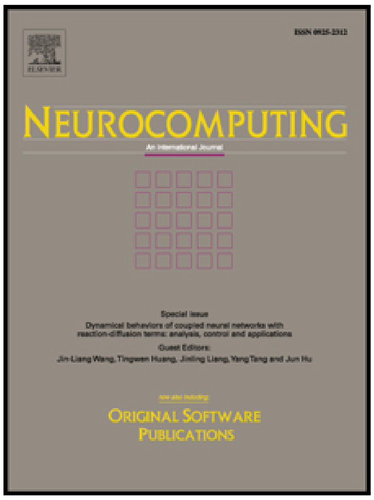
Accepted Manuscript



Wavelet Sampling and Generalization in Neural Networks

Zhiguo Zhang , Mark A. Kon

PII: S0925-2312(17)30809-3
DOI: 10.1016/j.neucom.2017.04.054
Reference: NEUCOM 18407

To appear in: *Neurocomputing*

Received date: 20 October 2016
Revised date: 14 March 2017
Accepted date: 27 April 2017

Please cite this article as: Zhiguo Zhang , Mark A. Kon , Wavelet Sampling and Generalization in Neural Networks, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2017.04.054

Wavelet Sampling and Generalization in Neural Networks

Zhiguo Zhang*
College of Automation, University of Electronic Science and Technology of China, Chengdu, China.
Email: zhiguozhang@uestc.edu.cn

Mark A. Kon
Department of Mathematics and Statistics, Boston University, Boston, MA, 02215, USA
Email: mkon@bu.edu

*Corresponding author:

Abstract: A new approach based on wavelet sampling is proposed to overcome overfitting in neural networks. Our approach optimizes input weights and network structure according to the empirical distribution of input training data. Thus only output weights are adjusted from training data errors. Using the fact that our algorithm trains input and output weights in independent procedures, our theorems demonstrate that it has rapid and global convergence. More importantly, we redefine a norm on l^2 space, corresponding to a useful new cost function. Using this cost function, the algorithm improves the ability of our networks to distinguish target functions from noise. In fact, we prove that this algorithm allows neural networks to act as wavelet filters, yielding good generalization, approximation and anti-noise capacities. Our simulations verify these theoretical results and simultaneously show the algorithm is robust to noise.

Keywords: Overfitting, Neural networks, Wavelet sampling theory

1. Introduction

In neural network learning, overfitting can be a serious problem. It is generally thought to be caused by model complexity and to be exacerbated by noise in training data. Model selection and early stopping are two effective ways to avoid overfitting, via limitation of network complexity [1, 11]. Using measures such as the Akaike [2] and Bayesian information criteria [3], model selection techniques apply statistical learning to select optimal numbers of neurons and their connectivity, improving generalization of networks greatly. On the other hand, early stopping is often used together with cross-validation. In these cases, samples are generally split into several sets for training, validating and testing a model based on cross-validation methods such as stratified *k*-fold [4, 5], exhaustive random [6], leave-*p*-out [7] and Monte Carto [8]. Training is stopped as soon as error on the validation set increases, insuring generalization.

Though model selection and early stopping (together with cross-validation) can help networks avoid overfitting by limiting complexity, the efficiency of such algorithms is often unsatisfactory. Additionally, through model selection, setting up networks with desired complexities is often computationally expensive [9, 10].

Based on the suggestion that overfitting is largely related to noise, filtering or suppression of noise is another obvious means to avoid overfitting. A number of methods (e.g. principal component analysis and reproducing kernels) have been used to avoid overfitting via noise removal [11, 12, 13]. However, such noise removal does not always lead to good generalization, and in some cases can even result in loss of details in estimation of target functions [9, 10].

Learning in networks also can be studied as a problem in optimal approximation [3, 14]. From this viewpoint, regularization is a powerful tool to overcome overfitting. Typically, a regularization term involving a differential operator is introduced into the cost function. With such terms, overly rapid decrease of training errors is mitigated and so overfitting is avoided [11, 15, 16].

However, such results in regularization training often depend strongly on choices of regularization terms [14]. Even for identical sets of training data, different regularization terms can lead to quite different training results. Hence, training results under regularization techniques might have the best-approximation property for a given cost function, but may not yield an optimal approximation of the target function for practical purposes.

Besides the abovementioned methods, there exist many other approaches to deal with the problem of overfitting. For example, generalization can be enhanced by improving robustness of neural networks with the aid of techniques known descriptively as dropout, autoencoder, adversarial training and adversarial committee [17-19]. Meanwhile, in [19-21], overfitting is avoided by adapting model dimensions, neural activation functions or neuron node properties via the use of soft activation function, compressed projection, pooling and pruning techniques. In addition, many optimization methods based on approaches involving evolutionary algorithms [22, 23], support vector machines [24], Kalman filters [25] and other approaches can be effective with selection of training data and network structures to ensure generalization. These methods are based on a number of heterogeneous viewpoints, but all of them can be related to work concerned with model optimization, noise removal or data density and have similar shortcomings in computational expense, detail loss and approximation