Neural Networks 73 (2016) 70–76

Contents lists available at ScienceDirect

Neural Networks

journal homepage: www.elsevier.com/locate/neunet

One pass learning for generalized classifier neural network

Buse Melis Ozyildirim [a,*], Mutlu Avci [b]

[a] Department of Computer Engineering, University of Cukurova, Adana, Turkey
[b] Department of Biomedical Engineering, University of Cukurova, Adana, Turkey

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form 13 October 2015
Accepted 16 October 2015
Available online 26 October 2015

Keywords:
GCNN
Generalized classifier neural network
Classification neural networks
Smoothing parameter
Variance

ABSTRACT

Generalized classifier neural network introduced as a kind of radial basis function neural network, uses gradient descent based optimized smoothing parameter value to provide efficient classification. However, optimization consumes quite a long time and may cause a drawback. In this work, one pass learning for generalized classifier neural network is proposed to overcome this disadvantage. Proposed method utilizes standard deviation of each class to calculate corresponding smoothing parameter. Since different datasets may have different standard deviations and data distributions, proposed method tries to handle these differences by defining two functions for smoothing parameter calculation. Thresholding is applied to determine which function will be used. One of these functions is defined for datasets having different range of values. It provides balanced smoothing parameters for these datasets through logarithmic function and changing the operation range to lower boundary. On the other hand, the other function calculates smoothing parameter value for classes having standard deviation smaller than the threshold value. Proposed method is tested on 14 datasets and performance of one pass learning generalized classifier neural network is compared with that of probabilistic neural network, radial basis function neural network, extreme learning machines, and standard and logarithmic learning generalized classifier neural network in MATLAB environment. One pass learning generalized classifier neural network provides more than a thousand times faster classification than standard and logarithmic generalized classifier neural network. Due to its classification accuracy and speed, one pass generalized classifier neural network can be considered as an efficient alternative to probabilistic neural network. Test results show that proposed method overcomes computational drawback of generalized classifier neural network and may increase the classification performance.

1. Introduction

Radial basis function (RBF) is used as kernel function in neural networks and machine learning algorithms such as general regression neural network (GRNN), probabilistic neural network (PNN), generalized classifier neural network (GCNN) and support vector machines (SVM). Kernel parameters and stability affect performances of neural networks directly. While stability becomes more important for recurrent neural networks such as stochastic neural network, kernel parameters are important for feed forward neural networks (Rakkiyappan, Chandrasekar, Lakshmanan, & Park, 2014; Rakkiyappan, Chandrasekar, Lakshmanan, Park, & Jung, 2013; Rakkiyappan, Zhu, & Chandrasekar, 2014). RBF has only one parameter affecting the performances of algorithms. This parameter is known as smoothing parameter or variance. Hence, smoothing parameter optimization has been important research topic for increasing performances of machine learning algorithms. Since larger smoothing parameter value smoothes the estimation, it is assumed that all data are represented by a single huge datum. On the other hand, smaller smoothing parameter value takes into account data only closest to input datum. Data distribution should be known for estimating optimum value of smoothing parameter, otherwise, heuristic methods are required (Specht, 1991). In Specht (1991), holdout method is suggested for obtaining optimal smoothing parameter value. In this method, output value for each sample is calculated by using particular smoothing parameter, then mean squared error is measured. The value of smoothing parameter providing the best result is chosen for that dataset. In Nose-Filho, Lotufo, and Minussi (2011), different values of smoothing parameters are generated and the one providing the smallest mean absolute percentage error chosen as optimal smoothing parameter. Genetic algorithm

* Corresponding author.
E-mail addresses: melis.ozyildirim@gmail.com (B.M. Ozyildirim), mavci@cu.edu.tr (M. Avci).

http://dx.doi.org/10.1016/j.neunet.2015.10.008
0893-6080/

frequently used optimization method for determining suitable smoothing parameter value (Mao, Tan, & Set, 2000). In Masters and Land (1997), separate sigma values are assigned to each pattern neuron and utilized gradient descent for obtaining optimal values. If the smoothing parameter value is smaller than the average Euclidean distance between pattern and its neighbors, it fits data closely (Li & Bovik, 2011). Therefore, in Lee, Lim, Yuen, and Lo (2004), average of Euclidean distances among corresponding pattern and its K-nearest neighbors is used for initializing the smoothing parameter before using gradient descent approach to obtain final value. Another work based on the idea of average Euclidean distances, obtains unique smoothing parameter for all patterns by calculating the average Euclidean distances between pattern neurons belonging to the same class (Wang, Jiang, Hu, & Li, 2012). In Li, Zecchin, and Maier (2014), different smoothing parameter estimators are compared for the hydrological and water resources application. One of these methods named as Gaussian Reference Rule (GRR) is based on the idea of minimizing the asymptotic mean integrated squared error (AMISE) for Gaussian distributed data. In contrast to GRR, Biased Cross Validation (BCV) approach assumes that data are normally distributed. Two-stage direct plug-in (DPI) uses 2nd order integrated squared density derivative to obtain optimal smoothing parameter and minimizes the AMISE. It is reported that while increasing the number of stages provides better approximation, computational cost also increases. BCVDPI approach is the hybrid of BCV and DPI methods. In this method, BCV provides advantages of independence on the Gaussian distribution. On the other hand, smoothed cross validation (SCV) tries to minimize mean integrated squared error (MISE). These approaches are named as kernel density estimators. Unlike kernel density estimators, calibration optimization based methods are also frequently used for obtaining smoothing parameter. These methods are used for measuring the error with generated smoothing parameter(s) (Li et al., 2014).

GCNN is another RBF based neural network for classification. It uses regression based convergence, diverge effect term and gradient descent based smoothing parameter optimization method to avoid overfitting and stuck neuron problems. Assigning 0.9 or 0.1 target values to each datum in accordance with its class provides suitable structure for regression based convergence. GCNN uses diverge effect term in N neurons of summation layer, which is exponential form of difference between current and maximum target values. Diverge effect term increases the effect of target value and separates data belong to different classes. Smoothing parameter optimization increasing the efficiency is provided by gradient descent learning in GCNN, it consumes quite a long time. This can be considered as a drawback of GCNN (Ozyildirim & Avci, 2013). In Ozyildirim and Avci (2014), logistic regression based learning is applied to GCNN; introduced as logarithmic learning GCNN (LGCNN); to reduce training time. Logarithmic learning uses cost function instead of squared error and provides continuous update for smoothing parameter even if training parameter classified correctly. Therefore, faster convergence is obtained. However, efficiency of GCNN still depends on the initial value of smoothing parameter and training time may still be problem for real time applications.

In this work, one pass learning method is proposed for GCNN and named as one pass GCNN (OGCNN). It calculates a smoothing parameter for each class. OGCNN utilizes standard deviation and mean value based two functions to determine the value of smoothing parameter. Smoothing parameter value is calculated through one of the defined functions in accordance with standard deviation of features. It uses thresholding to determine the function to be executed. The main aim of thresholding is to fit data even if it has large range of values. OGCNN is tested on 14 different UCI machine learning repository datasets and classification performance of it is compared with that of standard GCNN, LGCNN, PNN, RBFNN, and Extreme Learning Machine (ELM) with RBF kernel (Bennett & Mangasarian, 1992; Frank & Asuncion, 2010; Mangasarian, Setiono, & Wolberg, 1990; Mangasarian & Wolberg, 1990; Turing Institute, 1986; Wolberg & Mangasarian, 1990; Yeh, Yang, & Ting, 2008; Zhou & Huang, 2012). Test results show that OGCNN improves the classification speed of standard GCNN and LGCNN maximally more than a thousand times. It also provides maximally 144 times faster classification performance than that of RBFNN. In addition, it provides faster classification than ELM. Moreover, OGCNN provides better classification performance in maximally 89% than that of PNN and RBFNN. ELM provides maximally 3.19% better classification performances than that of OGCNN for some datasets, while OGCNN provides maximally 11.33% better classification than that of ELM for the rest. OGCNN also improves classification performance of standard GCNN and LGCNN in the range of 2.5% and 80% for some datasets. However, OGCNN causes maximum 5.32% decreases in classification performances for three datasets in comparison to that of GCNN and LGCNN. While obtained decrease in training time is considered, these decreases in classification performance can be negligible. It can be seen that proposed method offers a solution to computational time drawback of GCNN and LGCNN.

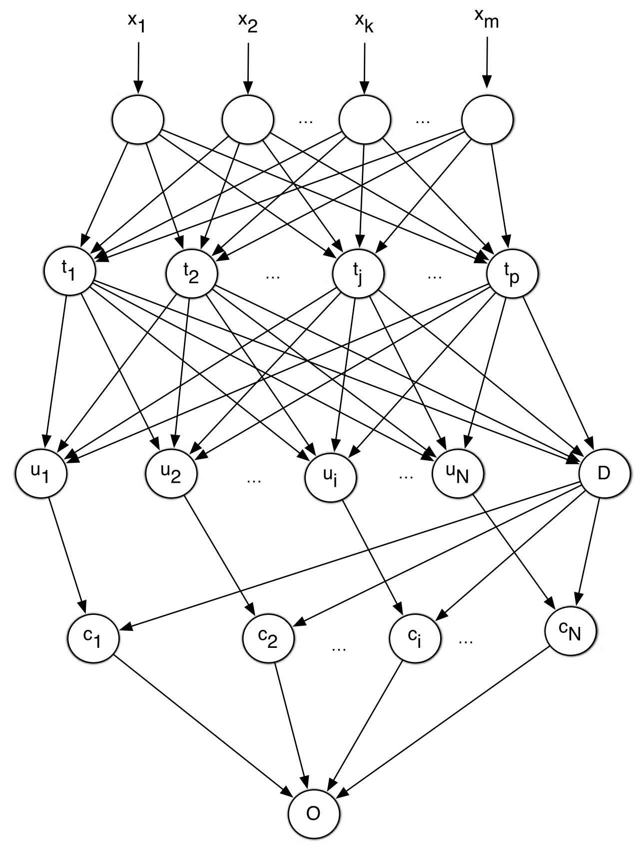

Fig. 1. GCNN architecture.

2. Generalized classifier neural network

GCNN is a five layered RBF based classifier neural network utilizing gradient descent approach and regression based classification. It optimizes smoothing parameter of RBF kernel through gradient descent approach. It has five layers named as input, pattern, summation, normalization and output as shown in Fig. 1.

Applied input vector x is transmitted to pattern layer through input layer. Pattern layer includes one neuron for each training