

Performance Analysis of Various Data Mining Classification Techniques using Neural Networks

Ravindra.S¹, Dr. Shankaraiah², Anitha.J³, Shylaja L.K⁴

Associate Professor, Department of E&C, City Engineering College, Bangalore, Karnataka, India¹

Professor, Department of E&C, SJCE, Mysore, Karnataka, India²

Assistant Professor, Department of MCA, Dr.AIT, Bangalore, Karnataka, India³

Assistant Professor, Department of MCA, Dr.AIT, Bangalore, Karnataka, India³

ABSTRACT: The issue of data classification is maybe one of the most generally studied in data mining as well as machine learning groups. This issue has been examined by many researchers from different disciplines more than several years. Utilizations of data classification includes a wide variety of different problems domains such as social networks, text data, biological data and multimedia data. Moreover, the issue may be experienced in various different situations like streaming data or uncertain data. In supervised learning classification method, the data is classified into predefined class label. This is most commonly used technique in data mining to build classification model form many input data sets; these methods generally build classification models, which are utilized to predict the future data trends. Neural network systems are not suitable for many data mining application directly, because the classification are made is not explicitly designed as symbolic rules, which are suitable for verification by humans. This paper going to join different neural network methods such as CHARM algorithm and top K-Rules, which are generally utilized in data mining application to increase the data mining result. The outperformance of all the methods with neural networks is tested on existing database as well as results are compared by computational complexity of these algorithms.

KEYWORDS: Artificial Neural Network, CHARM algorithm, Data Mining, K-rule mining.

I. INTRODUCTION

The data mining means, the knowledge extraction from expansive databases. It is an interdisciplinary range of exploration in databases, pattern recognition, data recovery, machine learning, parallel, perception and circulated processing. There are various applications in information mining, for example, client relationship administration, business sector and industry characterization, pharmacology, pharmaceutical, stock administration and science. The data mining method has many tasks such as association rule data mining method, prediction as well as classification scheme and finally clustering techniques. Classification [1] is a best method and it is most power full scheme in data mining to develop classification models from an input dataset. The classification techniques generally develop models and there are used to predict the future values or data trends [2, 3].

The main aim of predictive scheme is to develop a design by utilizing the better result of known datasets and it is predicted by the results obtained from unknown datasets using constructed model. Classification of Data mining is a procedure of recognizable proof in that crude information are changed over to sorted significant data. It can be isolated into two gatherings: supervised and unsupervised. Numerous characterization strategies including decision tree, neural system (NN), and support vector machine (SVM) and other principle based frameworks have been proposed. Neural system characterization is administered down to earth approach with heaps of achievement in a few grouping assignments. Notwithstanding, its arrangement proficiency and exactness is for the most part an issue, which talks

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

about in this paper. It can be utilized to model complex connections in the middle of inputs and yields or to discover examples.

Neural systems signifies to a brain analogy for data handling and how the mind really works. ANN regularly just called a neural system. It is a numerical model or computational model in light of organic neural systems and natural neural framework for information investigation. Neural systems give an exceptionally general method for approaching nearer issues. The development of a Neural Network incorporates structure, encoding, and review variables. Numerous Neural system models have been effectively connected to the arrangement of complex issues identified with signal handling, grouping, clustering highlight choice, information representation, information mining, and data combination. They made databases for finding prescient data, finding hidden patterns that specialists might miss because it lies outside their desires [4].

One of the data mining issues is classification. Different algorithms have been intended to handle the issue by specialists in distinctive fields, for example, numerical programming, machine learning, and measurements. They are additionally worried with the successful utilization of the accessible database strategies, for example, proficient information recovery instruments. With such concerns, most algorithms proposed are fundamentally in view of decision trees. The point is to synergistically consolidate the techniques with improving the general execution of the neural system grouping algorithms. The neural systems don't give express learning representation as principles, huge specimen size, over fitting information, complicated relations between the data variables and the output. The rest of this paper is organized as follows. Section II provides background knowledge on different data mining using neural network algorithms. Section III gives as specific related work done by different authors, section IV present the proposed framework as well as implementation as well as flow chart. Section V provides simulation results with discussion and finally Section VI concludes this paper along with future research direction.

II. RELATED WORK

There are many different neural network algorithms are presented by different authors in their papers. In this paper three different artificial neural network algorithms are used with data mining algorithms to improve the computational complexity of the data mining algorithm following are the three different neural network techniques used in this paper.

A. Feed forward Neural Network:

A feed forward neural system (FFNN) is the one of the most straightforward neural system method, for example, in Fig 1. Comprises of three layers: an information layer, hidden and yield layer. In each layer there are one or all the more handling components. Processing elements is intended to reproduce the neurons in the cerebrum and this is the reason they are frequently produces neurons. A handling component gets inputs from either the past layer or the outside world. During training, this weight is balanced. Data just goes in the forward bearing through the system - there are no input feedback loops.

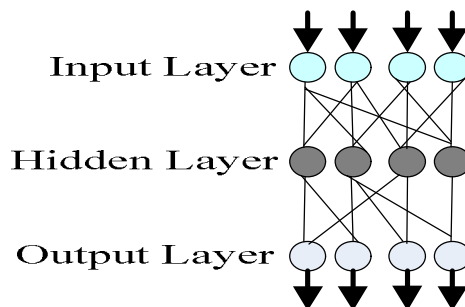


Fig 1. Feed forward NN flow.

The improved procedure for preparing a feed forward neural system is as per the following:

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

- Input information is introduced to the system and proliferated through the system until it achieves the yield layer of the system. This forward procedure creates a predicted yield.
- The Predicted yield is subtracted from the real yield the obtained error data for the system is calculated.
- The forward procedure begins again once back propagation has completed, and this cycle is proceeded until the error between the actual and predicted outputs is reduced.

B. The Back Propagation Algorithm:

Back propagation is a typical strategy for educating ANN systems how to perform a specific task. The back propagation scheme is utilized as a part of layered feedforward ANNs. This implies the artificial neurons are composed in layers, and send their signals "forward", and afterward the error are spread in reverse. The back propagation utilizes supervised learning, which implies that we give the calculation samples of the inputs and yields we need the system to process, and afterward the error is ascertained. The thought of the back propagation calculation is to decrease this error, until the ANN takes in the preparation information.

C. Radial basic function Network:

Fig 2. Shows the general idea of Radial basis function network is form the theory of approximation function. Here we have seen how the multi-layer perceptron network is hidden in the layer of sigmoidal units and it can learn to approximate the function. This network takes a different approach. The important feature are:

1. These are double -layer feed-forward systems.
2. Here, hidden nodes designed a set of radial basis functions.
3. Output is implemented using linear summation functions.
4. The training data is separated into two different stages: firstly, weights from input layer to hidden layer are determined, then secondly, weights from different hidden layer to output layer.
5. Here training of dataset is fast.
6. The networks are very good at interpolation.

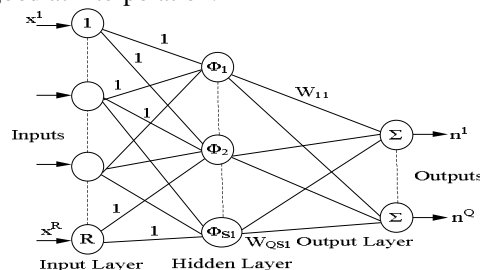


Fig 2. Radial Basis Function Network

III. PROPOSED SYSTEM WITH IMPLEMENTATION

Data mining is the procedure of removing legitimate, authentic and significant data from extensive databases is utilized to analyses the information in the scientific way database and group the information taking into account execution into normal, above normal and beneath normal classes. This paper proposed different data mining classification techniques using neuronal network algorithms. Here, the first techniques is association rule data mining classification techniques with neuronal network is discussed and secondly, the CHARM data mining classification method with neuronal network scheme is proposed.

A. Association rule for data mining with neural network:

The flowchart explains the proposed scheme, Association rule data mining with neural system appears in fig 3. In the first stage, the dataset is trained with known or prior information given to input layer. This digitized data is generated form association rule form the datasets.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

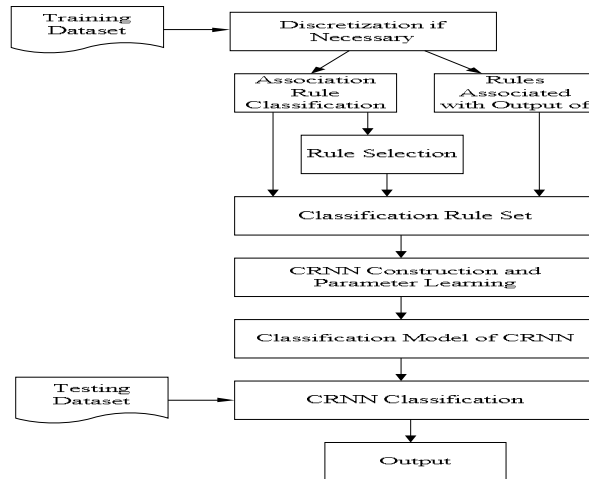


Fig 3. Association rule mining with neural network

The data classification set is obtained from discretization process. Classification rule NN (CRNN) development and parameter taking in set is then acquire form classification rule set. The yield of the CRNN grouping is the desired yield dataset.

B. CHARM data mining Algorithm with Neural Network:

CHARM is a proficient algorithm for identifying the arrangement of all regular close item-sets. There are various imaginative thoughts utilized in the improvement of CHARM; these include:

- CHARM all the while investigates both the item-set space and exchange space, over a novel IT-tree search space of the database.
- CHARM utilizes an exceedingly effective hybrid search technique that skips numerous levels of the IT-tree to rapidly distinguish the close item-sets, rather than enumerating numerous possible subsets.
- Diffsets definitely chop down the extent of memory required to store middle results.

The CHARM scheme with the neural network is shown in fig 4.

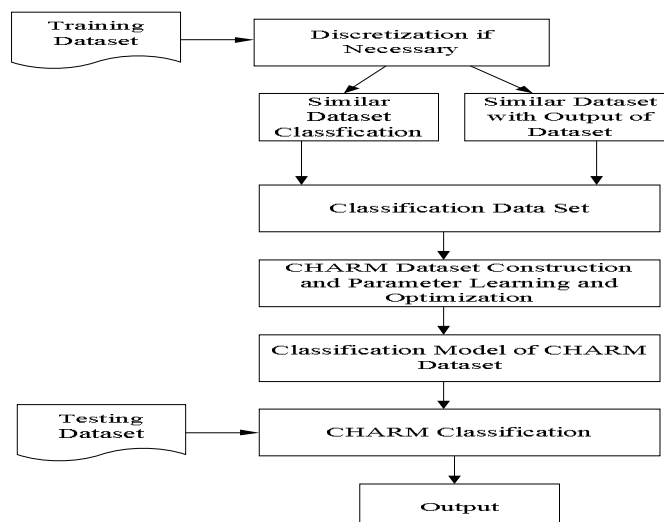


Fig 4. CHARM algorithms with neural network

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

IV. RELATED WORK

This section provides the specific work done by other authors on various data mining classification methods using neural networks. In this segment a brief review of literature relating to Data Mining as the subject of the examination is exhibited. It is an essential piece of learning and discovers applications in business, medication, science and building. It has been utilized to naturally find data in extensive databases.

Authors	Problem Focused	Techniques used	Performance parameters
S. Yasodha et al.[6]	Talent management problem	Support vector machine, Nearest neighbour algorithm.	Better accuracy, high performance.
Chakchai et al.[7]	For intrusion detection analysis	SVM, K-NN, NN, Decision Tree, Ripper Rule.	Cross-validation and confusion matrix.
G. M. Nasira et al.[8]	The complexity of vegetable price prediction.	Back-propagation NN.	Absolute error percentage, accuracy percentage.
A. Dewan et al.[9]	To determine and predict exact heart disease.	Back-propagation NN.	Reduce costs, correct treatment.
T. J. Peter et al.[10]	prediction of heart disease	An empirical study	Better predictor variables
R. Dash et al.[11]	To reduce large computation time.	Greedy polynomial neural network (GPNN), Polynomial Neural Network	Better processing time and its power.
F. Mili et al.[12]	Classification tasks of data mining	Functional link artificial neural network (FLANN).	Expansion function, best function.
A. S. Subaira et al.[13]	Intrusion detection, to reduce the strain.	Survey paper, SVM, GA, NN, FL, K-NN.	Highlighting an advantage & disadvantages of each techniques.
E. Kambal et al.[14]	To improve the process of assessing credit worthiness, credit scoring.	Decision Tree, ANN, GA and Principal Component Analysis (PCA).	highest accuracy, better scoring models
K. Anusha et al.[15]	Intrusion Detection, security issues in the MANET.	Intuitionistic fuzzy, RVM (Relevance Vector Machine) and NN.	Error rates, false positive & packet drop rates.
N.Hemageetha et al.[16]	Price prediction	Back-propagation neural network (BPNN) and Radial basis function neural network (RBF).	Fault tolerance. Efficient and accurate.
F. Thabtah et al.[17]	Decision making, the problem of Arabic text categorization.	C4.5, RIPPER, PART, and OneRule.	Error rate, Efficient and accurate.
A.J.Chamatkar et al.[18]	Data integrity, data classification.	CHARM Algorithm, Top K Rules mining and CM SPAM Algorithm	Accuracy, computational complexity.
P. Wlodarczak et al.[19]	Multimedia mining , achieve better resolution	Deep learning	Resolution, predicting the efficacy.
B. N. Lakshmi et al.[20]	To retrieve information effectively, enables data exploration, data analysis and prediction.	ANN, GA and decision trees.	Resolution, predicting the efficacy.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

V. RESULTS AND DISCUSSION

In this section, the parametric evaluation of the simulation of the proposed method is being evaluated and validated. In our research we have studied various algorithms, and found out the best ones in the field of data-mining. These algorithms. Great amount of work with respect to DM applications in different fields like medication, bioinformatics, meteorology and different fields are accessible. Data mining and Neural Network procedures have been utilized on information distribution centers to settle on more educated choices. Table.1 shows the existing works done by other author in data mining classification methods using neural networks are namely, CHARM Bit Set for Frequent Pattern Mining, Top K Rules for Association Rule Mining. The datasets is collected from online retailing e-commerce websites Flipkart and Amazon. Each dataset contains 1000, 2000, 5000 and 10000 entries, and is available for Frequent Pattern Mining, are namely, CHARM Bit Set for Frequent Pattern Mining, Top K Rules for Association Rule Mining. The datasets is collected from online retailing e-commerce websites Flipkart and Amazon. Each dataset contains 1000, 2000, 5000 and 10000 entries, and is available for Frequent Pattern Mining,

TABLE 2: TOP K RULES MINING (MINING TOP 10% RULES) WITH DIFFERENT NEURAL NETWORK TECHNIQUES

	Top Rules	K-BPP	CFF	RBF
Time 1000	20	33	7	13
Memory 1000	7.99	4.92	2.99	2.89
Computational complexity 1000	-----	30%	14%	14%
Time 2000	175	115	53	75
Memory 2000	15.50	7.50	4.60	2.98
Computational complexity 2000	-----	31%	13.5%	16%
Time 5000	132	5249	56	69
Memory 5000	10.58	6.5	5.0	2.6
Computational complexity 5000	-----	30.8%	12.8%	16.5%
Time 10000	235	120	64	85
Memory 10000	16.88	7.69	6.33	5.16
Computational complexity 10000	-----	33%	13%	15%

TABLE 3: CHARM MINING WITH DIFFERENT NEURAL NETWORK TECHNIQUES

	CHARM	BPP	CFF	RBF
Time 1000	41	12	10	18
Memory 1000	33.3	27.9	18.2	14.3
Computational complexity 1000	-----	30%	11%	14%
Time 2000	1	1	1	2
Memory 2000	22.3	12	7.05	5.8
Computational complexity 2000	-----	30%	12.9%	16.9%
Time 5000	7	3	1	3
Memory 5000	25	26	14	9.9
Computational complexity 5000	-----	28.9%	12.8%	16.5%
Time 10000	10	4	3	6
Memory 10000	56.8	27.6	24	18
Computational complexity 10000	-----	30.8%	12.8%	16.5%

As shown in table.2 and table.3, the computational complexity for the Top K Rule mining with Back Propagation is 32 %, with Radial Basis Function Networks is 16% and with Feed Forward is 13%. Computational complexity for CHARM Mining with Back Propagation is 30%, with Radial Basis Function Networks is 14% and with Feed Forward

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 6, Special Issue 12, July 2017

is 11%. So the computational complexity of the Feed Forward technique is less than the other technique of neural network for the data mining algorithms.

VI. CONCLUSION AND FUTURE DIRECTION

In this paper studies the Artificial Neural Network and how we can use ANN with the data mining concept. In this paper we are successfully able to integrate the Neural Network with the three data mining techniques. From the above result it's clear that time and memory complexity does not entirely depend on the length of the input dataset. For frequent pattern mining Radial Basis and Cascaded Feed-forward Networks are the best. Association Rule Mining follows the same trend of output as Frequent Pattern mining; Radial Basis and Cascaded Feed-forward Networks are the best. Sequential pattern mining is best done by Radial Basis Networks.

REFERENCES

- [1] Han, Jiawei, Micheline Kamber, and Jian Pei. Data mining: concepts and techniques: concepts and techniques. Elsevier, 2011.
- [2] Neelamadhab Padhy, Dr. Pragnyaban Mishra, Rasmita Panigrahi "The Survey of Data Mining Applications And Feature Scope" at International Journal of Computer Science, Engineering and Information Technology (IJCSSEIT), Vol.2, No.3, June 2012.
- [3] Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. "Application of data mining techniques in customer relationship management: A literature review and classification." Expert systems with applications 36, no. 2 (2009): 2592-2602.
- [4] Grimaldi, David, and Michael S. Engel. Evolution of the Insects. Cambridge University Press, 2005.
- [5] M. J. Zaki and K. Gouda. Fast vertical mining using Diffsets. Technical Report 01-1, Computer Science Dept., Rensselaer Polytechnic Institute, March 2001.
- [6] Yasodha, S.; Prakash, P.S., "Data mining classification technique for talent management using SVM," in Computing, Electronics and Electrical Technologies (ICCEET), 2012 International Conference on , vol., no., pp.959-963, 21-22 March 2012.
- [7] Chakchai So-In; Mongkonchai, N.; Aimtongkham, P.; Wijitsopon, K.; Rujirakul, K., "An evaluation of data mining classification models for network intrusion detection," in Digital Information and Communication Technology and its Applications (DICTAP), 2014 Fourth International Conference on , vol., no., pp.90-94, 6-8 May 2014.
- [8] Nasira, G.M.; Hemaageetha, N., "Vegetable price prediction using data mining classification technique," in Pattern Recognition, Informatics and Medical Engineering (PRIME), 2012 International Conference on , vol., no., pp.99-102, 21-23 March 2012.
- [9] Dewan, A.; Sharma, M., "Prediction of heart disease using a hybrid technique in data mining classification," in Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on , vol., no., pp.704-706, 11-13 March 2015.
- [10] Peter, T.J.; Somasundaram, K., "An empirical study on prediction of heart disease using classification data mining techniques," in Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on , vol., no., pp.514-518, 30-31 March 2012.
- [11] Dash, R.; Misra, B.B.; Dash, P.K.; Panda, G., "Greedy polynomial neural network for classification task in data mining," in Information and Communication Technologies (WICT), 2012 World Congress on , vol., no., pp.537-542, Oct. 30 2012-Nov. 2 2012.
- [12] Mili, F.; Hamdi, M., "A comparative study of expansion functions for evolutionary hybrid functional link artificial neural networks for data mining and classification," in Computer Applications Technology (ICCAT), 2013 International Conference on , vol., no., pp.1-8, 20-22 Jan. 2013.
- [13] Subaira, A.S.; Anitha, P., "Efficient classification mechanism for network intrusion detection system based on data mining techniques: A survey," in Intelligent Systems and Control (ISCO), 2014 IEEE 8th International Conference on , vol., no., pp.274-280, 10-11 Jan. 2014.
- [14] Kambal, E.; Osman, I.; Taha, M.; Mohammed, N.; Mohammed, S., "Credit scoring using data mining techniques with particular reference to Sudanese banks," in Computing, Electrical and Electronics Engineering (ICCEEE), 2013 International Conference on , vol., no., pp.378-383, 26-28 Aug. 2013.
- [15] Anusha, K.; Ezhilmaran, D., "Investigations on Classification Algorithms for Intrusion Detection System in MANETS," in Electronics, Communication and Computational Engineering (ICECCE), 2014 International Conference on , vol., no., pp.216-219, 17-18 Nov. 2014.
- [16] Hemaageetha, N.; Nasira, G.M., "Radial basis function model for vegetable price prediction," in Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on , vol., no., pp.424-428, 21-22 Feb. 2013
- [17] Thabtah, F.; Gharaibeh, O.; Abdeljaber, H., "Comparison of rule based classification techniques for the Arabic textual data," in Innovation in Information & Communication Technology (ISIICT), 2011 Fourth International Symposium on , vol., no., pp.105-111, Nov. 29 2011-Dec. 1 2011.
- [18] Chamatkar, A.J.; Butey, P.K., "Implementation of Different Data Mining Algorithms with Neural Network," in Computing Communication Control and Automation (ICCUBEA), 2015 International Conference on , vol., no., pp.374-378, 26-27 Feb. 2015.
- [19] Wlodarczak, P.; Soar, J.; Ally, M., "Multimedia data mining using deep learning," in Digital Information Processing and Communications (ICDIPC), 2015 Fifth International Conference on , vol., no., pp.190-196, 7-9 Oct. 2015
- [20] Lakshmi, B.N.; Raghunandhan, G.H., "A conceptual overview of data mining," in Innovations in Emerging Technology (NCOIET), 2011 National Conference on , vol., no., pp.27-32, 17-18 Feb. 2011.