Volume 8, No. 5, May-June 2017
International Journal of Advanced Research in Computer Science
RESEARCH PAPER
Available Online at www.ijarcs.info
ISSN No. 0976-5697

Radial Basis Function-Particle Swarme Optimization Algorithm for Feature Reduction (RBF-PSO)

Dr. M. Balamuragan
Associate professor,
School of computer science, Engineering and applications,
Bharathidasan University, Tiruchirappalli, Inida

R. Prasanka
Research Scholar,
School of computer science, Engineering and application
Bharathidasan University, Tiruchirappalli, India

Abstract: Feature Reduction is an important step in data mining technique which removes unwanted attributes. The feature selection technique is used in various research fields. In this paper a hybrid algorithm named Radial Basis Function-Particle Swarm Optimization (RBF-PSO) Algorithm is proposed by combining the two existing algorithms, Radial Basis Function and Particle Swarm Optimization Algorithm. The result obtained proves that the proposed algorithm work well for Feature Reduction when comparing with the existing algorithms.

Keywords: Data Mining; Feature Reduction; Particle Swarm Optimization; Radial Basis Function.

INTRODUCTION

The researchers in the term Data mining refers to “extracting” or “mining” the knowledge from large volume of data. Data mining which means discovery of extracting large hidden data, previously unknown patterns and relationships that are difficult to detect with traditional statistics. Mining is the core process involved in Knowledge Discovery Process. Knowledge Discovery is a process of getting high level knowledge from low level data. In data mining, the hefty data sets are a computational progression to ascertain patterns. Data mining techniques are the result of a long process of research and product development [1]. Data mining is a form of knowledge discovery essential for solving problems in a specific domain. Data mining is a step in the whole process of knowledge discovery which can be explained as a process of extracting or mining knowledge from large amounts of data [2]. The major challenge in the medical data mining is imprecision and uncertainty. Nowadays, offering cost worthy service is major problem faced by the health care organizations. The medical field has succeeded in identifying and predicting the disease with the aid of Data mining techniques.

Feature selection is the problem of selecting a subset of features without reducing the accuracy of representing the original set of features. Feature selection is used in many applications to remove irrelevant and redundant features where there are high dimensional datasets. These datasets may contain a high degree of irrelevant and redundant features that may decrease the performance of the classifiers. A Feature Selection algorithm explores the search space of different feature combinations to reduce the number of features and simultaneously optimize the classification performance. Feature selection is a multi-objective problem. Feature selection also known as Attribute subset selection, Variable selection , Feature subset selection is the process of selecting subset or essential features from the existing set of features. Feature selection in supervised learning has a main goal of finding a feature subset that produces higher classification accuracy.Novakovic, J., et al Feature selection can be able to define as a process that chooses a minimum subset of M features from the original set of N features therefore, the feature space is optimally reduced according to a certain evaluation criterion [3].Feature selection there are three categorization filter, wrapper and embedded As the dimensionality of an area extends the quantity of highlight N increments. Finding the best component subset is generally obstinate [4] and numerous issues identified with highlight determination have been appeared to be NP-hard [5]. Specialists have concentrated on different parts of highlight determination. Highlight determination calculations might be isolated into channels [6], wrappers [7] and installed approaches [8].

The remaining content of this paper is organized as follows. In Section II, brief literature on adaptive PSO and PSO-RBF hybridization are given. The proposed PSO hybrid algorithms are described in Section III. Section IV discusses the results before the concluding remarks in Section V

LITERATURE REVIEW

Y. Shi, R.C. Eberhart, V. Kadirkamanathan, K. Selvarajah, P.J. Fleming, et al Particle swarm optimization (PSO) is one of the most newly developed evolutionary algorithms, the researches on the mathematical foundations and the realistic applications of PSO have involved much attention [9-10]. Particle swarm optimization (PSO) is a swarm intelligence method developed in [11] stimulated by the common behavior of bird flocking and fish schooling.pso have been exposed to successfully optimize broad range of continues functions [12-13]. Particle swarm optimization as developed by the authors comprises a very simple method, and paradigms can be able to implement in a few lines of computer code. It requires only primitive mathematical operators, and is computationally inexpensive in terms of both memory requirements and speed. Early difficult have establish the implementation toward effective with several kinds of problems.

PSO has been applied to improve neural networks in various aspects, such as network connection weights, network architecture and learning algorithms. In recent years, there have been several papers reporting on the replacement of the back propagation algorithm by PSO for some neural network structures [14-15]. Particle Swarm Optimization is a population (swarm) Based optimization tool. Every single solution (called a particle) “flies” over the solution space in search for the optimal solution. The particles are evaluated using a fitness function to how close they are to the optimal solution [16].

RBF neural network's performance mainly depends on the center and the width of radical basis function of the hidden layer and the weight values of the output layer. Traditional RBF neural network learning strategy has significant drawbacks, which only finds the optimal solution in local space to determine the network structure parameters.Alejo et al.[17] RBF Neural Network is an extremely powerful type of feed forward Neural Network first perceived by Broomhead and Lowe. These feed-forward networks are trained using a supervised algorithm. They have been broadly used for classification and interpolation regression. The RBF Network is a three layer feed forward fully connected network, which uses RBFs as the only nonlinearity in the hidden layer neurons. The output layer has no nonlinearity and the connections of the output layer are only weighted, the connections from the input to the hidden layer are not weighted [18].Tsimboukakis and Tambouratz is et al. [19] RBF structure strictly comprises three layers. The first layer is the input layer that deals with the transmission of the inputs to the network. The second layer, known also as the hidden layer usually contains the activation function of the radial basis function. This activation function depends on the Euclidian distance from the input pattern vector to the center of hidden neuron. Finally, the third or the output layer usually contains a linear activation function. PSO-RBF finds the size of the network and the parameters that configure each neuron: center and width of its basis function. Supervised mean subtractive clustering algorithm has been proposed [20] to evolve RBF Networks and the evolved RBF acts as fitness evaluation function of PSO algorithm for feature selection. The method performs feature selection and RBF training simultaneously. PSO algorithm has been introduced [21] to train RBF Network related to automatic configuration of network architecture related to centers of RBF.

PROPOSED FRAME WORK

The data set collected from the UCI repository. This research paper proposes a Feature Reduction Framework (FRF) which contains dataset repository, data cleaner, feature reducer and hybridization manager. Dataset repository is used to store the different kind of datasets such as medical dataset, Wine dataset. Missing values are filled by data cleaner. Feature reducer is applied to reduce the unnecessary attributes from dataset. In order to obtain the optimal dataset, In reducer is applied to reduce the unnecessary attributes from dataset. Order to obtain the optimal dataset, the hybridization manager is used to hybrid RBF and PSO algorithm. This hybrid algorithm named Radial Basis Function and Particle Swarm Optimization (RBF-PSO) algorithm. The Figure 1 shows the Feature Reduction Framework for reducing the features in dataset.

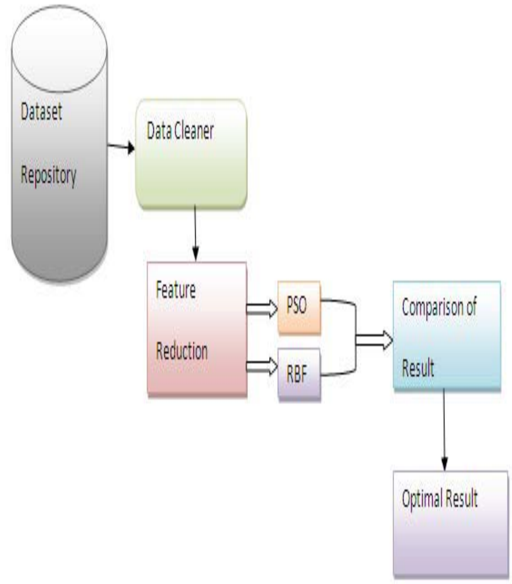


Figure1.Proposed Frame Work For Feature Reduction

Pseudo code RBF-PSO Algorithm

Step 1: Initialize particle weight & learning rate

Step 2: To step (3to12) until stopping condition is reached

Step 3: Calculate fitness value

Step4: Each input unit receives fitness value x_i & and send to the hidden unit (i=1 to n)

Step 5: Calculate the radial basis function

Step6: Send output signed for hidden unit to input of output layer units.

Step7: Select the centers for the radial basis function. The centers are selected from the set of input vectors. It should be noted that a sufficient number of centers have to be selected to ensure adequate sampling of the input vector space.

Step8: Calculate output from the hidden layer units Step 8: Calculate output from the hidden layer units

$$v_i(x_i) = \frac{\exp\left[-\sum_{j=1}^{r}(x_{ji} - \hat{x}_{ji})^2\right]}{\sigma_{i^2}}$$

where x_{ji} is the center of the RBF unit for input variables. σ_I is the width of the i^{th} RBF unit. X_{ji} is the j^{th} variable of input pattern

Step 9: Set new pbest

Step10: calculate the output of the Neural Network

$$y_{net} = \sum_{i=1}^{k} w_{im} v_i(x_i) + w_0$$

k-> number of hidden layer nodes

y_{net} -> output value of the m^{th} node in output layer for the n^{th} incoming pattern

w_{im}-> weight between i^{th} RBF unit and m^{th} output node.

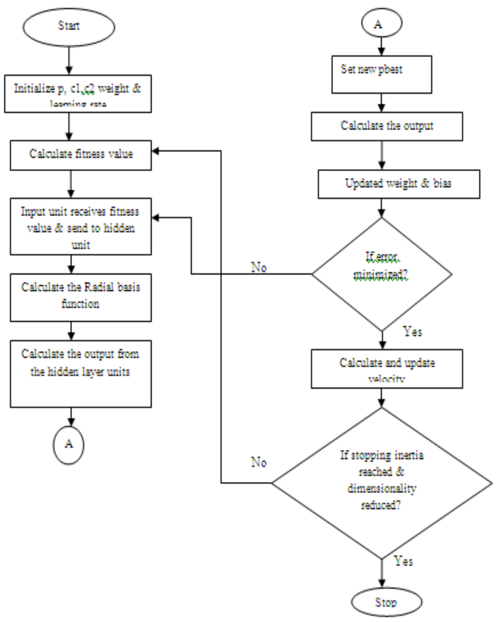
w_0-> biasing term at n^{th} output node

Step 11: calculate the error and test for stopping conditions.

Step 12: Calculate and update velocity

Step 13: Stop

Flow chart



IV.RESULT AND DISCUSSION

Dataset Name	Attributes	RBF	PSO	RBF-PSO
Diabetes	13	9	6	7
Lung -Cancer	15	7	9	6
Yeast	9	8	6	5
Hepatitis	10	7	8	6
Wine data	13	10	9	8
Liver data	14	11	8	9

Table 1.Comparision- Performance of Proposed Algorithm with the Existing System

Proper experimentation and comparison are carried to evaluate the performance of Radial Basis Function and Particle Swarm Optimization Algorithm. To evaluate the performance of RBF-PSO algorithm, it is compared with existing RBF, PSO techniques. The performance analysis of RBF-PSO algorithm is shown in figure 2. The RBF-PSO algorithm reduces the maximum features count in liver dataset compare with others datasets such as diabetes, lung cancer, yeast, hepatitis and wine data.

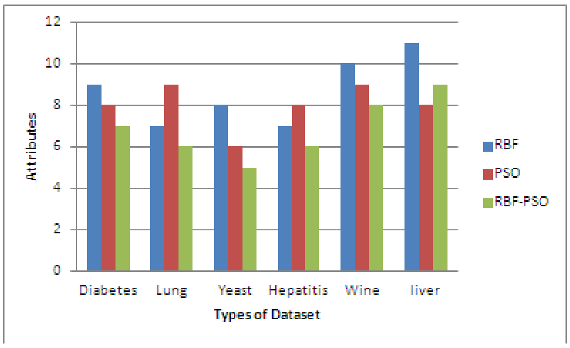


Figure 2.Comparision of the Proposed Algorithm with the Existing Terms In Respect to Feature Reduction

CONCLUSION

Feature reduction is a crucial and important step in Data Mining. The features should be reduced without affecting the knowledge extracted from the data set. This work combined two algorithms namely Radial Basis Function and Particle Swarm Optimization with different kind of datasets comprising the feature reduction technique and produce best result for the liver dataset. To improve the algorithm can be hydride with other techniques which is the feature scope of this work.

REFERENCES

[1] Bhagyashree Ambulkar and Vaishali Borkar “Data Mining in Cloud Computing”, MPGINMC, Recent Trends in Computing, ISSN 0975-8887, pp 23-26, 2012.

[2] J.Han and M. Kamber, Data mining concepts and techniques, Morgan Kaufmann, San Francisco 2006.

[3] Novakovic, J., 2010. The Impact of Feature Selection on the Accuracy of Naïve Bayes Classifier, 18th Telecommunications Forum TELFOR, page(s) 1114-1116, November 23-25, Serbia, Belgrade.

[4] Domingos, P., Pazzani, M., 1997. On the Optimality of the Simple Bayesian Classifier under Zero-One loss, Machine Learning, Volume29, page(s) 103-130, November/December 1997.

[5] Blum, A.L., Rivest, R.L., 1992. Training 3-node neural networks is NP-complete, Neural Networks, Volume 5, page(s) 117-127.

[6] Almuallim, H., Dietterich, T.G., 1991. “Learning with many irrelevant features”, in proceedings of AAAI-91, page(s) 547-552, Anaheim, California.

[7] Blum, A.L., Langley, P., 1997. Selection of Relevant Features and Examples in Machine Learning, Artificial Intelligence, Volume 97, page(s) 245-271.

[8] Y. Shi, R.C. Eberhart, “Empirical study of particle swarm optimization, Proceedings of Congress on Evolutionary computation”, IEEE Service Center, Piscataway, NJ, 1999.

[9] Kadirkamanathan, K. Selvarajah, P.J. Fleming, “Stability analysis of the particle dynamics in particle swarm optimizer”, IEEE Transaction on Evolutionary Computation 10 (3) (2006) 245–255.

[10] J. Kennedy, R. Eberhart, PSO optimization, in: Proceeding of IEEE International Conference Neural Networks, vol. IV, 1995, pp. 1941–1948.

[11] Leonard, J.A., Kramer,M.A.: “Radial basis function networks for classifying process faults”, Control Systems Magazine 11(3), 31–38 (1991).

[12] Liu, Y., Zheng, Q., Shi, Z., Chen, J.Training Radial Basis Function Networks with Particle Swarms. In: Yin, F.-L., Wang, J., Guo, C. (eds.) ISNN 2004. LNCS, vol. 3173, pp. 317–322. Springer, Heidelberg 2004.

[13] Chen, J., Qin, Z.: Training RBF neural networks with PSO and improved subtractive clustering algorithms. In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4233, pp. 1148–1155. Springer, Heidelberg 2006.

[14] Alejo, R., Garcia, V., Sotoca, J., Mollineda, R. and Sanchez, J. “Improving the Performance of the RBF Neural Networks Trained with Imbalanced Samples”. Computational and Ambient Intelligence, Vol. 4507, Springer, (2007), pp. 162-169.

[15] V.G. Gudise and G.K. Venayagamoorthy, “Comparison of Particle Swarm Optimization and Back propagation as Training Algorithms for Neural Networks”, Proceedings of the Swarm Intelligent Symposium, Indianapolis, IN, USA, April 24-26, 2003, pp 110-117.

[16] X. Hu, Y. Shi and R. Eberhart, “Recent Advances in Particle Swarm”, Proceedings of the Congress on Evolutionary

Computation", Portland, OR, USA, June 19-23, 2004, Vol. 1, pp 90-97.

[17] F. Van den Bergh and A.P. Engelbrecht, "Cooperative Learning in Neural Networks using Particle Swarm Optimizers", South African Computer Journal, Vol. 26, 2000, pp 84-90.

[18] Leonard, J.A., Kramer,M.A.: "Radial basis function networks for classifying process faults".Control Systems Magazine 11(3), 31–38 (1991)

[19] Tsimboukakis, N. and Tambouratzis, G. "Neural Networks for Author Attribution". Proceedings of the Fuzzy Systems Conference, FUZZ-IEEE 2007. IEEE, pp. 1-6.

[20] J., Qin, Z.: Training "RBF neural networks with PSO and improved subtractive clustering algorithms". In: King, I., Wang, J., Chan, L.-W., Wang, D. (eds.) ICONIP 2006. LNCS, vol. 4233, pp. 1148–1155. Springer, Heidelberg 2006.

[21] Liu, Y., Zheng, Q., Shi, Z., Chen, J.Training "Radial Basis Function Networks with Particle Swarms". In: Yin, F.-L., Wang, J., Guo, C. (eds.) ISNN 2004. LNCS, vol. 3173, pp. 317–322. Springer, Heidelberg 2004.

