the copyright holder concerned. It is the author's responsibility to take these in writing.

Approval for reproduction/modification of any information (including figures and tables) published elsewhere must be obtained by the authors/copyright holders before submission of the manuscript. Contributors (Authors) are responsible for any copyright fee involved.

3. SUBMISSION OF MANUSCRIPTS

Manuscripts should be uploaded via this online submission page. The online submission is most efficient method for submission of papers, as it enables rapid distribution of manuscripts and consequently speeds up the review procedure. It also enables authors to know the status of their own manuscripts by emailing us. Complete instructions for submitting a paper is available below.

Manuscript submission is a systematic procedure and little preparation is required beyond having all parts of your manuscript in a given format and a computer with an Internet connection and a Web browser. Full help and instructions are provided on-screen. As an author, you will be prompted for login and manuscript details as Field of Paper and then to upload your manuscript file(s) according to the instructions.

To avoid postal delays, all transaction is preferred by e-mail. A finished manuscript submission is confirmed by e-mail immediately and your paper enters the editorial process with no postal delays. When a conclusion is made about the publication of your paper by our Editorial Board, revisions can be submitted online with the same procedure, with an occasion to view and respond to all comments.

Complete support for both authors and co-author is provided.

4. MANUSCRIPT'S CATEGORY

Based on potential and nature, the manuscript can be categorized under the following heads:

Original research paper: Such papers are reports of high-level significant original research work.

Review papers: These are concise, significant but helpful and decisive topics for young researchers.

Research articles: These are handled with small investigation and applications.

Research letters: The letters are small and concise comments on previously published matters.

5. STRUCTURE AND FORMAT OF MANUSCRIPT

The recommended size of original research paper is less than seven thousand words, review papers fewer than seven thousands words also.Preparation of research paper or how to write research paper, are major hurdle, while writing manuscript. The research articles and research letters should be fewer than three thousand words, the structure original research paper; sometime review paper should be as follows:

Papers: These are reports of significant research (typically less than 7000 words equivalent, including tables, figures, references), and comprise:

(a)Title should be relevant and commensurate with the theme of the paper.

(b) A brief Summary, "Abstract" (less than 150 words) containing the major results and conclusions.

(c) Up to ten keywords, that precisely identifies the paper's subject, purpose, and focus.

(d) An Introduction, giving necessary background excluding subheadings; objectives must be clearly declared.

(e) Resources and techniques with sufficient complete experimental details (wherever possible by reference) to permit repetition; sources of information must be given and numerical methods must be specified by reference, unless non-standard.

(f) Results should be presented concisely, by well-designed tables and/or figures; the same data may not be used in both; suitable statistical data should be given. All data must be obtained with attention to numerical detail in the planning stage. As reproduced design has been recognized to be important to experiments for a considerable time, the Editor has decided that any paper that appears not to have adequate numerical treatments of the data will be returned un-refereed;

(g) Discussion should cover the implications and consequences, not just recapitulating the results; conclusions should be summarizing.

(h) Brief Acknowledgements.

(i) References in the proper form.

Authors should very cautiously consider the preparation of papers to ensure that they communicate efficiently. Papers are much more likely to be accepted, if they are cautiously designed and laid out, contain few or no errors, are summarizing, and be conventional to the approach and instructions. They will in addition, be published with much less delays than those that require much technical and editorial correction.

The Editorial Board reserves the right to make literary corrections and to make suggestions to improve briefness.

It is vital, that authors take care in submitting a manuscript that is written in simple language and adheres to published guidelines.

Format

Language: The language of publication is UK English. Authors, for whom English is a second language, must have their manuscript efficiently edited by an English-speaking person before submission to make sure that, the English is of high excellence. It is preferable, that manuscripts should be professionally edited.

Standard Usage, Abbreviations, and Units: Spelling and hyphenation should be conventional to The Concise Oxford English Dictionary. Statistics and measurements should at all times be given in figures, e.g. 16 min, except for when the number begins a sentence. When the number does not refer to a unit of measurement it should be spelt in full unless, it is 160 or greater.

Abbreviations supposed to be used carefully. The abbreviated name or expression is supposed to be cited in full at first usage, followed by the conventional abbreviation in parentheses.

Metric SI units are supposed to generally be used excluding where they conflict with current practice or are confusing. For illustration, 1.4 l rather than 1.4 × 10-3 m3, or 4 mm somewhat than 4 × 10-3 m. Chemical formula and solutions must identify the form used, e.g. anhydrous or hydrated, and the concentration must be in clearly defined units. Common species names should be followed by underlines at the first mention. For following use the generic name should be constricted to a single letter, if it is clear.

Structure

All manuscripts submitted to Global Journals Inc. (US), ought to include:

Title: The title page must carry an instructive title that reflects the content, a running title (less than 45 characters together with spaces), names of the authors and co-authors, and the place(s) wherever the work was carried out. The full postal address in addition with the e-mail address of related author must be given. Up to eleven keywords or very brief phrases have to be given to help data retrieval, mining and indexing.

Abstract, used in Original Papers and Reviews:

Optimizing Abstract for Search Engines

Many researchers searching for information online will use search engines such as Google, Yahoo or similar. By optimizing your paper for search engines, you will amplify the chance of someone finding it. This in turn will make it more likely to be viewed and/or cited in a further work. Global Journals Inc. (US) have compiled these guidelines to facilitate you to maximize the web-friendliness of the most public part of your paper.

Key Words

A major linchpin in research work for the writing research paper is the keyword search, which one will employ to find both library and Internet resources.

One must be persistent and creative in using keywords. An effective keyword search requires a strategy and planning a list of possible keywords and phrases to try.

Search engines for most searches, use Boolean searching, which is somewhat different from Internet searches. The Boolean search uses "operators," words (and, or, not, and near) that enable you to expand or narrow your affords. Tips for research paper while preparing research paper are very helpful guideline of research paper.

Choice of key words is first tool of tips to write research paper. Research paper writing is an art.A few tips for deciding as strategically as possible about keyword search:

- One should start brainstorming lists of possible keywords before even begin searching. Think about the most important concepts related to research work. Ask, "What words would a source have to include to be truly valuable in research paper?" Then consider synonyms for the important words.
- It may take the discovery of only one relevant paper to let steer in the right keyword direction because in most databases, the keywords under which a research paper is abstracted are listed with the paper.
- One should avoid outdated words.

Keywords are the key that opens a door to research work sources. Keyword searching is an art in which researcher's skills are bound to improve with experience and time.

Numerical Methods: Numerical methods used should be clear and, where appropriate, supported by references.

Acknowledgements: Please make these as concise as possible.

References

References follow the Harvard scheme of referencing. References in the text should cite the authors' names followed by the time of their publication, unless there are three or more authors when simply the first author's name is quoted followed by et al. unpublished work has to only be cited where necessary, and only in the text. Copies of references in press in other journals have to be supplied with submitted typescripts. It is necessary that all citations and references be carefully checked before submission, as mistakes or omissions will cause delays.

References to information on the World Wide Web can be given, but only if the information is available without charge to readers on an official site. Wikipedia and Similar websites are not allowed where anyone can change the information. Authors will be asked to make available electronic copies of the cited information for inclusion on the Global Journals Inc. (US) homepage at the judgment of the Editorial Board.

The Editorial Board and Global Journals Inc. (US) recommend that, citation of online-published papers and other material should be done via a DOI (digital object identifier). If an author cites anything, which does not have a DOI, they run the risk of the cited material not being noticeable.

The Editorial Board and Global Journals Inc. (US) recommend the use of a tool such as Reference Manager for reference management and formatting.

Tables, Figures and Figure Legends

Tables: Tables should be few in number, cautiously designed, uncrowned, and include only essential data. Each must have an Arabic number, e.g. Table 4, a self-explanatory caption and be on a separate sheet. Vertical lines should not be used.

Figures: Figures are supposed to be submitted as separate files. Always take in a citation in the text for each figure using Arabic numbers, e.g. Fig. 4. Artwork must be submitted online in electronic form by e-mailing them.

Preparation of Electronic Figures for Publication

Even though low quality images are sufficient for review purposes, print publication requires high quality images to prevent the final product being blurred or fuzzy. Submit (or e-mail) EPS (line art) or TIFF (halftone/photographs) files only. MS PowerPoint and Word Graphics are unsuitable for printed pictures. Do not use pixel-oriented software. Scans (TIFF only) should have a resolution of at least 350 dpi (halftone) or 700 to 1100 dpi (line drawings) in relation to the imitation size. Please give the data for figures in black and white or submit a Color Work Agreement Form. EPS files must be saved with fonts embedded (and with a TIFF preview, if possible).

For scanned images, the scanning resolution (at final image size) ought to be as follows to ensure good reproduction: line art: >650 dpi; halftones (including gel photographs) : >350 dpi; figures containing both halftone and line images: >650 dpi.

Color Charges: It is the rule of the Global Journals Inc. (US) for authors to pay the full cost for the reproduction of their color artwork. Hence, please note that, if there is color artwork in your manuscript when it is accepted for publication, we would require you to complete and return a color work agreement form before your paper can be published.

Figure Legends: Self-explanatory legends of all figures should be incorporated separately under the heading 'Legends to Figures'. In the full-text online edition of the journal, figure legends may possibly be truncated in abbreviated links to the full screen version. Therefore, the first 100 characters of any legend should notify the reader, about the key aspects of the figure.

6. AFTER ACCEPTANCE

Upon approval of a paper for publication, the manuscript will be forwarded to the dean, who is responsible for the publication of the Global Journals Inc. (US).

6.1 Proof Corrections

The corresponding author will receive an e-mail alert containing a link to a website or will be attached. A working e-mail address must therefore be provided for the related author.

Acrobat Reader will be required in order to read this file. This software can be downloaded

(Free of charge) from the following website:

www.adobe.com/products/acrobat/readstep2.html. This will facilitate the file to be opened, read on screen, and printed out in order for any corrections to be added. Further instructions will be sent with the proof.

Proofs must be returned to the dean at dean@globaljournals.org within three days of receipt.

As changes to proofs are costly, we inquire that you only correct typesetting errors. All illustrations are retained by the publisher. Please note that the authors are responsible for all statements made in their work, including changes made by the copy editor.

6.2 Early View of Global Journals Inc. (US) (Publication Prior to Print)

The Global Journals Inc. (US) are enclosed by our publishing's Early View service. Early View articles are complete full-text articles sent in advance of their publication. Early View articles are absolute and final. They have been completely reviewed, revised and edited for publication, and the authors' final corrections have been incorporated. Because they are in final form, no changes can be made after sending them. The nature of Early View articles means that they do not yet have volume, issue or page numbers, so Early View articles cannot be cited in the conventional way.

6.3 Author Services

Online production tracking is available for your article through Author Services. Author Services enables authors to track their article - once it has been accepted - through the production process to publication online and in print. Authors can check the status of their articles online and choose to receive automated e-mails at key stages of production. The authors will receive an e-mail with a unique link that enables them to register and have their article automatically added to the system. Please ensure that a complete e-mail address is provided when submitting the manuscript.

6.4 Author Material Archive Policy

Please note that if not specifically requested, publisher will dispose off hardcopy & electronic information submitted, after the two months of publication. If you require the return of any information submitted, please inform the Editorial Board or dean as soon as possible.

6.5 Offprint and Extra Copies

A PDF offprint of the online-published article will be provided free of charge to the related author, and may be distributed according to the Publisher's terms and conditions. Additional paper offprint may be ordered by emailing us at: editor@globaljournals.org .

You must strictly follow above Author Guidelines before submitting your paper or else we will not at all be responsible for any corrections in future in any of the way.

Before start writing a good quality Computer Science Research Paper, let us first understand what is Computer Science Research Paper? So, Computer Science Research Paper is the paper which is written by professionals or scientists who are associated to Computer Science and Information Technology, or doing research study in these areas. If you are novel to this field then you can consult about this field from your supervisor or guide.

TECHNIQUES FOR WRITING A GOOD QUALITY RESEARCH PAPER:

1. Choosing the topic: In most cases, the topic is searched by the interest of author but it can be also suggested by the guides. You can have several topics and then you can judge that in which topic or subject you are finding yourself most comfortable. This can be done by asking several questions to yourself, like Will I be able to carry our search in this area? Will I find all necessary recourses to accomplish the search? Will I be able to find all information in this field area? If the answer of these types of questions will be "Yes" then you can choose that topic. In most of the cases, you may have to conduct the surveys and have to visit several places because this field is related to Computer Science and Information Technology. Also, you may have to do a lot of work to find all rise and falls regarding the various data of that subject. Sometimes, detailed information plays a vital role, instead of short information.

2. Evaluators are human: First thing to remember that evaluators are also human being. They are not only meant for rejecting a paper. They are here to evaluate your paper. So, present your Best.

3. Think Like Evaluators: If you are in a confusion or getting demotivated that your paper will be accepted by evaluators or not, then think and try to evaluate your paper like an Evaluator. Try to understand that what an evaluator wants in your research paper and automatically you will have your answer.

4. Make blueprints of paper: The outline is the plan or framework that will help you to arrange your thoughts. It will make your paper logical. But remember that all points of your outline must be related to the topic you have chosen.

5. Ask your Guides: If you are having any difficulty in your research, then do not hesitate to share your difficulty to your guide (if you have any). They will surely help you out and resolve your doubts. If you can't clarify what exactly you require for your work then ask the supervisor to help you with the alternative. He might also provide you the list of essential readings.

6. Use of computer is recommended: As you are doing research in the field of Computer Science, then this point is quite obvious.

7. Use right software: Always use good quality software packages. If you are not capable to judge good software then you can lose quality of your paper unknowingly. There are various software programs available to help you, which you can get through Internet.

8. Use the Internet for help: An excellent start for your paper can be by using the Google. It is an excellent search engine, where you can have your doubts resolved. You may also read some answers for the frequent question how to write my research paper or find model research paper. From the internet library you can download books. If you have all required books make important reading selecting and analyzing the specified information. Then put together research paper sketch out.

9. Use and get big pictures: Always use encyclopedias, Wikipedia to get pictures so that you can go into the depth.

10. Bookmarks are useful: When you read any book or magazine, you generally use bookmarks, right! It is a good habit, which helps to not to lose your continuity. You should always use bookmarks while searching on Internet also, which will make your search easier.

11. Revise what you wrote: When you write anything, always read it, summarize it and then finalize it.

12. Make all efforts: Make all efforts to mention what you are going to write in your paper. That means always have a good start. Try to mention everything in introduction, that what is the need of a particular research paper. Polish your work by good skill of writing and always give an evaluator, what he wants.

13. Have backups: When you are going to do any important thing like making research paper, you should always have backup copies of it either in your computer or in paper. This will help you to not to lose any of your important.

14. Produce good diagrams of your own: Always try to include good charts or diagrams in your paper to improve quality. Using several and unnecessary diagrams will degrade the quality of your paper by creating "hotchpotch." So always, try to make and include those diagrams, which are made by your own to improve readability and understandability of your paper.

15. Use of direct quotes: When you do research relevant to literature, history or current affairs then use of quotes become essential but if study is relevant to science then use of quotes is not preferable.

16. Use proper verb tense: Use proper verb tenses in your paper. Use past tense, to present those events that happened. Use present tense to indicate events that are going on. Use future tense to indicate future happening events. Use of improper and wrong tenses will confuse the evaluator. Avoid the sentences that are incomplete.

17. Never use online paper: If you are getting any paper on Internet, then never use it as your research paper because it might be possible that evaluator has already seen it or maybe it is outdated version.

18. Pick a good study spot: To do your research studies always try to pick a spot, which is quiet. Every spot is not for studies. Spot that suits you choose it and proceed further.

19. Know what you know: Always try to know, what you know by making objectives. Else, you will be confused and cannot achieve your target.

20. Use good quality grammar: Always use a good quality grammar and use words that will throw positive impact on evaluator. Use of good quality grammar does not mean to use tough words, that for each word the evaluator has to go through dictionary. Do not start sentence with a conjunction. Do not fragment sentences. Eliminate one-word sentences. Ignore passive voice. Do not ever use a big word when a diminutive one would suffice. Verbs have to be in agreement with their subjects. Prepositions are not expressions to finish sentences with. It is incorrect to ever divide an infinitive. Avoid clichés like the disease. Also, always shun irritating alliteration. Use language that is simple and straight forward. put together a neat summary.

21. Arrangement of information: Each section of the main body should start with an opening sentence and there should be a changeover at the end of the section. Give only valid and powerful arguments to your topic. You may also maintain your arguments with records.

22. Never start in last minute: Always start at right time and give enough time to research work. Leaving everything to the last minute will degrade your paper and spoil your work.

23. Multitasking in research is not good: Doing several things at the same time proves bad habit in case of research activity. Research is an area, where everything has a particular time slot. Divide your research work in parts and do particular part in particular time slot.

24. Never copy others' work: Never copy others' work and give it your name because if evaluator has seen it anywhere you will be in trouble.

25. Take proper rest and food: No matter how many hours you spend for your research activity, if you are not taking care of your health then all your efforts will be in vain. For a quality research, study is must, and this can be done by taking proper rest and food.

26. Go for seminars: Attend seminars if the topic is relevant to your research area. Utilize all your resources.

27. Refresh your mind after intervals: Try to give rest to your mind by listening to soft music or by sleeping in intervals. This will also improve your memory.

28. Make colleagues: Always try to make colleagues. No matter how sharper or intelligent you are, if you make colleagues you can have several ideas, which will be helpful for your research.

29. Think technically: Always think technically. If anything happens, then search its reasons, its benefits, and demerits.

30. Think and then print: When you will go to print your paper, notice that tables are not be split, headings are not detached from their descriptions, and page sequence is maintained.

31. Adding unnecessary information: Do not add unnecessary information, like, I have used MS Excel to draw graph. Do not add irrelevant and inappropriate material. These all will create superfluous. Foreign terminology and phrases are not apropos. One should NEVER take a broad view. Analogy in script is like feathers on a snake. Not at all use a large word when a very small one would be sufficient. Use words properly, regardless of how others use them. Remove quotations. Puns are for kids, not grunt readers. Amplification is a billion times of inferior quality than sarcasm.

32. Never oversimplify everything: To add material in your research paper, never go for oversimplification. This will definitely irritate the evaluator. Be more or less specific. Also too, by no means, ever use rhythmic redundancies. Contractions aren't essential and shouldn't be there used. Comparisons are as terrible as clichés. Give up ampersands and abbreviations, and so on. Remove commas, that are, not necessary. Parenthetical words however should be together with this in commas. Understatement is all the time the complete best way to put onward earth-shaking thoughts. Give a detailed literary review.

33. Report concluded results: Use concluded results. From raw data, filter the results and then conclude your studies based on measurements and observations taken. Significant figures and appropriate number of decimal places should be used. Parenthetical remarks are prohibitive. Proofread carefully at final stage. In the end give outline to your arguments. Spot out perspectives of further study of this subject. Justify your conclusion by at the bottom of them with sufficient justifications and examples.

34. After conclusion: Once you have concluded your research, the next most important step is to present your findings. Presentation is extremely important as it is the definite medium though which your research is going to be in print to the rest of the crowd. Care should be taken to categorize your thoughts well and present them in a logical and neat manner. A good quality research paper format is essential because it serves to highlight your research paper and bring to light all necessary aspects in your research.

Informal Guidelines of Research Paper Writing

Key points to remember:

- Submit all work in its final form.
- Write your paper in the form, which is presented in the guidelines using the template.
- Please note the criterion for grading the final paper by peer-reviewers.

Final Points:

A purpose of organizing a research paper is to let people to interpret your effort selectively. The journal requires the following sections, submitted in the order listed, each section to start on a new page.

The introduction will be compiled from reference matter and will reflect the design processes or outline of basis that direct you to make study. As you will carry out the process of study, the method and process section will be constructed as like that. The result segment will show related statistics in nearly sequential order and will direct the reviewers next to the similar intellectual paths throughout the data that you took to carry out your study. The discussion section will provide understanding of the data and projections as to the implication of the results. The use of good quality references all through the paper will give the effort trustworthiness by representing an alertness of prior workings.

Writing a research paper is not an easy job no matter how trouble-free the actual research or concept. Practice, excellent preparation, and controlled record keeping are the only means to make straightforward the progression.

General style:

Specific editorial column necessities for compliance of a manuscript will always take over from directions in these general guidelines.

To make a paper clear

· Adhere to recommended page limits

Mistakes to evade

- Insertion a title at the foot of a page with the subsequent text on the next page
- Separating a table/chart or figure - impound each figure/table to a single page
- Submitting a manuscript with pages out of sequence

In every sections of your document

· Use standard writing style including articles ("a", "the," etc.)

· Keep on paying attention on the research topic of the paper

· Use paragraphs to split each significant point (excluding for the abstract)

· Align the primary line of each section

· Present your points in sound order

· Use present tense to report well accepted

· Use past tense to describe specific results

· Shun familiar wording, don't address the reviewer directly, and don't use slang, slang language, or superlatives

· Shun use of extra pictures - include only those figures essential to presenting results

Title Page:

Choose a revealing title. It should be short. It should not have non-standard acronyms or abbreviations. It should not exceed two printed lines. It should include the name(s) and address (es) of all authors.

Abstract:

The summary should be two hundred words or less. It should briefly and clearly explain the key findings reported in the manuscript--must have precise statistics. It should not have abnormal acronyms or abbreviations. It should be logical in itself. Shun citing references at this point.

An abstract is a brief distinct paragraph summary of finished work or work in development. In a minute or less a reviewer can be taught the foundation behind the study, common approach to the problem, relevant results, and significant conclusions or new questions.

Write your summary when your paper is completed because how can you write the summary of anything which is not yet written? Wealth of terminology is very essential in abstract. Yet, use comprehensive sentences and do not let go readability for briefness. You can maintain it succinct by phrasing sentences so that they provide more than lone rationale. The author can at this moment go straight to shortening the outcome. Sum up the study, with the subsequent elements in any summary. Try to maintain the initial two items to no more than one ruling each.

- Reason of the study - theory, overall issue, purpose
- Fundamental goal
- To the point depiction of the research
- Consequences, including definite statistics - if the consequences are quantitative in nature, account quantitative data; results of any numerical analysis should be reported
- Significant conclusions or questions that track from the research(es)

Approach:

- Single section, and succinct
- As a outline of job done, it is always written in past tense
- A conceptual should situate on its own, and not submit to any other part of the paper such as a form or table
- Center on shortening results - bound background information to a verdict or two, if completely necessary
- What you account in an conceptual must be regular with what you reported in the manuscript
- Exact spelling, clearness of sentences and phrases, and appropriate reporting of quantities (proper units, important statistics) are just as significant in an abstract as they are anywhere else

Introduction:

The **Introduction** should "introduce" the manuscript. The reviewer should be presented with sufficient background information to be capable to comprehend and calculate the purpose of your study without having to submit to other works. The basis for the study should be offered. Give most important references but shun difficult to make a comprehensive appraisal of the topic. In the introduction, describe the problem visibly. If the problem is not acknowledged in a logical, reasonable way, the reviewer will have no attention in your result. Speak in common terms about techniques used to explain the problem, if needed, but do not present any particulars about the protocols here. Following approach can create a valuable beginning:

- Explain the value (significance) of the study
- Shield the model - why did you employ this particular system or method? What is its compensation? You strength remark on its appropriateness from a abstract point of vision as well as point out sensible reasons for using it.
- Present a justification. Status your particular theory (es) or aim(s), and describe the logic that led you to choose them.
- Very for a short time explain the tentative propose and how it skilled the declared objectives.

Approach:

- Use past tense except for when referring to recognized facts. After all, the manuscript will be submitted after the entire job is done.
- Sort out your thoughts; manufacture one key point with every section. If you make the four points listed above, you will need a least of four paragraphs.

- Present surroundings information only as desirable in order hold up a situation. The reviewer does not desire to read the whole thing you know about a topic.
- Shape the theory/purpose specifically - do not take a broad view.
- As always, give awareness to spelling, simplicity and correctness of sentences and phrases.

Procedures (Methods and Materials):

This part is supposed to be the easiest to carve if you have good skills. A sound written Procedures segment allows a capable scientist to replacement your results. Present precise information about your supplies. The suppliers and clarity of reagents can be helpful bits of information. Present methods in sequential order but linked methodologies can be grouped as a segment. Be concise when relating the protocols. Attempt for the least amount of information that would permit another capable scientist to spare your outcome but be cautious that vital information is integrated. The use of subheadings is suggested and ought to be synchronized with the results section. When a technique is used that has been well described in another object, mention the specific item describing a way but draw the basic principle while stating the situation. The purpose is to text all particular resources and broad procedures, so that another person may use some or all of the methods in one more study or referee the scientific value of your work. It is not to be a step by step report of the whole thing you did, nor is a methods section a set of orders.

Materials:

- Explain materials individually only if the study is so complex that it saves liberty this way.
- Embrace particular materials, and any tools or provisions that are not frequently found in laboratories.
- Do not take in frequently found.
- If use of a definite type of tools.
- Materials may be reported in a part section or else they may be recognized along with your measures.

Methods:

- Report the method (not particulars of each process that engaged the same methodology)
- Describe the method entirely
- To be succinct, present methods under headings dedicated to specific dealings or groups of measures
- Simplify - details how procedures were completed not how they were exclusively performed on a particular day.
- If well known procedures were used, account the procedure by name, possibly with reference, and that's all.

Approach:

- It is embarrassed or not possible to use vigorous voice when documenting methods with no using first person, which would focus the reviewer's interest on the researcher rather than the job. As a result when script up the methods most authors use third person passive voice.
- Use standard style in this and in every other part of the paper - avoid familiar lists, and use full sentences.

What to keep away from

- Resources and methods are not a set of information.
- Skip all descriptive information and surroundings - save it for the argument.
- Leave out information that is immaterial to a third party.

Results:

The principle of a results segment is to present and demonstrate your conclusion. Create this part a entirely objective details of the outcome, and save all understanding for the discussion.

The page length of this segment is set by the sum and types of data to be reported. Carry on to be to the point, by means of statistics and tables, if suitable, to present consequences most efficiently.You must obviously differentiate material that would usually be incorporated in a study editorial from any unprocessed data or additional appendix matter that would not be available. In fact, such matter should not be submitted at all except requested by the instructor.

Content

- Sum up your conclusion in text and demonstrate them, if suitable, with figures and tables.
- In manuscript, explain each of your consequences, point the reader to remarks that are most appropriate.
- Present a background, such as by describing the question that was addressed by creation an exacting study.
- Explain results of control experiments and comprise remarks that are not accessible in a prescribed figure or table, if appropriate.
- Examine your data, then prepare the analyzed (transformed) data in the form of a figure (graph), table, or in manuscript form.

What to stay away from

- Do not discuss or infer your outcome, report surroundings information, or try to explain anything.
- Not at all, take in raw data or intermediate calculations in a research manuscript.
- Do not present the similar data more than once.
- Manuscript should complement any figures or tables, not duplicate the identical information.
- Never confuse figures with tables - there is a difference.

Approach

- As forever, use past tense when you submit to your results, and put the whole thing in a reasonable order.
- Put figures and tables, appropriately numbered, in order at the end of the report
- If you desire, you may place your figures and tables properly within the text of your results part.

Figures and tables

- If you put figures and tables at the end of the details, make certain that they are visibly distinguished from any attach appendix materials, such as raw facts
- Despite of position, each figure must be numbered one after the other and complete with subtitle
- In spite of position, each table must be titled, numbered one after the other and complete with heading
- All figure and table must be adequately complete that it could situate on its own, divide from text

Discussion:

The Discussion is expected the trickiest segment to write and describe. A lot of papers submitted for journal are discarded based on problems with the Discussion. There is no head of state for how long a argument should be. Position your understanding of the outcome visibly to lead the reviewer through your conclusions, and then finish the paper with a summing up of the implication of the study. The purpose here is to offer an understanding of your results and hold up for all of your conclusions, using facts from your research and generally accepted information, if suitable. The implication of result should be visibly described. Infer your data in the conversation in suitable depth. This means that when you clarify an observable fact you must explain mechanisms that may account for the observation. If your results vary from your prospect, make clear why that may have happened. If your results agree, then explain the theory that the proof supported. It is never suitable to just state that the data approved with prospect, and let it drop at that.

- Make a decision if each premise is supported, discarded, or if you cannot make a conclusion with assurance. Do not just dismiss a study or part of a study as "uncertain."
- Research papers are not acknowledged if the work is imperfect. Draw what conclusions you can based upon the results that you have, and take care of the study as a finished work
- You may propose future guidelines, such as how the experiment might be personalized to accomplish a new idea.
- Give details all of your remarks as much as possible, focus on mechanisms.
- Make a decision if the tentative design sufficiently addressed the theory, and whether or not it was correctly restricted.
- Try to present substitute explanations if sensible alternatives be present.
- One research will not counter an overall question, so maintain the large picture in mind, where do you go next? The best studies unlock new avenues of study. What questions remain?
- Recommendations for detailed papers will offer supplementary suggestions.

Approach:

- When you refer to information, differentiate data generated by your own studies from available information
- Submit to work done by specific persons (including you) in past tense.
- Submit to generally acknowledged facts and main beliefs in present tense.

Administration Rules Listed Before Submitting Your Research Paper to Global Journals Inc. (US)

Please carefully note down following rules and regulation before submitting your Research Paper to Global Journals Inc. (US):

Segment Draft and Final Research Paper: You have to strictly follow the template of research paper. If it is not done your paper may get rejected.

- The **major constraint** is that you must independently make all content, tables, graphs, and facts that are offered in the paper. You must write each part of the paper wholly on your own. The Peer-reviewers need to identify your own perceptive of the concepts in your own terms. NEVER extract straight from any foundation, and never rephrase someone else's analysis.

- Do not give permission to anyone else to "PROOFREAD" your manuscript.

- Methods to avoid Plagiarism is applied by us on every paper, if found guilty, you will be blacklisted by all of our collaborated research groups, your institution will be informed for this and strict legal actions will be taken immediately.)
- To guard yourself and others from possible illegal use please do not permit anyone right to use to your paper and files.

Criterion for Grading a Research Paper (Compilation)
by Global Journals Inc. (US)

Please note that following table is only a Grading of "Paper Compilation" and not on "Performed/Stated Research" whose grading solely depends on Individual Assigned Peer Reviewer and Editorial Board Member. These can be available only on request and after decision of Paper. This report will be the property of Global Journals Inc. (US).

Topics	Grades		
	A-B	**C-D**	**E-F**
Abstract	Clear and concise with appropriate content, Correct format. 200 words or below	Unclear summary and no specific data, Incorrect form Above 200 words	No specific data with ambiguous information Above 250 words
Introduction	Containing all background details with clear goal and appropriate details, flow specification, no grammar and spelling mistake, well organized sentence and paragraph, reference cited	Unclear and confusing data, appropriate format, grammar and spelling errors with unorganized matter	Out of place depth and content, hazy format
Methods and Procedures	Clear and to the point with well arranged paragraph, precision and accuracy of facts and figures, well organized subheads	Difficult to comprehend with embarrassed text, too much explanation but completed	Incorrect and unorganized structure with hazy meaning
Result	Well organized, Clear and specific, Correct units with precision, correct data, well structuring of paragraph, no grammar and spelling mistake	Complete and embarrassed text, difficult to comprehend	Irregular format with wrong facts and figures
Discussion	Well organized, meaningful specification, sound conclusion, logical and concise explanation, highly structured paragraph reference cited	Wordy, unclear conclusion, spurious	Conclusion is not cited, unorganized, difficult to comprehend
References	Complete and correct format, well organized	Beside the point, Incomplete	Wrong format and structuring

INDEX

T

V

W

save our planet

Global Journal of Computer Science and Technology

Visit us on the Web at www.GlobalJournals.org | www.ComputerResearch.org
or email us at helpdesk@globaljournals.org

ISSN 9754350