AL-SR-1992- 0026

ARMSTRONG LABORATORY

BACKPROPAGATION AND EEG DATA (U)

Paul E. Morton, Lt Col, Ph.D., M.D.
Glenn F. Wilson, Ph.D.

Armstrong Aerospace Medical Research Laboratory
Human Engineering Division

October 1988

Final Report for Period 19 December 1988 to 30 June 1989

Approved for public release; distribution is unlimited.

AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

NOTICES

When US Government drawings, specifications, or other data are used for any purpose other than a definitely related Government procurement operation, the Government thereby incurs no responsibility nor any obligation whatsoever, and the fact that the Government may have formulated, furnished, or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication or otherwise, as in any manner licensing the holder or any other person or corporation, or conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

Please do not request copies of this report from Armstrong Aerospace Medical Research Laboratory. Additional copies may be purchased from:

National Technical Information Service
5285 Port Royal Road
Springfield, Virginia 22161

Federal Government agencies and their contractors registered with Defense Technical Information Center should direct requests for copies of this report to:

Defense Technical Information Center
Cameron Station
Alexandria, Virginia 22314

TECHNICAL REVIEW AND APPROVAL

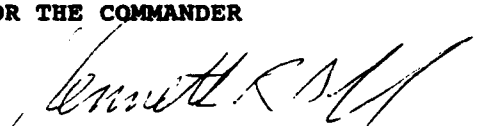
AL-SR-1992-0026

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

The voluntary informed consent of the subjects used in this research was obtained as required by Air Force Regulation 169-3.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER



KENNETH R. BOFF, PhD, Chief
Human Engineering Division
Crew Systems Directorate
Armstrong Laboratory

Oct 1988 Final Report (Aug 1989-Aug 1990)

NUMBERS

Backpropagation and EEG Data (U)

PE 62202F
PR 7184
TA 14
WU D8

Lt Col Paul E. Morton
Glenn F. Wilson

Armstrong Aerospace Medical Research Laboratory
Human Engineering Division, AFSC, HSD
Wright-Patterson AFB OH 45433-6573

PERFORMING ORGANIZATION REPORT NUMBER

AL-SR-1992- 0026

SPONSORING/MONITORING AGENCY REPORT NUMBER

Published in Fourth Annual AAAI Conference Proceedings, 25-27 Oct 88, Volume 2, pages 81-86.

DISTRIBUTION CODE

Approved for public release;
distribution is unlimited.

The development of neural networks has pursued a myriad of different courses reflecting the interests of a large number of researchers from highly varied backgrounds. This paper would like to focus on one point of this 'many faceted gem,' as Stephen Grossberg [1] described the field. The point of focus will be to address some of the practical results of applying a backpropagation trained net to raw electroencephalogram (EEG) data. Much important work on more efficient training rules has been done; however, equally critical is consideration of the information content of the data, the net size, number of hidden nodes and order of training data. This paper explores some of the training issues raised by applying backpropagation to this very complex data.

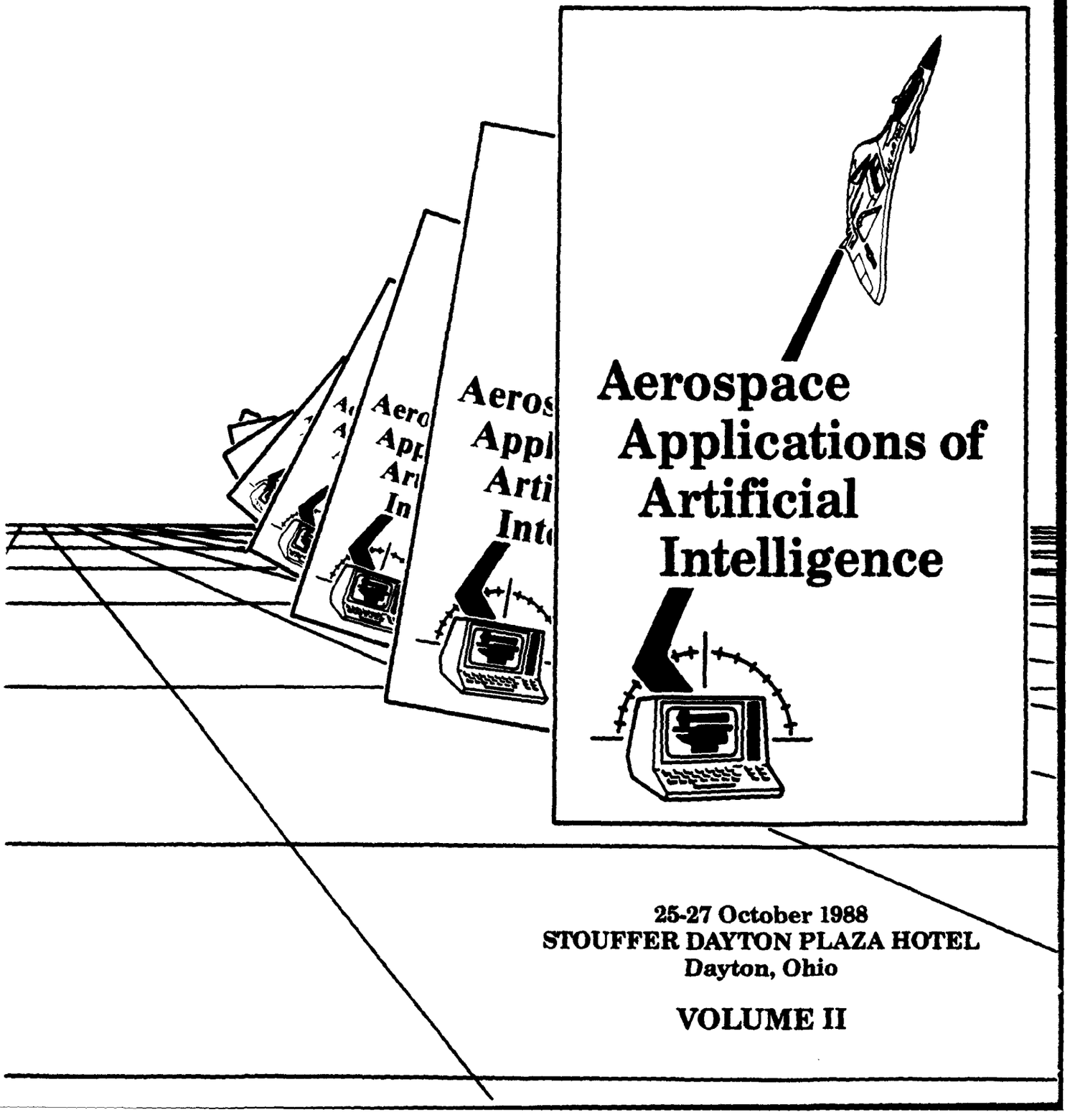
Neural nets, EEG, evoked response
Classification

NUMBER OF PAGES
6

PRICE CODE

SECURITY CLASSIFICATION OF REPORT	SECURITY CLASSIFICATION OF THIS PAGE	SECURITY CLASSIFICATION OF ABSTRACT	LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Fourth Annual AAAI Conference
PROCEEDINGS



25-27 October 1988
STOUFFER DAYTON PLAZA HOTEL
Dayton, Ohio

VOLUME II

Backpropagation and EEG data

P. Morton G. Wilson

October 25, 1988

1 Introduction

The development of neural networks has pursued a myriad of different courses reflecting the interests of a large number of researchers from highly varied backgrounds. This paper would like to focus on one point of this 'many faceted gem', as Stephen Grossberg [1] described the field. The point of focus will be to address some of the practical results of applying a backpropagation trained net to raw electroencephalogram (EEG) data. Much important work on more efficient training rules has been done; however, equally critical is consideration of the information content of the data, the net size, number of hidden nodes and order of training data [4]. This paper explores some of the training issues raised by applying backpropagation to this very complex data.

2 Purpose

For a long time much work has been done to develop objective methods of measuring mental workload. The value of this would be immense especially in the design and evaluation of new technology. Unfortunately, due to the difficulty of processing physiologic data and extracting subtle patterns from the noise, an objective workload metric is still not a reality. Neural nets may provide a way to process the noisy and complex physiologic data to generate a useful metric. As a first step this paper describes some of the problems applying a neural net to EEG data generated by the so called Audio Oddball Paradigm [5]. The net used was varied greatly in size and in specifics of training. Future work will use these nets to evaluate the potential utility of the data contained in the EEG as processed by the net in moving toward a useful workload metric.

3 Methods

The data were collected from three different leads placed on a subjects head all in the midline spaced front to back with a common ground. The data were low pass filtered at 30 Hz and digitally sampled at 200 Hz. Epochs of 1 second in length were synchronized with the audio tone stimuli. Two audio tones of 1000 Hz and 1050 Hz were presented to the subject with the higher tone presented rarely

(about10–30% of the time). The tone duration was 0.5 seconds and several seconds between tones given. The occurrence of the rare (higher) tone was randomnized except that no two rare were allowed to come back to back. The resultant EEG evoked responses were digitized and over several sessions large data sets of single trial evoked response EEG signals were collected. These data were then processed into training and test sets by discarding randomly many of the frequent responses so that the ratio of frequent to rare responses would be between 1:1 and 3:1. Standard backpropagation [3] trained perceptrons were implemented using these data. A sigmoid transfer function was used and the learning rate was kept constant throughout.

3.1 Initial efforts

Initially a subject was given the Audio-Oddball test where the rare tone occurred only 10% of the time. This gives a slightly higher quality evoked response but requires more time to collect a large number of rare responses. A set of 83 training vectors and 33 test vectors with about a 3:1 ratio of frequent to rare responses was assembled. A three layer net having 400 inputs, 60 nodes in the first hidden layer, 40 nodes in the second hidden layer and 2 output nodes (written 400x60x40x2) was constructed. The 400 inputs represent two channels of EEG data where each channel represents 200 points to cover the one second epoch. This was trained on the training set and got 79% correct classification of the test set. Based on this result the next step was to collect a larger training set and expand to three channels of inputs.

3.2 Large 3 layer nets

In order to generate a larger data base the percent of rares in the Audio-Oddball test was increased to 30% and three channels were collected. A training set of 269 vectors and a test set of 67 were randomly generated each with 3 channels. The ratio of the rares to frequents was about 40% by discarding some of the frequent events randomly. As shown in table 1 several sizes of large nets were tried. The 600x200x200x2 net after more than a week of computer time showed no trend toward convergence at all. Interestingly, if the net is first trained with the two average vectors for rare and frequent responses and then with the individual vectors of the test set it does converge as reflected in Table 2 below. Also reducing the number of hidden nodes by 50 caused convergence regardless of whether the first hidden layer or the second was reduced. Surprisingly, with further reduction in total hidden nodes the net showed improved performance until very small layers were used, then the performance degraded and finally the net would not converge. One limitation to this approach is long training times for these large nets.

3.3 Training with Average vectors

To improve generalization and shorten training times the entire training set was

Net Size	*Classification Rate after...*		
	average vectors	*5% training set*	*100% training set*
600x200x200x2	55.2%	.	74.6%
600x150x100x2	67.2%	73.1%	62.7%
600x100x150x2	68.7%	65.7%	.
600x80x50x2	68.7%	.	.
600x70x20x2	70.2%	82.1%	79.1%
600x60x30x2	76.1%	76.1%	71.6%
600x60x20x2	73.1%	79.1%	68.7%
600x50x20x2	67.2%	79.1%	.
600x40x20x2	67.2%	.	.
600x30x60x2	72.6%	74.6%	.
600x20x70x2	62.7%	.	.
600x10x10x2	67.2%	.	.

Table 2: Nets trained on average data.

Net Size	*Classification Rate*
600x200x200x2	does not converge
600x150x200x2	58.2%
600x200x150x2	59.7%
600x45x45x2	73.1%
600x10x10x2	70.2%
600x5x1x2	68.2%
600x1x1x2	does not converge

Table 1: Results With 3 Layer Nets

used to compute two average vectors. One represented the average of all the rare responses and the other frequent responses. Next two very different approaches were followed to make use of the average vectors: The first approach was to use the same large 3 layered net and first train with the average data to a modest average error (.01–.05). Then this partially trained net was trained further with just a few of the training vectors and then the entire training set was used to train the net even further. Next its performance was measured by the test set. Table 2 shows these results. In some cases the second step of training with the partial training set was omitted. Not all net sizes were investigated past training with the average vectors. The second approach was to divide the net into two parts, one having only one hidden layer and the other having only an output node. The first net was 600x50x600 and was trained on the entire training set to produce either the average rare or average frequent vector as its output. The second one layer net was trained with the two average vectors only to generate the correct output (i.e. 0 for frequent and 1 for rare). The two nets were then concatenated to form a 600x50x600x1 net. The test set was then used as input to this

combined net and its correct classification rate was 69%.

3.4 Two layer nets

Next the effect of using an even smaller net with only two layers was explored. A net of size 600x5x2 was trained on the average vectors and then applied to the test set with a 69% correct classification rate. This was astounding in light of the huge difference in the size of the nets. This performance was comparable to the best results obtained with much larger nets and much more training.

3.5 Review of the data

After the above work the accuracy of the classification of the test set was stuck at about the 70–80% mark. This produced a desire to find new directions for improvement. In looking at the results of the net on the training set where very high correct classification rates were expected, it was noted that some of the training vectors were consistently misclassified. To explore this observation, the entire training set was graphed and reviewed. It was noted that quite a few of the individual training vectors did not look as expected but rather in some cases had the opposite appearance, that is some rares looked like frequent responses and some of the frequents looked like rare responses. This may in fact be due to an inappropriate response by the subject. The subject may at first actually think a rare tone is a frequent and then after the evoked response has been recorded decide it is actually a rare. In an attempt to address this confusion the training set and test set were graded into 4 classes: True and false for both rare and frequent responses. Next a series of 2 layer nets with 4 outputs were trained and tested. Table 3 shows these results. The best accuracy of 65% is with 80 hidden nodes. This result suggests that there is something in the raw EEG data that divides it in 4 or more classes. We are now in the process of using a Kohonen self organizing net to sort the training set into several classes and attempting to refine the quality of the training data. This will no doubt improve the performance and increase our understanding of the original signals.

Net Size	*Classification Rate*
600x10x4	58.2%
600x20x4	59.7%
600x30x4	52.0%
600x40x4	56.7%
600x50x4	56.7%
600x60x4	55.2%
600x70x4	59.7%
600x80x4	65.0%
600x100x4	53.7%

Table 3: 2 Layers / 4 Classes

4 Conclusions

Neural networks can be trained to differentiate between one of two single trial evoked response EEGs at better than 82%. This is remarkable in light of the noise and confusion involved in real EEG data. A

great deal more effort will be required to achieve a useful mental workload metric; however, this effort establishes the usefulness of neural nets for this type of data. The following areas yielded the following conclusions based on the above data:

4.1 Size of the data set

More training data are not necessarily better! When the size of the training file was expanded from 87 to 216 the performance of the nets decreased. The best results were obtained by using only the first 10 of the vectors in the training set. For most cases the performance decreased with the entire training set as seen in table 2.

4.2 Size of the net

One size fits all nets! The several ranges of nets explored all showed the same size of hidden layer where performance was optimum. Outside this point the nets performance varied within a small range until it fell off steeply at the extremes of size. This suggests that trying nets over a wide range of sizes may be necessary if optimum performance is required. If only modest performance is needed this may be achieved with much less experimentation. The optimized size for all the nets was about 90 total hidden nodes. For 2 hidden layers these 90 nodes worked best when divided 70:20 (3.5:1) between first and second hidden layers. The large nets train in fewer steps but require much more time per step when implemented on a standard computer. There was a tendency for the order of training data to be critical on nets with over 350 hidden nodes as evidenced by the non-convergence of the 600x200x200x2 net unless first trained with the two average vectors. Lippmann [4] noted similar results on speech data with similar nets. He pointed out that using about twice the number of nodes as required by the number of classification regions would give good performance. This raises the question: Is the number of nodes required for empiric optimum performance a rough estimation of the relative number of classification regions inherent in the data? For some problems this would be a useful result and it deserves further exploration.

4.3 Using average data

All you need is a few good averages! The use of average data was often better than the entire training set and requires a fraction of the number of training steps and time. A selected set of the training vectors does improve the results but these are not as easy to define as an average. Several authors [1, 4] have described enhanced training by using the vectors that lie on the boundaries between classification regions. This likely explains why training with a few of the training vectors after the average ones increases performance in almost all cases.

4.4 Concatenated nets

Concatenated nets work! Novel training approaches worked equally well here. In some problems concatenated nets may provide the only way to a solution. This

may also allow the training of specific layers for specific functions as demonstrated by the 600x50x600x1 concatenated net described above. For complicated problems this may provide useful output from the middle of the net as well as from the output layer. It could also give some insight into the meaning of hidden layer weights.

4.5 The data

Fuzzy data was a bear! There is no 'gold standard' for EEG data to use as a benchmark. The failure of the nets to do better is in part related to a lack of precise knowledge of the data. For the 4 class problem some improvement is made by looking closely at the vectors in the training set which failed to train. Some of these vectors may be reclassified on review and may improve the classification rate of the test set after training. It may be that transforming the evoked responses to the frequency domain may improve classification. The data is now being explored with self organizing Kohonen nets and it appears there are several clusters of responses in even this simple problem. Clearly, to expect to able to handle complex real world data we must build on this start and bring to bear the power of neural networks to gain insights into the data to be classified. The early results noted above provide some flicker of promise to do this.

References

[1] S. Grossberg. *Presidents Address* First Annual INNS Meeting, 1988, Boston.

[2] S. Ahmad and G. Tesauro. *A Study of Scaling and Generalization in Neural Networks* 1988, Abstracts of First Annual INNS Meeting, Boston Vol. 1 Supp. 1, Pergamon Press.

[3] R. Lippmann, *An Introduction to Computing with Neural Nets*, IEEE ASSP Magazine, Apr 1987, pg 4-22.

[4] W. Huang and R. Lippmann, *Comparisons Between Neural Net and Conventional Classifiers*, ICNN, San Diego, Jun 1987.

[5] W. Ritter et. al. *Manipulation of ERP Manifestations of Informational Processing Stages*, Science Vol 218 pg 909-911, 1982.