A New Definition of Sensitivity for RBFNN and Its Applications to Feature Reduction*

Xizhao Wang and Chunguo Li

Faculty of Mathematics and Computer Science, Hebei University, Baoding, Hebei, China
{wangxz,licg}@mail.hbu.edu.cn

Abstract. Due to the existence of redundant features, the Radial-Basis Function Neural Network (RBFNN) which is trained from a dataset is likely to be huge. Sensitivity analysis technique usually could help to reduce the features by deleting insensitive features. Considering the perturbation of network output as a random variable, this paper defines a new sensitivity formula which is the limit of variance of output perturbation with respect to the input perturbation going to zero. To simplify the sensitivity expression and computation, we prove that the exchange between limit and variance is valid. A formula for computing the new sensitivity of individual features is derived. Numerical simulations show that the new sensitivity definition can be used to remove irrelevant features effectively.

1 Introduction

As a sort of neural networks, the Radial-Basis Function Neural Network (RBFNN) is usually used to approximating a very complex and smooth function. In its basic form, the structure of RBFNN involves three layers, i.e., the input layer, the hidden layer and the output layer, with entirely different roles. The input layer accepts the information from the environment; the second layer (the only hidden layer) applies a nonlinear transformation to the accepted information; and the output layer supplies the response of the network. The input layer is made up of sensory units, the hidden layer nonlinear neurons, and the output layer pure linear neurons [1].

RBFNNs are able to approximate any smooth function within the required accuracy. According to RBFNN's training procedure, the hidden layer keeps adding neurons one by one until the required accuracy is reached. The network obtained is likely to be very huge. It is due to the high dimension of the hidden space when the training data have redundant information. If the redundant information could be deleted before training, the network size and performance would be improved. Sensitivity analysis of the network consequently arises and becomes one of the most important means for redundant information removal of neural networks [2–5].

The sensitivity analysis of neural networks has been investigated for over 30 years. During this period, a number of useful methodologies were put forward to investigate the sensitivity of Multilayer Perceptrons (MLP) [3–6]. One of the most popular techniques is to delete redundant inputs of MLPs using partial derivative [3, 4]. Another

* This research work is supported by NSFC (60473045) and Natural Science Foundation of Hebei Province (603137).

J. Wang, X. Liao, and Z. Yi (Eds.): ISNN 2005, LNCS 3496, pp. 81–86, 2005.

popular technique is to consider the weight perturbation [5, 6]. For example, Piché [6] investigated the effects of weight errors upon a statistical model of an ensemble of Madalines instead of the effects upon specific known networks.

Recently, the investigation of sensitivity analysis for RBFNNs has been found [7]. Sensitivity analysis plays an extremely important role in removing irrelevant features and improving the performance of RBFNNs. Due to the simplicity and the good approximation property of RBFNNs, the study on RBFNN's sensitivity has attracted more and more research fellows [8].

This paper aims at giving a new definition of sensitivity of RBFNNs. Considering the perturbation of network output as a random variable, this paper defines a new sensitivity formula which is the limit of variance of output perturbation with respect to the input perturbation going to zero. The exchange of limit and variance is proved here to be valid. A formula for computing the new sensitivity of individual features is derived. Numerical simulations show that the new sensitivity definition can be used to remove irrelevant features effectively.

2 A New Sensitivity Definition

Here we focus on a new definition of sensitivity: sensitivity based on variance of output perturbation. Statistically, the variance of one random variable measures the deviation from its center.

A trained RBFNN can be expressed as follows:

$$f(x_1, x_2, \cdots, x_n) = \sum_{j=1}^{m} w_j \left(\exp\left(\frac{\sum_{i=1}^{n} (x_i - u_{ij})^2}{-2 v_j^2} \right) \right) \tag{1}$$

where n is the number of features, m the number of centers, $(u_{1j}, u_{2j}, \cdots, u_{nj})$ the j-th center, v_j the spread of the j-th center, and w_j the weight of the output layer, $j = 1, 2, \cdots, m$.

Definition. The magnitude of sensitivity of the first feature is defined

$$S(x_1) = \underset{\Delta x \to 0}{Lim} \left(Var\left(\frac{f(x_1 + \Delta x, x_2, \cdots, x_n) - f(x_1, x_2, \cdots, x_n)}{\Delta x} \right) \right) \tag{2}$$

where $x_1, x_2, \cdots, x_n$ are n independent random variables with a joint distribution $\Phi(x_1, x_2, \cdots, x_n)$, $Var(y)$ means the variance of random variable y, and the input perturbation Δx is a real variable (not a random variable).

Similarly, the sensitivity definition for other features can be given.

If the limit works before the integral (i.e., *Var*) in Eq. (2), the definition can be simplified by using the partial derivative expression:

$$\frac{\partial f}{\partial x_1} = \underset{\Delta x \to 0}{Lim} \frac{f(x_1 + \Delta x, x_2, \cdots, x_n) - f(x_1, x_2, \cdots, x_n)}{\Delta x} \tag{3}$$

Mathematically, it is a very important Limit-Integral exchange problem.

Theorem. The limit operation and the variance operation in Eq.(2) can be exchanged, that is,

$$\underset{\Delta x\to 0}{Lim}\left(Var\left(\frac{f(x_1+\Delta x,x_2,\cdots,x_n)-f(x_1,x_2,\cdots,x_n)}{\Delta x}\right)\right)=Var\left(\frac{\partial f}{\partial x_1}\right) \tag{4}$$

where the function *f* is defined as Eq.(1) and $x=\left(x_1,x_2,\cdots,x_n\right)$ is a random vector with the density function $\varphi\left(x\right)$. The first and second order moments of the random vector x are supposed to be finite.

Proof. Consider a series of numbers $\Delta_k \to 0\,(k\to\infty)$ to replace the continuous $\Delta x\to 0$. Let

$$g_k=\frac{f(x_1+\Delta_k,x_2,\cdots,x_n)-f(x_1,x_2,\cdots,x_n)}{\Delta_k} \tag{5}$$

According to Eq.(1), the function g_k can be expressed as

$$g_k=\sum_{j=1}^{m} w_j\left(\exp\left(\frac{\sum_{i=1}^{n}(x_i-u_{ij})^2}{-2v_j^2}\right)\right)\left(\frac{1}{\Delta_k}\left(\exp\left(\frac{2\Delta_k(x_1-u_{1j})+\Delta_k^2}{-2v_j^2}\right)-1\right)\right) \tag{6}$$

We now first prove

$$\underset{k\to\infty}{Lim}\int g_k\varphi\left(x\right)dx=\int\left(\underset{k\to\infty}{Lim}\,g_k\right)\varphi\left(x\right)dx \tag{7}$$

Consider the last term in Eq.(6) and let

$$h_k=\left(\frac{1}{\Delta_k}\left(\exp\left(\frac{2\Delta_k(x_1-u_{1j})+\Delta_k^2}{-2v_j^2}\right)-1\right)\right), \tag{8}$$

We have

$$\begin{aligned}\lim_{k\to\infty} h_k &= \lim_{\Delta\to 0}\left(\frac{\dfrac{d}{d\Delta}\exp\left(\dfrac{2\Delta(x_1-u_{1j})+\Delta^2}{-2v_j^2}\right)}{\dfrac{d}{d\Delta}(\Delta)}\right)\\ &=\lim_{\Delta\to 0}\left(\exp\left(\frac{2\Delta(x_1-u_{1j})+\Delta^2}{-2v_j^2}\right)\cdot\frac{(x_1-u_{1j})+\Delta}{-v_j^2}\right)=\frac{(x_1-u_{1j})}{-v_j^2}\end{aligned} \tag{9}$$

It implies that $\left|h_k\right|\le\left|\dfrac{(x_1-u_{1j})}{-v_j^2}\right|+C$ holds well for all *k* where C is a constant. Therefore, from Eq.(6), we can obtain that

$$\begin{aligned}|g_k\varphi(x)| &= \left|\sum\nolimits_{j=1}^{m} w_j \left(\exp\left(\frac{\sum\nolimits_{i=1}^{n}(x_i-u_{ij})^2}{-2v_j^{\ 2}}\right)\right) h_k\varphi(x)\right| \\ &\le \sum\nolimits_{j=1}^{m}|w_j|\left(\left|\frac{(x_1-u_{1j})}{-v_j^{\ 2}}\right|+C\right)\varphi(x)\end{aligned} \tag{10}$$

Noting that the right of Eq.(10) is independent of k and the first order moments of x are supposed to exist, we have

$$\int\left(\sum\nolimits_{j=1}^{m}|w_j|\left(\left|\frac{(x_1-u_{1j})}{-v_j^{\ 2}}\right|+C\right)\varphi(x)\right)dx<\infty \tag{11}$$

which results in the correctness of Eq. (7), according to the integral convergence theorem.

Similar to the proof of Eq. (7), we can prove that

$$\underset{k\to\infty}{Lim}\int g_k^2\varphi(x)\,dx=\int\left(\underset{k\to\infty}{Lim}\,g_k^2\right)\varphi(x)\,dx \tag{12}$$

Further, paying attention to the following fundamental equalities on random variables,

$$Var(g_k)=E\left(g_k^{\ 2}\right)-\left(E(g_k)\right)^2=\int g_k^{\ 2}\varphi(x)\,dx-\left(\int g_k\varphi(x)\,dx\right)^2 \tag{13}$$

we obtain from Equations (7), (12) and (13) that

$$\begin{aligned}\underset{k\to\infty}{Lim}\,Var(g_k) &= \underset{k\to\infty}{Lim}\left(E\left(g_k^{\ 2}\right)-\left(E(g_k)\right)^2\right) \\ &= \underset{k\to\infty}{Lim}\int g_k^{\ 2}\varphi(x)\,dx-\left(\underset{k\to\infty}{Lim}\int g_k\varphi(x)\,dx\right)^2 \\ &= \int\left(\underset{k\to\infty}{Lim}\,g_k\right)^2\varphi(x)\,dx-\left(\int\left(\underset{k\to\infty}{Lim}\,g_k\right)\varphi(x)\,dx\right)^2 \\ &= E\left(\underset{k\to\infty}{Lim}\,g_k\right)^2-\left(E\left(\underset{k\to\infty}{Lim}\,g_k\right)\right)^2 \\ &= Var(\underset{k\to\infty}{Lim}\,g_k)\end{aligned} \tag{14}$$

Noting that

$$\underset{k\to\infty}{Lim}\,g_k=\frac{\partial f}{\partial x_1} \tag{15}$$

and $\{\Delta_k\}$ is an arbitrary sequence going to zero, we complete the proof of the theorem.

3 The Computational Formula of Sensitivity for RBFNNs

To numerically calculate the new sensitivity for each feature, the following equations are followed. Noting that

$$\frac{\partial f}{\partial x_1}=\sum_{j=1}^{m} w_j\left(\frac{\left(x_1-u_{1j}\right)}{-v_j}\exp\left(\frac{\sum_{i=1}^{n}\left(x_i-u_{ij}\right)^2}{-2v_j}\right)\right) \tag{16}$$

and Eq.(4), we have

$$Var\left(\frac{\partial f}{\partial x_1}\right)=Var\left(\sum_{j=1}^{m} w_j\left(\frac{\left(x_1-u_{1j}\right)}{-v_j^{2}}\exp\left(\frac{\sum_{i=1}^{n}\left(x_i-u_{ij}\right)^2}{-2v_j^{2}}\right)\right)\right) \tag{17}$$

Eq.(17) can also be expressed as

$$Var\left(\frac{\partial f}{\partial x_1}\right)=\sum_{j=1}^{m}\frac{w_j^2}{v_j^4}\left\{\begin{array}{l}\left(\dfrac{\sigma_1^2 v_j^3\left(2\sigma_1^2+v_j^2\right)+v_j^5\left(\mu_1-u_{1j}\right)^2}{\left(2\sigma_1^2+v_j^2\right)^{5/2}}\right)\exp\left(\dfrac{\sum\limits_{i=2}^{n}\left(\sigma_i^2\left(\sigma_i^2-v_j^2\right)+\left(\mu_i-u_{ij}\right)^2\left(2\sigma_i^2-v_j^2\right)\right)}{v_j^4}-\dfrac{2\left(\mu_1-u_{1j}\right)^2}{\left(2\sigma_1^2+v_j^2\right)^2}\right)\\ -\left(\dfrac{v_j^6\left(\mu_1-u_{1j}\right)^4}{\left(\sigma_1^2+v_j^2\right)^3}\right)\exp\left(\dfrac{\sum\limits_{i=2}^{n}\left(\sigma_i^2\left(\sigma_i^2-2v_j^2\right)+2\left(\mu_i-u_{ij}\right)^2\left(\sigma_i^2-v_j^2\right)\right)}{2v_j^4}-\dfrac{\left(\mu_1-u_{1j}\right)^2}{\left(\sigma_1^2+v_j^2\right)^2}\right)\end{array}\right\} \tag{18}$$

where $x_1,x_2,\cdots,x_n$ are supposed to be independent normal distributions with means $\mu_1,\mu_2,\cdots,\mu_n$ and variances $\sigma_1,\sigma_2,\cdots,\sigma_n$ respectively. Similarly, the computational formula for computing other features sensitivity can be derived.

For irrelevant feature deletion, we need the following formula

$$E\left(\frac{\partial f}{\partial x_1}\right)=\sum_{j=1}^{m}\frac{w_j}{-v_j^2}\left\{\left(\frac{v_j^3\left(\mu_1-u_{1j}\right)^2}{\left(\sigma_1^2+v_j^2\right)^{3/2}}\right)\exp\left(\begin{array}{l}\dfrac{\sum_{i=2}^{n}\left(\sigma_i^2\left(\sigma_i^2-2v_j^2\right)+2\left(\mu_i-u_{ij}\right)^2\left(\sigma_i^2-v_j^2\right)\right)}{4v_j^4}\\ -\dfrac{\left(\mu_1-u_{1j}\right)^2}{2\left(\sigma_1^2+v_j^2\right)^2}\end{array}\right)\right\} \tag{19}$$

It is the expectation of output's partial derivative with respect to individual features. The feature deletion policy takes the following steps. First, sort the features by sensitivity in descending order. Second, choose the features with the smallest sensitivity constituting a set A. Third, select the features with the smallest expectation from A to eliminate.

A number of simulations are conducted on some UCI databases. The simulation result shows that the training and testing accuracy can be kept if features with small sensitivity are deleted. Without losing any accuracy, Table 1 shows the results of deleting features according to our feature deletion policy. It is worth noting that, for the irrelevant feature deletion, we use both the variance (i.e., the sensitivity) and the expectation of the partial derivative. One question is whether or not the expectation is important for the irrelevant feature deletion? It remains to be studied further.

Table 1. The sensitivity simulations using UCI databases for input feature selection.

Database	Total Feature Number	Eliminated Feature Number
Sonar	60	12
Wine	13	3
Zoo	13	4
Ionosphere	34	5

4 Conclusions

A new sensitivity definition for RBFNNs is given and a formula for computing the new sensitivity is derived in this paper. The simulation result shows that the training and testing accuracy can be kept if features with stable and small expectation are deleted for most selected datasets.

References

1. Haykin, S.: Neural Networks: A Comprehensive Foundation (Second Edition). Prentice-Hall, Incorporation 2001
2. Winter, R.: Madaline Rule II: A New Method for Training Networks of Adalines. Ph. Thesis, Stanford University, Stanford, CA 1989
3. Zurada, J.M.: Perturbation Method for Deleting Redundant Inputs of Perceptron Networks. Neurocomputingc **14** (1997) 177-193
4. Choi, J.Y., Choi, C.: Sensitivity Analysis of Multilayer Perceptron with Differentiable Activation Functions. IEEE Transactions on Neural Networks, **3** (1992) 101-107
5. Zeng, X., Yeung, D.S.: Sensitivity Analysis of Multilayer Perceptron to Input and Weight Perturbations. IEEE Transactions on Neural Networks, **12** (2001) 1358-1366
6. Stephen, W., Piché.: The Selection of Weight Accuracies for Madalines. IEEE Transactions on Neural Networks, **6** (1995) 432-445
7. Karayiannis, N.B.: Reformulated Radial Basis Neural Networks Trained by Gradient Decent. IEEE Transactions on Neural Networks, **10(3)** (1999) 657-671
8. Karayiannis, N.B.: New Development in the Theory and Training of Reformulated Radial Basis Neural Networks. IEEE International Joint Conference on Neural Network, (2000) 614-619