Минно-геоложки университет "Св. Иван Рилски"
Годишник, том 47, свитък I, Геология и геофизика, София, 2004, стр. 283-286

ON THE PARAMETERS SELECTION OF THE NEURAL NETWORK CLASSIFIER FOR REMOTELY SENSED MULTICHANNEL DATA

H. Nikolov [1], N. Nikolova-Jeliazkova [2]
[1] Solar-Terrestrial Influences Laboratory, BAS
[2] Institute of Parallel Processing, BAS

ABSTRACT. Neural networks have been used as a good general learning tool in data processing in solving large number of problems in many areas of scientific research. Since the neural networks falls in the category of supervised methods of classification the importance of the parameters of the network and training constraints are recognized as one of the key factors that affect the considerably on network performance. Different approaches could be adopted for tuning the parameters of the network, but there is not a rule of the thumb which is always valid. In this paper we propose some general rules, extracted from our experience classifying multispectral data from remote sensing experiments with neural networks, defining the role and the nature of each parameter. This rules target two basic topics in the methodology – first minimize the time needed for training the network (initial weights, momentum etc.), which allows the user to experiment with different structures. The second one is to find the most convenient structure of the network (number of layers, hidden neurons etc.) which suits the data used. We propose new characteristic features of the studied objects to be derived from the multispectral data, resulting in increase of the accuracy.

ВЪРХУ ПОДБОРА НА ПАРАМЕТРИ НА НЕВРОННА МРЕЖА ИЗПЪЛНЯВАЩА РОЛЯТА НА КЛАСИФИКАЦИОННА ПРОЦЕДУРА ЗА ДАННИ ПОЛУЧЕНИ ПРИ ДИСТАНЦИОННИ ИЗСЛЕДВАНИЯ

Х. Николов [1], Н. Николова-Желязкова [2]
[1] Централна лаборатория по слънчево земни въздействия, БАН
[2] ИПОИ, БАН

РЕЗЮМЕ. Невронни мрежи се използват често за решаване на широк кръг задачи в научни и научно-приложни изследвания, при които е необходимо класификация на обекти с предварително обучение. Последното условие налага внимателен подбор на параметрите на невронната мрежа (модела) и ограниченията при обучение на модела, тъй като от тях зависи в голяма степен и крайният резултат. За достигане на окончателните параметри на модела (брой слоеве, тегловни коефициенти, топология) и скростта при обучението му се използват различни стратегии, но общоприето и общовалидно мнение сред научните среди не се е налажило. В тази статия предлагаме един успешен подход, установил ролята на споменатите параметри, който се базира на опита на авторите при обработка и класификация на многоспектрални данни от дистанционни изследвания на подстилащата повърхност на Земята. Две са основните цели залагнали при разработването на предлагания подход – подбор на оптимална топология на модела и минимално време за обучението му. Предложени са за използване нови характеризиращи признаци на изследваните обекти, базирани на многоспектрални данни, повишаващи точността при класификация.

Introduction

The ultimate goal of using remotely sensed data (spectrometric, radar) is to extract information from them about the properties of the substance of the earth's surface and of the atmosphere together with their geographical relationships. As final result of processing the data a thematic map of the target region is produced.

During the past 20 years, statistical classification methods, such as the minimum distance and the maximum likelihood classifiers, have been widely used. Generally two approaches are used at this point – unsupervised and supervised classification. However, statistical methods have their restrictions, related particularly to the distribution assumptions and limitations in the input data types. In the past decade, the artificial neural network approach, theoretically a more sophisticated and robust method of image classification has been introduced and employed in remote sensing applications. Although this approach has been used in a wide range of scientific disciplines for a variety of applications since the early 1980s, their use in remote sensing area is relatively new, dating only from the early 1990s. Studies have shown that artificial neural networks (ANNs) are more robust than conventional statistical methods in terms of producing classification results with higher accuracies and requiring fewer training samples. The most important characteristic of ANNs is perhaps their non-parametric nature, assuming no *a priori* knowledge, particularly of the frequency distribution of the data. Because of their adaptability and their ability to produce high-quality results, the use of artificial neural networks has spread in the scientific community at large, leading to an increasing amount of research in the remote sensing field.

Researchers have applied neural network classifiers to remotely sensed data for several different purposes. For example; Benediktsson *et al*. (1990), Kanellopoulos *et al*. (1992), Paola and Schowengerdt (1995a), compared the results of maximum likelihood classification, which is the most elaborate statistical method of image classification, with artificial neural network classifiers, and found that ANNs can produce more accurate results than a maximum likelihood

classifier. Articles by Paola and Schowengerdt (1995b) and Atkinson and Tatnall (1997) review the use of artificial neural networks for remote sensing data.

Method and data

Artificial neural networks are heuristic algorithms, in that they can learn from experience via samples and can subsequently be applied to recognize new data. ANN has been proved to be more robust compared to conventional statistical classifiers in recognizing patterns from noisy and complex data and in estimating their nonlinear relationships.

ANNs have been found to be effective in identifying patterns and other underlying data structures in multidimensional data, such as the remotely sensed data. They have some unique advantages, such as their non-parametric nature, arbitrary decision boundary capabilities, tolerant to noisy data, and ability to generalize from training data. In addition, unlike traditional statistical methods, such as the maximum likelihood classifier, ANNs permit the use of a range of data types, including categorical data. It has also been reported that artificial neural networks can classify small training datasets better than conventional statistical classifiers.

The main idea behind ANNs is to create a processing system that behaves as close as possible to the human brain. The structure that was widely adopted, and will be consider this point forward, is a feedforward network of equal processing elements arranged in layers. This structure is known as multilayer perceptron, which as four-layer model is shown on fig.1.

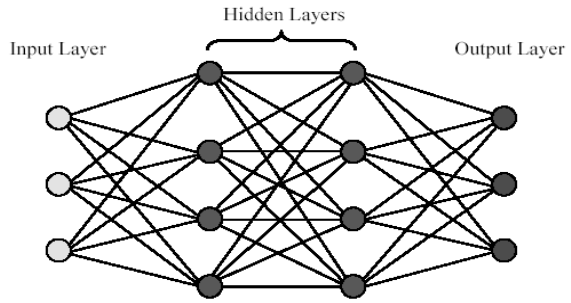


Fig. 1. Four layered model with three inputs

As activation function for the nodes of the network any differentiable nonlinear function could be used, but two are widely used – sigmoid and 'tanh'. The main difference in the use of these activation functions is that whilst for sigmoid input data and output classes are coded in the [0 1] range, for the 'tanh' function they are given in a [-1 1] range. As can be noticed, the 'tanh' function represents the data in a broader range, which may have positive effect in the performance of the network.

In this study a suggestion is made that final classification results of the ANN, is not be in a form of ordered binary number (for example 000100), but adding some "fuzziness" giving representation as fractional number from interval (0;1), starting from the uppermost output node. The one corresponding to the class with highest probability and zeroes are for the rest. In our case we discriminate between 8 major classes that are to found in the area of study.

Although artificial neural network classification methods are more robust than conventional statistical approaches, they suffer from a number of drawbacks, related in particular to the long training time requirement, determining the most efficient network structure for a particular problem, and inconsistent results due to the use of random initial inter-node weights. Most importantly, the structure of the network has a direct effect on training time and classification accuracy. There are also problems stemming from the nature of steepest-descent based learning algorithms. In the next sections of the paper we will provide guidelines how to overcome these disadvantages.

In this study the multispectral data are taken from ETM+ instrument flying onboard remote sensing satellite Landsat 7. They cover partially of Central North region of Bulgaria. In our experiments we used spectral 6 bands and the NDVI (K. Campbell 2002). The training samples were obtained from shape files consulting modern topographic map of the same area at scale 1:25000. This manner input-output pairs for each class for training the ANN was constructed.

The training algorithm is the core of an artificial neural network (ANN) application as it is necessary to make the network nodes and weights capable of performing a useful task by learn the internal structure of the data. There are many learning strategies developed for different neural network models, however, for training feed-forward neural networks the most popular technique is the backpropagation algorithm introduced by Rumelhart *et al*. (1986). According to Werbos (1995), it has been used in about 70% of ANN applications. He defines backpropagation as a procedure for efficiently calculating the derivatives of some output quantity of a nonlinear system, with respect to all inputs and parameters of that system, through calculations proceeding backwards from outputs to inputs. The main aim of the training process is to delineate the decision boundaries in the feature space thus defining classes of interest.

In our experiments the backpropagation algorithm, also known as generalised delta rule, which is an iterative gradient descent training procedure, was used. Principally it is carried out in two stages. In the first stage, after all the network weights have been randomly initialised, the input data are presented to the network and propagated forward to estimate the output value for each pattern set. In the second stage, the difference (error) between known and estimated output is fed backward through the network and the weights are changed in such a way that the difference is minimized. The whole process is repeated with weights being recalculated at every iteration until the error is minimal, or else lower than a given threshold value.

Results

In general, the application of ANNs requires some important decisions to be taken by the user, specifically a remote sensing researcher, which seriously affect the accuracy of the resulting classification. In terms of the factors involved, these decisions may be divided into two main groups: external factors and internal factors. External factors include the image resolution (spatial and radiometric resolution) and sample choice. However, internal factors are the choices of an appropriate network size (structure), initial weights, number of iterations, transfer function, and learning rate. While internal factors result from the limitations of the MLP and the backpropagation learning algorithm, external factors given here are specifically

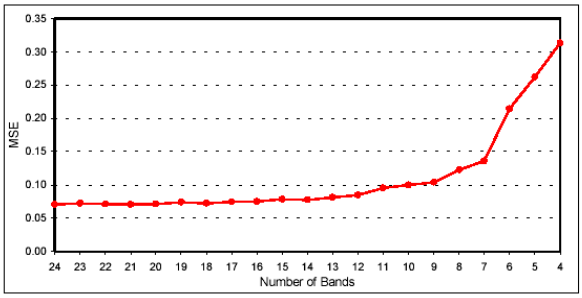
caused by the issues associated only with remote sensing related studies. These parameters need to be understood and adequately resolved in order to produce good results using ANNs.

The effectiveness of ANNs was evaluated in this order:

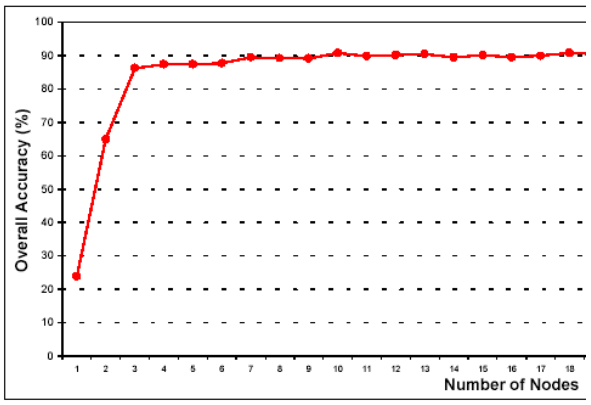
- Number of Input Nodes
- Number of Hidden Nodes
- Learning Rate and Momentum
- Initial Weight Range
- Number of Training Samples
- Output Encoding

Brief discussion about the influence of every one of the parameters follows.

Number of Input Nodes - generally corresponds to the number of independent variables (S. Warren 1994). In this study the number of the input features was selected to be equal to 7 (see before). This number was considered by the authors to be sufficient based on literature survey (??) and own experiments (Cospar).



Number of Hidden Nodes – they have direct impact on the determination of the decision boundaries in the feature space. The authors agree with Hand (1997), 'a network with two hidden layers allows convex regions to be combined, producing no convex, even disconnected regions'. In practice this allows better separation even the classes doesn't form only one hypersphere. The question is not only to find the optimum number of hidden layer nodes but also to determine the optimum number of hidden layers. Our suggestion is to have two hidden layers with 15 nodes on the first one and 9 on the second. This allowed us to separate classes as shown on fig.3.



Learning Rate and Momentum – both parameters try to avoid the main disadvantage of the backpropagation learning algorithm namely its slow convergence. The learning rate is related with how large the step toward the minima of the error surface as function of weights. Here a trade-off between large training time (small rate) and unstable, oscillating system (large rate) should be found. Our experiments (diploma) proved that suitable values for it lie in the interval [0.1-0.2].

Initial Weight Range – setting the initial weights of the internode links to a random values, a starting location on the multidimensional error surface is defined, which means different starting point. Setting them to zeroes means that no learning will take place and high values (above 1) results in slow learning. Although a lot of research have been carried out in order to examine the effect of different initial weight configurations, to date there is no method universally accepted for the determination of an optimum range. We got accuracy above 65 % with initial values for the weights from the interval [-0.1;+0.1]. (diploma).

Number of Training Samples - The number of training samples employed at the learning stage has a significant impact on the performance of any classifier. This issue is perhaps more important for neural networks than for conventional statistical classifiers since their performance is totally dependent from the training data presented. Because of this fact, neural network models are sometimes called datadependent methods. Higher accuracy requires the use of more features (i.e. spectral bands or their combinations) requires more training samples. This relationship can be thought of as a linear one. One good rule is proposed by Klimasauskas (1993) suggesting using five training samples for each weight in the network can be applied to determine the number of training samples needed.

Output Encoding – it must be considered before the training process starts since the output encoding the real world features or classes are represented in the network in a special way. Output encoding techniques suggested by researchers have not been fully investigated and compared in terms of the classification accuracy produced. The conventional way of representing the classes in neural networks is to allocate one output node for each land cover type, and assign 1 to the node that corresponds to a particular class and 0 to the nodes that represent other classes. For example, if there are four classes to be classified, a code of 0 1 0 0 is used to represent the class 2 in the network. The advantage of this approach is that the values of the output nodes calculated for test data can be interpreted as *a posterior* probabilities of membership since they are in the 0-1 range. It is argued that the use of such encoding for outputs may be problematic as the values of 0 and 1 can only be produced by the network when the weights have infinitively large positive and negative values. The standard sigmoid activation function can only yield these extreme values for the inputs of $\pm\infty$.

Conclusions

All suggestions derived from this study could be summarized as follows: feedforward network trained by backpropagation with number of inputs equal to features used, two hidden layers with initial weights ranging [-0.1;+0.1] and outputs coded as scaled fractional number.

References

Benediktsson, J. A., Swain, P. H. and Ersoy, O. K., 1990 Neural network approaches versus statistical methods in

classification of multisource remote sensing data. IEEE Transactions on Geoscience and Remote Sensing, 28, 540-551.

D. J Hand, 1997 Construction and Assessment of Classification Rules, New York: John Wiley & Sons.

K. Campbell 2002, "Remote Sensing," Encyclopedia of Environmetrics, (A. H. El-Shaarawi and W. W. Piegorsch, eds.), Volume 3, John Wiley and Sons Ltd, West Sussex England, pp.1744-1755.

Kanellopoulos, I., Varfis, A., Wilkinson, G. G. and Mégier, J., 1992 Land-cover discrimination in SPOT HRV imagery using an artificial neural network – a 20-class experiment. International Journal of Remote Sensing, 13, 917-924.

Paola,J. D. and Schowengerdt, R. A. (1995a) A detailed comparison of backpropagation neural network and maximum-likelihood classifiers for urban land use classification. IEEE Transactions on Geoscience and Remote Sensing, 33, 981-996.

Paola, J. D. and Schowengerdt, R. A. (1995b) A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. International Journal of Remote Sensing, 16, 3033-3058.

S. Warren, 1994, Neural Networks and Statistical Models, Proceedings of the Nineteenth Annual SAS Users Group International Conference, April, SAS Institute Inc., Cary, NC, USA.

Klimasauskas, C. C. (1993) Applying neural networks. In Neural Networks in Finance and Investing, edited by R. R. Trippi and E. Turban. (Cambridge: Probus), 47-72.

Препоръчана за публикуване от
катедра "Приложна геофизика", ГПФ