Radial Basis Function (RBF) Neural Network Classification based on Consistency Evaluation Measure

Aye Mya Thandar, Myo Kay Khaing
University of Computer Studies, Yangon, Myanmar

ABSTRACT

Many researchers have been applied artificial neural networks in clinical diagnosis, image analysis, signal analysis, interpretation and various classification problems. Among artificial neural networks, RBF neural network has a single hidden layer and it is used to classify complex problems, whereas an MLP may have one or more hidden layers. Many feature selection methods have become important preprocessing steps to improve training performance and accuracy before classification. Consistency-based feature selection is an important category of feature selection research. This paper presents about RBF neural network classification based on consistency measure for medical datasets. There are irrelevant features in medical dataset and it becomes easier to train RBF network by removing unnecessary features. Therefore, this paper shows higher accuracy, better network performance and less time complexity by using RBF classifier based on consistency based feature selection.

Keywords

Radial Basis functions, consistency subsets, feature selection, Best first search

1. Introduction

There are several neuron models, network architectures and learning algorithm and this paper presents about RBF neural network. In a neural network, the hidden units form a set of "functions" that compose a random "basis" for the input patterns and these functions are called radial basis functions. RBF networks are known for requiring a much shorter training period, but they have the disadvantage of requiring good coverage of the input space by radial basis functions. If RBF neural network were trained by excessive redundant information, network training will be overloaded, training time will be increased, and even the network generalization ability will be reduced. The major drawback of RBF neural network is the large dimensionality which requires optimal feature subset. It is important to choose a small subset of features that is necessary and enough to describe the target concept. Therefore feature selection methods are necessary to reduce large dimensionality in training network before data calculation and classification.

Feature selection is the process of removing features from the data set that are irrelevant with respect to the task that is to be performed. Feature selection can be extremely useful in reducing the dimensionality of the data to be processed by the classifier [2]. There are many evaluation measures for feature selection. This paper describes RBF neural network based on consistency evaluator for selection to find the best subset of features that not only maximizes the classification accuracy but minimizes the number of features.

2. Related work

Consistency-based feature selection was proposed as a solution to this problem (eg. Zhao and Liu, 2007) and several consistency measures are proposed (6) (Liue et al., 1998; Pawlak, 1991; Arauzo-Azofra.et al., 2008; Shin and Xu, 2009) (10). In consistency-based feature selection, consistency measures are used to evaluate relevance of feature subsets. A consistency measure is intuitively defined as a metric to measure the distance of a feature subset from the consistent state.

Two researchers developed credit scoring based on Neural Networks and feature selection (4). The benefits of credit scoring involve reducing the credit analysis cost, enabling faster credit decisions, closer monitoring of existing accounts. But they used Multi Layer Perception (MLP) and it is more complex in term of weight and network training compared to RBF network.

Three researchers expressed that Radial Basis Function(FBF) Networks form a class of Artificial Neural Network(ANNs), which has certain advantages over other types of ANNs, such as better approximation capabilities, simpler network structures and faster learning algorithms(8).They used Particle Swarm Optimization to enhance RBF Network. Shifei Ding, Xinzheng Xu and Hong Zhu studied Artificial Neural Networks based on Genetic algorithm (9). They achieved a better performance and simplified the network topology by using genetic algorithm.

3. Feature Selection methods

Feature selection is often an essential data preprocessing step prior to applying a classification. The term feature selection is taken to refer to algorithms that output a subset of the input feature set. If information is irrelevant, the data is noisy and unreliable and input data is more difficult to train. Reducing dimensionality of data reduces the size of the hypothesis space and allows algorithm to operate faster and more effectively. The best effect of feature selection is to improve classification accuracy. Feature selection algorithms may be widely categorized into three groups: filter, wrapper method and embedded methods.

There are two important parts in feature selection: evaluation of a candidate feature subset and search through the feature space. There is a wide range of evaluation functions used in feature selection. Evaluation function may be deterministic or non-deterministic, and sometimes they are probabilistic estimates of a theoretical measure. They are distance measures, information gain attribute

measures, gain ratio measures, dependence measures, consistency measures, classifier error rate measures, chi - squared attribute measure and relief attribute measures. Among feature selection methods, consistency evaluation focuses on inconsistency measure according to which a feature subset is inconsistent. Search methods used in feature selection are best first, forward selection, race search, random search, genetic search, ranker, exhaustive search and rank search. We use best first search to find the best subset of measurement in consistency evaluation.

Best first search is an AI strategy that allows backtracking along the search path like greedy hill climbing, best first moves through the search space by making local changes to the current feature subset. However, unlike hill climbing, if the path being explored begins to look less promising, the best first search can backtrack to a more promising previous subset and continue the search from there. Given enough time, a best first search will explore the entire search space, so it is common to use a stopping criterion.
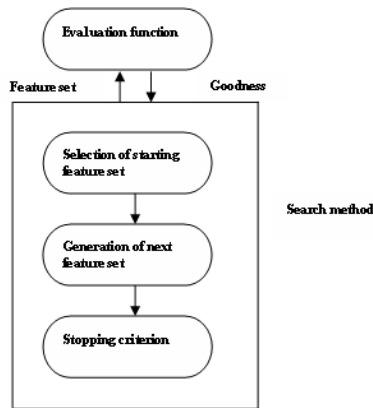

Figure 1.Feature selection process

4. The Proposed RBF network based on consistency subset evaluator

In RBF network, the position of the center can be selected from the training data set in a random way. RBF neural networks have one input layer, one hidden layer and one output layer. The hidden layer calculates the norm of the input from the neuron. It passes the norm through a non-linear activation function. The linear layer does the linear weighted addition of the outputs of the hidden neurons.

All features for medical datasets are evaluated by using the function of RBF neural network. Consistency measure is evaluated to satisfy the best subset of feature generation in order to find inconsistent datasets. After evaluation measure of features, best first search is used to optimize training data. Network training is complex and nonlinear to calculate error rate and result. Therefore, it becomes easier to train RBF neural network after reducing irregular features by using consistency based feature selection.
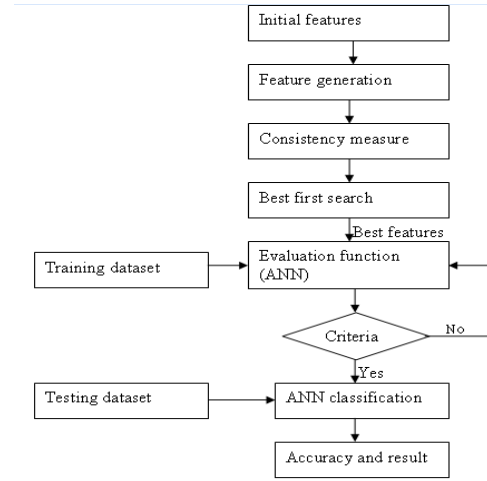

Figure 2. RBF neural network classifier based on consistency subset evaluator

5. RBF Neural Networks

The radial-basis function networks (RBF) address the problem of curve-fitting, which is approximation in high-dimensional spaces. Learning in this case is equivalent to finding an interpolating surface in the multidimensional space that provides a best fit to the training data, measured by preselected statistical criteria. There are three layers called input, hidden/kernel layer and output node. Each hidden unit represents a single radial basis function, with associated center position and width. Such hidden units are sometimes referred to as centroids or kernels. Each output unit performs a weighted summation of the hidden units, using the w_j s as weights. The training is typically done in two phases first fixing the width and centers and then the weights. The radial-basis functions technique suggests constructing of interpolation functions F of the following form:

$$F(\mathrm{x}) = \Sigma_{i=1}^{N} w_i \, \varphi(\, \| \mathbf{x} - \mathbf{x}_i \| \,)$$

Where: $\varphi(\, \| \mathbf{x} - \mathbf{x}_i \| \,)$ is a set of nonlinear radial-basis functions, $\mathbf{x}_i$ are the centers of these functions, and $\|.\|$ is the Euclidean norm.

The unknown weights can be found by solving the following linear matrix equation:

$$\mathbf{w} = (\, \mathbf{\Phi}^{T}\mathbf{\Phi} \,)^{-1} \, \mathbf{\Phi}^{T} \, \mathbf{y}$$

Where: $\mathbf{w}$ is the weight vector, $\mathbf{y}$ is the response vector, and $\mathbf{\Phi}$ is the design matrix

$$\mathbf{\Phi} = \{ \, \varphi_{ei} \mid \varphi_{ei} = \varphi(\, \| \mathbf{x}_e - \mathbf{x}_i \| \,), \, (\, e,i \,)=1,2,...,N \, \}$$
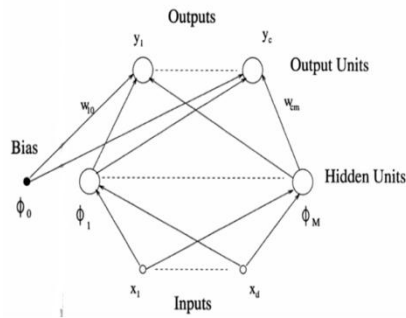
Figure3. RBF neural networks

6. Consistency measure

Consistency measure is also known as "filter" methods because of it's independence from any particular classifier that may use the selected features output by the feature selection method. In consistency-based feature selection, training instances are projected onto the subset of attributes and then the consistency of the subset is evaluated. It is therefore common practice to use consistency-based subset evaluator in conjunction with a search algorithm that looks for the smallest subset with consistency.

A pattern is considered inconsistent if there exists at least two instances such that they match all but their class labels; for example, an inconsistency is caused by instances (0 1, 1) and (0 1, 0) where the two features take the same values in the two instances while the class attribute varies which is the last value in the instance.

The inconsistency criterion is computed as follows: (1) suppose there are k possible class labels $label_1, label_2, \ldots, label_k$ in a certain dataset which contains N instances;(2) suppose there are J distinct combinations of attribute values for a subset s of attributes without considering the class labels of the instances; (3) suppose that D_i is the number of occurrences (or matching instances without considering the class labels of the instances) of the ith D_i instances, c_1 instances belong to class label $label_1$, c_2 instances belong to $label_2, \ldots$, and ck instances belong to class label $label_k$ such that $c_1+c_2+\ldots+c_k=D_i$ and let $M_i = max\{c_1, c_2, \ldots, c_k\}$. Then the inconsistency count is given by:

$$Inconsistency\ count = (D_i - M_i)$$

In other words, for matching instances without considering the class labels of the instances in a subset s of attributes, the more the class labels match, and the less inconsistency is the subset s with respect to the class. The inconsistency rate of an attribute subset s is given by the sum of all inconsistency counts divided by the total number of instances:

$$Inconsistency_s = \frac{\sum_{i=1}^{J} D_i - M_i}{N}$$

$$Consistency = 1 - Inconsistency$$

7. Medical Datasets

We use 4 medical datasets such as breast cancer, lymph, sick-thyroid and hepatitis from UCI repository.

7.1 Breast Cancer

The purpose of the breast cancer dataset is to classify a tumor as either benign or malignant based on cell descriptions gathered by microscopic examination. It contains 9 attributes and 699 examples of which 485 are benign examples and 241 are malignant examples.

7.2 Lymphography (lymph)

This is a small medical data set containing 148 instances with 18 nominal features. The task is to distinguish healthy patients from those with metastases or malignant lymphoma. The values for class attribute are normal find, metastases, malign lymph and fibrosis. This is the one of three medical domains (the others being Primary Tumor and Breast Cancer) provided by the University Medical Centre, Institute of Oncology, Lyubljana, Yugoslavia.

7.3 sick-thyroid

There are 30 attributes, two classes and 3163 instances in this dataset. We use this from UCI repository database.

7.4 Hepatitis

This data set is at UCI repository database. It contains 155 training dataset and 140 testing dataset. There are 20 attributes which are used as input neurons in this paper.

8. Experimental results of proposed system

We use four medical datasets to implement this proposed system which shows better experiments. After feature selection methods are evaluated, ranker search is used to find best feature subset except consistency method which use best first search. According to feature selection results, we can see that consistency method can reduce inconsistent features to find the best feature subsets by using the best first search.

We find that the results produced by RBF are not optimal before using consistency based feature selection. We can not only improve the accuracy of RBF neural network but also reduce time complexity by using consistency method.

Accuracy is estimated as the number of correct class predictions, divided by the total number of test samples. In this paper, RBF neural network classifier is used to classify the attributes selected by consistency based feature selection method. We have to know true positive, true negative, false positive and false negative to calculate accuracy. True positive answer denotes correct classifications of positive case (true positive-TP). True negative answer denotes correct classifications of positive case (true negative-TP). False positive answer denotes incorrect classifications of negative cases into class positive (False positive-FP). False negative answer denotes incorrect classifications of negative cases into class positive (False negative-FP).The classification accuracy measures the proportion of correctly classified cases:

Accuracy = (TP+TN) / (TP+TN+FP+FN)

Table1.Attributes comparison between feature selection methods after feature selection

Dataset	ReliefF	infoGain	gain ratio	chi-square	relief	consistency
Hepatitis	19	19	19	19	19	12
sick-thyroid	29	29	29	29	29	12
Breast Cancer	12	12	12	12	12	3
Lymphography (lymph)	16	16	16	16	16	7

Table2. Accuracy and time comparison between RBF and RBF-consistency

Medical datasets	hepatitis	sick-thyroid	lymph	breast cancer
RBF	80.65%	90.74%	76.81%	74.71%
RBF-consistency	84.10%	93.92%	79.44%	85.19%
RBF(time complexity)	0.47	0.17	0.05	0.03
RBF-consistency(time complexity)	0.03	0.14	0.02	0.02
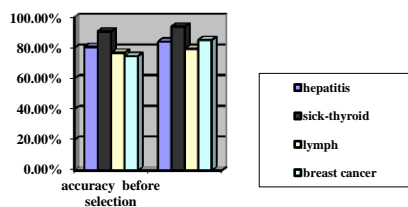
Figure4. Accuracy of RBF before and after feature selection
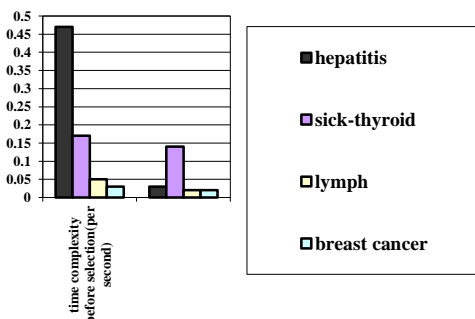
Figure5. Time complexity of RBF before and after feature selection

9. Conclusion and future works

We can remove unnecessary features and reduce network complexity.Therefore, it becomes easier to train RBF network more efficiently and easily in classification of medical datasets with best results. As there is lots of research about neural network methods, researchers are interested in using neural networks for many purposes. Therefore, there are many research works related with neural networks as future work. One of the future work is neural network can be improved by using many optimization measures, for example, genetic algorithm.

References

[1]. Antonio Arauzo-Azofra. Jose Manuel Benitez."Consistency measures for feature selection",J Intell Inf Syst,30:273-292, DOI 10.1007/s 10844-0037-0, (2008).

[2] C.N.Hsu, H. J. Huang and S. Dietrich "The ANNIGMA-Wrapper Approach to Fast Feature Selection for Neural Nets", IEEE Transactions on System, Man and Cybernetics, Part B, vol. 32, no. 2, pp.207-212, (2004).

[3] Chotirat Ann Ratanamahatana & Dimitrios Gunopulos "Scaling up the Naïve Bayesian Classifier: Using Decision Trees for Feature Selection", University of California, . (2004).

[4]. Dr.M.G.R,Dr.Jay B.Simha, "Evaluation of Feature Selection Methods for Predictive Modeling Using Neural Networks in Credits Scoring", Int.J.Advanced Networking and Applications, Volume:02,Issue:03,Pages:714-718(2010).

[5] Huan Liu, Hiroshi Motoda, Feature Selection: An Ever Evolving Frontier in Data Mining, JMLR: Workshop and Conference Proceedings 10: 4-13 The Fourth Workshop on Feature Selection in Data Mining ,(2010).

[6]. Liu et al., H.Liu,H, Motoda, and M.Dash. A monotonic measure for optimal feature selection. In Proceedings of European Conference on Machine Learning,(1998).

[7]. Pawlak.Rough Sets,Theoretical aspects of reasoning about data. Kluwer Academic Publishers.(1991).

[8] .Sultan Norman,Siti Mariyam Shamsuddin, and Aboul Ella Hassanien, "Hybrid Learning Enhancement of RBF Network with Partivle Swarm Optimization", Foundations of Comput. Intel. Vol.1,SCI 201,pp.381-397,Springer-Verlag Berlin Heidelberg ,(2009).

[9]. Shifei Ding,Xinzheng Xu,Hong Zhu,"Studies on Optimization Algorithms for some Artificial Neural Networks Based on Genetic Algorithm(GA)",Journal of computers,Vol.6,No. 5,May (2011).

[10].K.Shin and X.M.Xu. Consistency-based feature selection. In 13th International Conference on Knowledge-Based and Intelligent Information & Engineering System,(2009).

[11].Anderson, J.A.(2003) . An Introduction to neural networks. Prentice Hall, Chaeng, B. and Titterington, D. M. (1994). Neural networks: A review from a statistical perspective. Statistical Science, 9, 2-54.