Diagnosis and Modelling of Alzheimer's Disease through Neural Network Analyses of PET Studies

J.S. Kippenhan and J.H. Nagel

Department of Biomedical Engineering
University of Miami, Coral Gables, FL

ABSTRACT

The back-propagation neural network algorithm was applied to the analysis of regional patterns in cerebral function, as demonstrated in positron emission tomography (PET). A trained network was able to successfully distinguish PET scans of normal subjects from PET scans of Alzheimer's Disease patients. It is concluded that the combination of PET and neural networks is a useful diagnostic tool for Alzheimer's Disease. A new paradigm for back-propagation learning is discussed which emphasizes its similarity to template matching. It is demonstrated that, under certain circumstances, the back-propagation network can be used as an estimation tool, as well as a classification tool, i.e., a trained neural network can indicate the criteria by which its classifications are performed.

INTRODUCTION

The recent wave of interest in neural networks has spread across several disciplines, including the diverse fields of mathematics, psychology, artificial intelligence and signal processing. Neural network algorithms such as the back-propagation network [1] are now being widely applied, even though the means by which they work are often not completely understood.

Back-Propagation Neural Networks: The back-propagation algorithm can be successfully applied with relative ease and generality; these factors have made it a very popular approach to pattern recognition problems. With this algorithm, the user is not required to perform prior statistical analyses or to devise classification rules. The network, in a sense, performs these tasks itself, "learning" to classify by example. During the supervised learning phase, examples of previously-classified patterns are repeatedly presented, and network connection weights are adjusted until an appropriate input-output mapping is achieved.

Functional Imaging and Cerebral Disease: Positron emission tomography (PET) has greatly enhanced studies of the functional effects of various brain disorders. PET images are based on cerebral metabolism, and thus enable regional quantification of brain function. The diagnostic power of PET depends on its ability to indicate deficits or other abnormalities in function.

Recent work [2,3] has been directed toward the discovery of characteristic functional abnormalities in PET scans of patients with Alzheimer's Disease (AD). The abnormalities exhibited in such problems, involving the complex and little-understood interactions of multiple brain regions, can be expected to be rather complicated. To date, no distinguishing pattern of abnormality in AD cases has been found. Considerable variability exists in the patterns of regional metabolism in normal PET scans, making it difficult to distinguish between a "normal" and "abnormal" scan when using standard statistical methods.

THEORY

Back-propagation networks can perform impressive feats of learning. If viewed as a "black box" process, this learning ability can seem quite mysterious. Though the learning process is seen as somewhat heuristic, it is essentially based on statistical associations [4,5]. This statistical basis is a non-parametric one; no assumptions regarding the shape of input distributions (e.g. gaussian) are made. This property can be advantageous, allowing back-propagation classifiers to be robust in the face of non-gaussian distributions.

There are a few accepted paradigms for the back-propagation network's learning process. The learning process of networks with back-propagation architecture has been compared to multiple regression analysis[4]. From another perspective, the network learns to synthesize hyperplanes of separation, thus partitioning the input space into hypervolumes which form regions containing the separate classes [5]. From a "mapping" perspective, points in the "input space" are **mapped**, by successive transformations of coordinate systems, into a space in which linear discrimination can be performed. These paradigms describe the network's learning process from different, equally valid point of view.

A main objective of the work presented here is to investigate the power and utility of an additional paradigm; i.e., a perspective from which the network's processing can be viewed as a **correlation analysis** or **template matching** process. From this point of view, the weights associated with individual hidden units serve as "templates" whose features are appropriately adjusted during back-propagation learning.

The back-propagation algorithm, based on the generalized delta rule, is described elsewhere[1,5]. The main topic of interest here is the interaction between the first "hidden" layer and the input layer. In the notation of Rumelhart, *et al.*[1], the output of a hidden unit (after presentation of a given pattern) is obtained by applying a semi-linear transfer function f to its input, i.e.

$$o_{pj} = f_j(net_{pj}) \qquad \text{Eq. 1}$$

where the subscript j refers to the unit number and the subscript p designates the p^{th} pattern. The input is given by

$$net_{pj} = \sum w_{ji} o_{pi} + \theta_j, \qquad \text{Eq. 2}$$

where w_{ji} refers to the weight from the j^{th} unit to the i^{th} input component, o_{pi} refers to the i^{th} component of the p^{th} pattern, and θ_j is a bias, or threshold.

The input to the transfer function of each hidden unit can be expressed as the inner product of the unit's current weight vector **W** with an input pattern vector X, i.e.,

$$net_{pj} = \mathbf{W} \cdot \mathbf{X} \qquad \text{Eq. 3}$$

This relationship accommodates the usual practice of including a bias, or threshold, since each of the above vectors is considered to have been augmented by an additional component. The additional component of X is set equal to one. The additional component (weight) included in **W** is applied to the extra "input" component (equal to one) and is adjusted just like the other weights.

The common practice in implementing the back-propagation network is to have one output unit per class, and to train the network so that output values are high (close to one) when a pattern of the appropriate class is presented, and low (close to zero) otherwise. If each hidden unit is considered to be a "template matching" device, then this would provide the network with a basis for detection based on high or low outputs. Vector theory says that the inner product of two vectors is maximized when the vectors are parallel. If a hidden unit is presented with an input pattern representing an n-dimensional vector in the input space, the unit's output will be maximized when its weight vector is similar in "direction" to the input. In this way, a network with N hidden units can serve as a detector of N features, with the output of each feature detector weighted, in the next layer, according to that feature's importance to the classification process.

The geometric interpretation of the above paradigm is as follows: each set of weights specifies a hypersurface (a semi-linear hyperplane). The unaugmented weight vector (not including the threshold weight) points in a direction normal to the hyperplane, and the threshold or bias serves to specify the hyperplane's "intercepts". It can be shown that the analogy to template matching is most exact for zero-mean input vectors. Zero-mean input patterns will reside in a hyperplane that passes through the origin. In this case, if a network synthesizes hyperplanes which separate the input space according to classes, it will also synthesize vectors that will point either toward or away from the patterns thus separated. The template matching model breaks down somewhat when more than one class region lie along the same "direction" from the origin. In this case, the threshold value (which serves to "translate" the separation hyperplane) becomes more influential.

CH2936-3/90/0000-1449 $01.00

The above discussion points to the ability of the back-propagation network to perform as an "estimation" tool, in addition to its proven value as a classification tool. In a sense, it can perform "supervised clustering". The network algorithm has an advantage over simple, unsupervised clustering in its ability to accommodate cases in which single classes occupy more than one distinct region.

METHODS

Data was obtained from [F-18] FDG PET scans of normal and AD subjects. Data sets were prepared from studies of fifty-two subjects: twenty-two subjects who had been clinically diagnosed with Alzheimer's Disease, and thirty age-matched normal subjects. Duara, *et al.*[2] describe the methods of data preparation. Data for each patient consisted of eight values, representing cerebral glucose metabolism in the eight lobes of the brain: left and right frontal, parietal, temporal and occipital. The mean of each data set was calculated and removed, and the resulting eight-dimensional pattern vectors were used to train a two-layer back-propagation neural network with two hidden units, as shown in Figure 1.

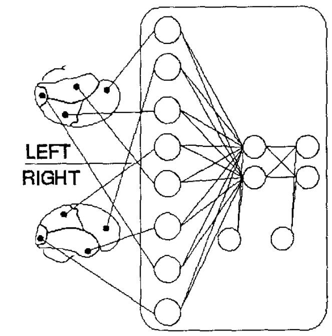

Figure 1: Illustration of the method by which data from brain PET studies is presented to the neural network.

For classification purposes, the network was trained with data from twenty-six subjects (fifteen normal, eleven AD), representing one-half of the above subject group. The trained network was then tested on the remaining half.

RESULTS

The back-propagation neural network classifier proved to be successful in distinguishing normal PET scans from those of patients diagnosed as "probable AD". Within the test group, the trained network's classification agreed with the clinical diagnosis in 24 of 26 cases. One patient from each class was misclassified, resulting in a sensitivity for AD of 91%, and a specificity of 93%. The accuracy of the network's classification is compared to that of more traditional methods in Figure 2. As shown in this figure, the neural network is more immune to false positives, as indicated by the higher specificity values, than are the other methods.

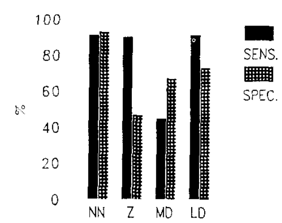

Figure 2: Comparison of classification methods by sensitivity and specificity. Methods are: NN - Neural Network, Z - Z-score method[2], MD - Minimum Distance, LD - Linear Discriminant.

To test the network's usefulness as an estimation tool, the entire group of subjects, with the exception of the two subjects which had been mis-classified, was used to train the network. This larger group was used to achieve greater generality. The weight vectors of hidden units were examined after the network had converged to a high degree, i.e. outputs were close to their target values for all subjects.

It should be noted that weight vectors will repeatedly converge to the same values only under these fairly strict conditions. Results indicate that weight vectors will consistently converge to the same pattern only when (1) all inputs have zero mean, and (2) the network has converged for all input patterns.

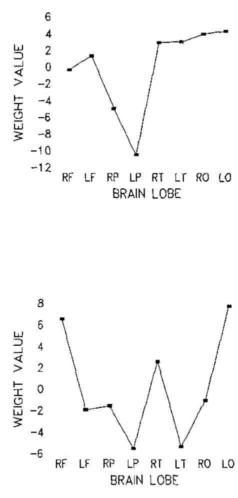

Figure 3: Plots of weight vectors of hidden units 1 (above) and 2 (below) within the trained network.

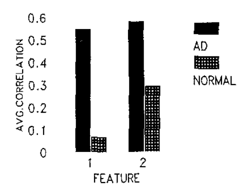

Figure 4: Comparisons of feature correlations with normal and AD patterns.

For the network trained with the above group of subjects, the weight vectors of the two hidden units consistently converged to the patterns shown in Figure 3. It is interesting to note that each of these patterns contains attributes that had been previously observed clinically, such as asymmetry with left impairment, decrease in parietal function, and sparing of the occipital region[2]. The network can be seen to have "learned" which attributes and combinations of attributes were important for diagnosis.

Figure 4 demonstrates that the mean statistical correlation of both weight vectors with AD patterns are higher than correlations of like weight vectors with normal patterns. Thus, each of these weight vectors can be regarded as "features" likely to be found in AD PET scans. Results such as these may prove to be useful for "modelling" AD and other cerebral diseases, i.e. investigating their physiological nature and origins.

REFERENCES

[1] Rumelhart, D.E., Hinton, G.E. and R.J. Williams. Learning internal representations by error propagation, *Parallel Distributed Processing*, Vol.1, Rumelhart, D.E. and J.L. McClelland and the PDP Research Group, MIT Press, 1986, pp. 318-364.

[2] Duara, R., Barker, W., Loewenstein, D., Pascal, S., and B. Bowen. Sensitivity and specifity of positron emission tomography and magnetic resonance imaging studies in alzheimer's disease and multi-infarct dementia. *European Neurology*. Vol. 29(suppl 3), pp.9-15, 1989.

[3] Kippenhan, J.S., W. Barker, S. Pascal, R. Duara. A neural-network classifier applied to PET scans of normal and Alzheimer's Disease patients. Accepted Soc of Nuc Med 37th Annual Meeting, June, 1990.

[4] Stone, G.O. An analysis of the delta rule and the learning of statistical associations. *Parallel Distributed Processing*, Vol.1, Rumelhart, D.E. and J.L. McClelland and the PDP Research Group, MIT Press, 1986, pp. 444-459.

[5] Pao, Yoh-Han. *Adaptive Pattern Recognition and Neural Networks*. Addison-Wesley, 1989.

Jonathan Shane Kippenhan
P.O. Box 248294
Department of Biomedical Engineering
University of Miami
Coral Gables, FL 33124
(305) 284-2442