Multilayer Perceptron for analyzing satellite data

Kamal Al-Rawi, and Raed Shadfan

Department of Computer Science, Faculty of Information Technology, Petra University, P.O. Box 961343, Amman 11196, Jordan.
e-mail: kamalr@uop.edu.jo
e-mail: k_alrawi@yahoo.com

Abstract

Different ANN architectures of MLP have been trained by BP and used to analyze Landsat TM images. Two different approaches have been applied for training: an ordinary approach (for one hidden layer $M-H_1-L$ & two hidden layers $M-H_1-H_2-L$) and one-against-all strategy (for one hidden layer $(M-H_1-1)xL$, & two hidden layers $(M-H_1-H_2-1)xL$). Classification accuracy up to 90% has been achieved using one-against-all strategy with two hidden layers architecture. The performance of one-against-all approach is slightly better than the ordinary approach.

Key words
MLP,
Landsat,
Classification

Article info
Received: Aug. 2004
Accepted: Feb. 2005
Published: Dec. 2011

البيرسبترون متعدد الطبقات لتحليل بيانات الاقمار الاصطناعية

كمال الراوي و رائد شدفان

قسم علوم الحاسبات، كلية تقانة المعلومات ، جامعة البتراء، الاردن

الخلاصة

معماريات مختلفة مِن الشبكات النيرونية البرسبترونية متعددة الطبقات قد عُلِّمَت بواسطة الانتشار الخَلفي واستخدِمَت لتحليل صور القمر الإصطناعي لاندسات. لقد استخدمت طريقتين للتعليم: الطريقة الإعتيادية (ذات الطبقة المخفية الواحدة $M-H_1-L$ وذات الطبقتان المَخفيتان $M-H_1-H_2-L$) و طريقة واحد مقابل الكل (ذات الطبقة المخفية الواحدة $M-H_1-1$ وذات الطبقتان المخفيتان $M-H_1-H_2-1$). دقة التصنيف وصلت إلى ٩٠% باستخدام طريقة واحد مقابل الكل للمعمارية ذات الطبقتان المخفيتان. أداء طريقة واحد مقابل الكل كانت أفضل قليلاً من الطريقة الاعتيادية.

Introduction

The principle of the Back Propagation (BP) Artificial Neural Networks (ANN) has been developed by Werbos [1]. In 1980's the algorithm of the BP ANN was further developed independently by many authors Le Cun [2], Parker [3], and Rumelhart *et al.* [4].

Computer analysis of satellite images has been a very active area of research Benediktsson *et al* [5]. Conventional classification is usually deployed for classifying patterns in these images. However, work in the last decade involved the use of neural networks in place of conventional classifiers. The main advantage of neural networks over conventional classifiers as Maximum Likelihood Classifiers (MLC) that they are non-parametric.

Multi-Layer Perceptron (MLP), with BP learning, is the most commonly used neural network in the literature to classify remotely sensed data. While some authors have reported that conventional classifiers perform better than MLP Mulder & Spreeuwers [6], Solaiman & Mouchot [7], many authors have reported that MLP performed better than MLC in classifying remotely sensed data Hepner *et al*. [8], Heerman & Khazenie [9], Paola & Schowengerdt [10], Yoshida & Omatu [11].

Objectives

The objective of this work is to construct an automated system for analyzing Landsat satellite data using MLP. Different ANN architectures of MLP class are tested. Two different approaches are applied for learning, an ordinary training using fully interconnected multi-output nets tagged as (M-H_1-L & M-H_1-H_2-L) and one-against-all training using a collection of single-output nets tagged as (M-H_1-1, & M-H_1-H_2-1).

Data

The data for this study were obtained by Landsat Thematic Mapper taken for the area around the Spanish City of "Talavera de la Reina". The data contains 6 features for around 65,000 exemplars; each corresponds to different electromagnetic wavelength. The spots corresponds to a total of thirteen different classes: meadow, wheat, alfalfa, mountains, three types of fallow land, two types of natural vegetation, forest, irrigated land, wetland, and river. Typically, these classes contain non-linear separations.

1. **Architecture**

The network topology used in this study is based on fully connected feed-forward ANNs. The number of nodes in the input layer is equal to the number of features presented by the data, while the number of nodes in the output layer (L) is equal to the number of classes that this data map to. At least one hidden layer must be added to the architecture in order to treat the non-linear separation among classes. Several networks with one and two hidden layers, with different number of nodes in each hidden layer, have been used. The architecture for a multilayer perceptron with two hidden layers (M-H1-H2-L) is shown in figure-1. M represents the number of nodes in the input layer, H_1 represents the number of nodes in the first hidden layer, H_2 represents the number of nodes in the second hidden layer, and L represents the number of nodes in t the output layer.

2. **Learning and testing**:

Two training schemes have been adopted in this work. The first, termed as ordinary training, is based on training fully connected feed-forward networks, each with 13 outputs corresponding to the 13 classes of the image. The network is trained to score 1 at the output that corresponds to the correct class, and to score 0 at all the other outputs. The second, termed as one-against-all, is based on training 13 different single-output networks as shown in figure-2, where each network corresponds to one class. The problem presented to each network is a simple 2-classes problem: one class that is associated with every network against all other classes. Eventually, these single-output networks are used to produce 13 outputs. In both training schemes, the final class is determined by choosing the output which has the highest score.

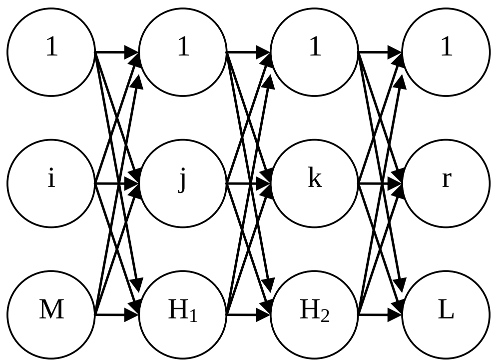


Fig.-1: The architecture of an ordinary MLP with two hidden layers. Number of nodes in the input layer, the first hidden layer, the second hidden layer, and the output layer are M, H1, H2, and L, respectively

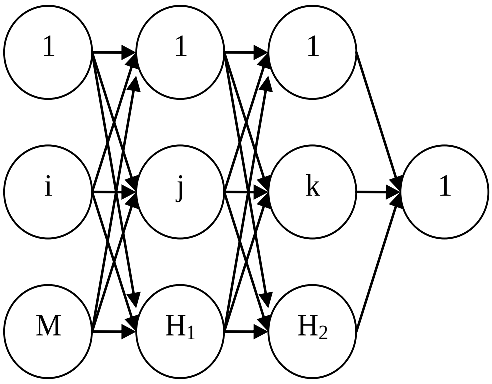


Fig.-2: The architecture of the one-against-all network. Thirteen of this ANNs have been trained. One for each class. During classification, we present the input for each one. The class will be adapted from the one with the highest score

The data was randomly mixed and split into training and testing sets with 10% and 90% of the whole data respectively. The training set includes around 6500 exemplars evenly distributed amongst the 13 classes: 500 exemplars from each class. During experiments, learning rates greater than 0.1 resulted in networks being caught in local minima. As a safety measure, the learning rate used in this study was set to 0.001. The number of epochs was chosen to be 50,000 iterations to compensate for the relatively low learning rate assigned.

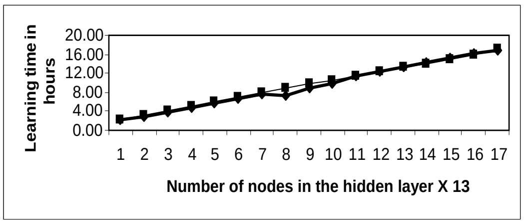
Results

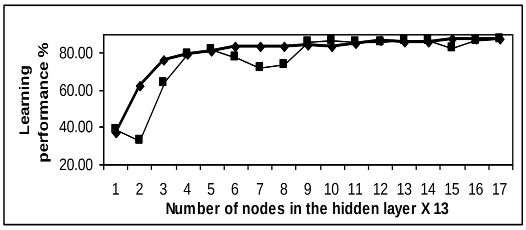
Many runs have been conducted for the two schemes using single hidden layer networks. The performance of the one-against-all scheme was a little better than the ordinary one. The results of these runs are shown in figure-3. The training and testing performance using (6-208-13) and (6-16-1)x13 network topologies were (87%, 82%) and (87%, 83%) respectively. While it was expected that the learning time for the one-against-all scheme to be lower than the ordinary scheme, due to the fewer connections, the learning time was about 16 hours for both approaches. A possible explanation for this result is the communication time calling 13 networks from the computer memory in the one-against-all networks is compensating for the computation time consumed by the processor to resolve the extra connections in the ordinary networks. However, further investigation must be conducted to reach a more solid conclusion. These runs have been conducted on a computer with 1.8 MHz processor, 128 MB RAM, and 40 GB hard disk.

More runs have been conducted, with two hidden layer networks, using one-against-all scheme only. The results of some of these runs are listed in table-1. The performance for training and testing, using (6-92-43-1) x13 topology was 93% and 90% respectively. However, the learning time was more than 130 hours. The learning time has been reduced to about 47 hours using architecture (6-42-10-1) x 13, while the learning and testing performance have been reduced by less than 2%. The classification performance at the class level were class-1 91%, class-2 87%, class-3 96%, class-4 93%, class-5 93%, class-6 92%, class-7 96%, class-8 85%, class-9 85%, class-10 94%, class-11 88%, class-12 84%, and class-13 100%.

Table-1: Some runs for different architectures using two hidden layers. One-against-all approach is used. The best classification performance is 90%. The training time for this run is 131 hours. A more practical run, with classification performance 88%, required 47 hours.

INPUT NODES	H1 NODES	H2 NODES	OUTPUT NODES	LEARNING PERFORMANCE%	TESTING ERFORMANCE%	LEARNING TIME/ hr
6	8	50	1	87	84	52
6	12	62	1	90	86	70
6	16	82	1	91	87	90
6	21	19	1	90	87	39
6	25	26	1	91	88	49
6	29	43	1	91	88	68
6	33	64	1	92	89	94
6	37	85	1	92	89	119
6	42	10	1	92	88	47
6	46	25	1	92	88	68
6	50	45	1	92	89	90
6	58	81	1	93	89	134
6	63	4	1	92	89	61
6	67	21	1	93	89	84
6	71	42	1	93	89	109
6	79	81	1	93	90	158
6	84	5	1	93	89	80
6	88	24	1	93	89	105
6	92	43	1	93	90	131
6	96	62	1	93	90	157
6	100	82	1	93	89	186





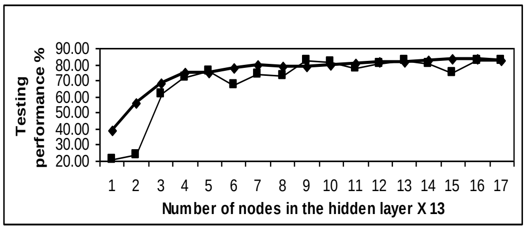


Fig.-3: The comparison of one-against-all versus the ordinary architecture. The upper chart shows the training time for one-against-all is slightly better than the ordinary one. The training and testing performance for one-against-all are little better than the ordinary one. Dark lines represent the one-against-all while the light lines represent the ordinary scheme.

The same data has been analyzed by Al-Rawi et al. [12] using Supervised ART-II that has been constructed by Al-Rawi et al. [13]. A similar performance was obtained. However, the training time was in minutes rather than days as the case of MLP. The training time using 9000 exemplars (one epoch is used) was few minutes. The training and testing performance was 94% and 86%, respectively.

Conclusions

1-The MLP gives a good performance when analyzing satellite images. However, it cannot be implemented in real time analysis due to very long training time.

2- One-against-all scheme is preferable over ordinary approach. When trying to classify a new data that has subset classes of the trained classes, one can use a subset of the trained ANNs that corresponds to the needed classes only.

3- Classification performance for MLP is little better than Supervised ART-II, however, Supervised ART-II can be implemented for real problem tasks.

Acknowledgement

We gratefully thank Professor Santiago Ormeno Villajos, Dept. of Topographic & Cartographic Engineering, University Polytechnic of Madrid, for the Landsat TM images. Our thank to Professor Consuelo Gonzalo Martin and her group, Faculty of Information Technology, University Polytechnic of Madrid, for their efforts on preparation the Landsat images.

References

[1] P. J. Werbos, Ph.D. thesis, Harvard University,Cambridge,MA,USA,(1974).
[2] Y. Le Cun, in Disordered Systems and Biological Organization, E. Bienenstock, (F. Fogelman Souli, and G. Weisbruch, Eds.), France, Spring-Verlag, (1986) 233-340.
[3] D. Parker, MIT, Cambridge, MA, USA, technical report TR-87 (1986).
[4] D. E Rumelhart,., Hinton, G. E., and Williams, R. J., *Parallel distributed Processing: Explorations in the Microstructure of Cognition* (D. E. Rumelhart and J. L. McClelland, Eds), MIT Press, Cambridge, Massachusetts, (1986) 318-362.
[5] J. A. Benediktsson, Swain, P. H., and Ersoy, O. K., *IEEE Transaction on Geoscience and Remote Sensing*, **28** (1990) 540-552.
[6] N. J. Mulder, and Spreeuwers, L., *International Geoscience and Remote Sensing Symposium (IGARSS´91).* Espo, Finland, (1991) 2211-2213.
[7] B. Solaiman, and Mouchot, M. C, *International Geoscience. and Remote Sensing Symposium (IGARSS´94),* Pasadena, CA, USA, (1994) 1413-1415.
[8] G. F. Hepner, Logan,T., Ritter, N., and Bryant, N., *Photogrammetric Engineering & Remote Sensing*, **56** (1990) 469-473.
[9] P. D Heermann, and Khazenie, N., *IEEE Transaction on Geoscience and Remote Sensing,* **30** (1992) 81-88.
[10] J. D Paola, and Schowengerdt, R. A, *Photogrammetric Engineering & Remote Sensing*, **63** (1997) 535-544.
[11] T. Yoshida, and Omatu, S., IEEE Transaction on Geoscience and Remote Sensing, **32** (1994) 1103-1109.
[12] K. R Al-Rawi, Gonzalo, C., and Martinez, E., Remote Sensing in the 21st Century: Economic and Environmental Applications, (Casanova ed.), Balkema, Rotterdam (2000) 229-235.
[13] K. R Al-Rawi, Gonzalo, C., and Arquero, A., European Symposium on Artificial Neural Network ESANN´99, Bruges, Belgium (1999) 289-294.