

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank) 2. REPORT DATE 3. REPORT TYPE AND DATES COVERED
FINAL/01 NOV 89 TO 31 OCT 92

4. TITLE AND SUBTITLE
STATISTICAL BENCHMARKS FOR NEURAL NETWORK PERFORMANCE

5. FUNDING NUMBERS

6. AUTHOR(S)
TERRENCE L. FINE AND THOMAS W. PARKS

2304/CS
AFOSR-90-0016

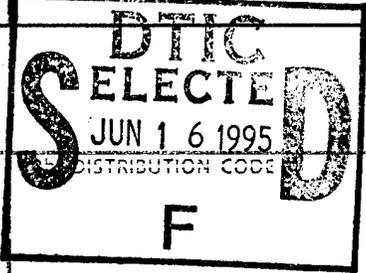
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
CORNELL UNIVERSITY
SCHOOL OF ELECTRICAL ENGINEERING
ENGINEERING AND THEORY CENTER BUILDING
ITHACA, NY 14853-3801

8. PERFORMING ORGANIZATION REPORT NUMBER
AFOSR-TR-95-0290

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NM
110 DUNCAN AVE, SUTE B115
BOLLING AFB DC 20332-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFOSR-TR-95-0390
AFOSR-90-0016

11. SUPPLEMENTARY NOTES



12a. DISTRIBUTION/AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

13. ABSTRACT (Maximum 200 words)
Much of our effort was devoted to establishing statistical performance bounds for neural networks, acting as pattern classifiers, through improvements to Vapnik-Chervonenkis their (VCT) This theory addresses the interrelationships between the complexity of a network, the amount of training data, and the statistical reliability/performance of the trained networks on independent testing data, and the statistical reliability/performance of the trained network on independent testing data. The troubling chasm between the predictions of VCT and the experience of practitioners remains to be crossed. Our extensive research into reductions of the sample size estimates produced by VCT and into other improvements of the VCT arguments have yet to yield results of practical significance that can serve as advice to neural network designers In the last year of this program we also initiated work on the use of neural networks as time series forecaster. Our work on time series forecasting has been more successful. The continuation of this work under NSF sponsorship has resulted in a state-of-the-art forecaster for short-term electric load prediction.

DTIC QUALITY INSPECTED 3

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
UNCLASSIFIED

18. SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. SECURITY CLASSIFICATION OF ABSTRACT
UNCLASSIFIED

20. LIMITATION OF ABSTRACT
SAR(SAME AS REPORT)

Recd 4/22/93 Accp 4/14/95

School of Electrical Engineering

Engineering and Theory Center Building FAX: 607-255-9072
Ithaca, NY 14853-3801

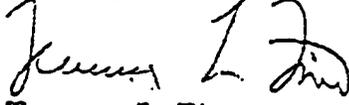
7 May 1993

ATTN: Ms. Louise Harrison
AFOSR/PKO
Bldg. 410, Rm C-124
Bolling AFB, DC 20332-6448

Dear Ms. Harrison:

I enclose two copies of my final technical report on AFOSR-90-0016. This report has been delayed and was requested on 16 April by Cindy G. Dahlstrom, Chief, Contracts/Grants, Administration Division. Please forward these copies to whomever requires them.

Sincerely yours,



Terrence L. Fine
Professor

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

19950614 051

Final Technical Report on AFOSR-90-0016
Statistical Benchmarks for Neural Network Performance
P.I.: Terrence L. Fine, Thomas W. Parks

Period Covered: 1 November 1990-31 October 1992
PM: Dr. A. Nachman

Abstract

Much of our effort was devoted to establishing statistical performance bounds for neural networks, acting as pattern classifiers, through improvements to Vapnik-Chervonenkis theory (VCT). This theory addresses the interrelationships between the complexity of a network, the amount of training data, and the statistical reliability/performance of the trained network on independent testing data. The troubling chasm between the predictions of VCT and the experience of practitioners remains to be crossed. Our extensive research into reductions of the sample size estimates produced by VCT and into other improvements of the VCT arguments have yet to yield results of practical significance that can serve as advice to neural network designers. In the last year of this program we also initiated work on the use of neural networks as time series forecasters. Our work on time series forecasting has been more successful. The continuation of this work under NSF sponsorship has resulted in a state-of-the-art forecaster for short-term electric load prediction.

Summary

As explained previously, at the outset of research we revised our program objectives to concentrate on probabilistic studies of:

the ability an artificial neural network (ANN) to generalize, the core of the benchmarking problem being approached through Vapnik-Chervonenkis theory (VCT);

the development of ANNs using discrete-valued nonlinear nodes such as the classical linear threshold units;

the ability of neural networks to implement nonlinear forecasters.

Secondary objectives were to develop a graduate program in electrical engineering at Cornell in the area of ANNs. Our ANN research group expanded to include several undergraduates, one M.Eng.Elec., two M.S., and two Ph.D. students. We continue to offer the only course in ANN at Cornell as well as a companion course on pattern classification that surveys standard probabilistic approaches. This summer we expect to see that completion of the two M.S. programs, with the students advancing to the Ph.D. program, and completion of one of the Ph.D. programs.

We turn now to review some of our results and this review will overlap with our previous interim technical report. Our study of VC theory attempted to understand its

inadequacy as a guide to practice; VC theory predicts a need for training sample sizes that are orders of magnitude greater than those used happily by practitioners. We made several attempts, reported in Fine [1991], to lower the VC upper bound estimates of training sample size n required to select a net η^* whose error probability performance \mathcal{E}^* is within ϵ of the lowest error probability \mathcal{E}^0 achievable by nets in a given family/architecture \mathcal{N} having a VC capacity/dimension V ; VC theory estimates $n = O(\frac{V}{\epsilon^2} \ln \frac{1}{\epsilon})$. Even a small net architecture can have $V > 100$ and $\epsilon < .1$ is far from stringent. VC theory then suggests that successful training will require $n = O(10^5)$ a far larger training set than is used in all but character recognition programs.

Our failures to reduce the VC upper bounds caused us to reconsider this problem and led us to show that for any VC capacity V there exist architectures $\mathcal{N}_1, \mathcal{N}_2$ of this capacity such that for \mathcal{N}_1 we need $n = O(\frac{V}{\epsilon^2})$ to select a net η^* whose performance is within ϵ of the performance of the best net η^0 in \mathcal{N}_1 while for \mathcal{N}_2 we require a sample size $n = O(\frac{V}{\epsilon})$. This latter result is within a factor of $\frac{1}{\epsilon} \log \frac{1}{\epsilon}$ of the VC bound. More recent research improved these estimates by studying the baseline case of a single linear-threshold-unit (perceptron) operating on normally distributed data. We found a necessary relationship between these parameters of

$$n = \frac{v}{70\epsilon^2}.$$

Hence, any universal VC-type bound must be at least as great as this and it must be $O(\frac{v}{\epsilon^2})$.

In the meantime it was reported that the mathematician M. Talagrand had obtained the 'best possible VC bounds' and we were finally able to obtain a preprint of his paper this past winter. Talagrand obtains the best possible exponents and his results prove that the $\log \frac{1}{\epsilon}$ factor is not needed. However, his constants are undetermined and it is impossible to obtain quantitative conclusions from his results without a large amount of difficult work. Talagrand recognizes this and leaves the necessary calculations to "those with a taste for it". It is clear from our result quoted above for the perceptron, that the issue has become one of the 'constants'. Our lower bound and his upper bound are both $O(\frac{v}{\epsilon^2})$. Our research on this problem continues, although we have yet to discern a general argument for reducing the VC bounds. Indeed, we have come to suspect, and are attempting to verify, that the VC bounds are roughly correct for the problem they address of uniform approximation. The gap between theory and practice may arise from the difference in sample sizes required to guarantee that all networks will have training set performance within ϵ of their of their true statistical performance and the sample size required to guarantee this for just the network selected by the training algorithm. The difficulty is in assessing this quantity.

Our second direction of research is in the application of neural networks to prediction and to classification. We have studied the use of discrete-valued nodes (binary or ternary valued) as pattern classifiers (Fine [1991]). We are also engaged in an application of neural networks, having the usual logistic units, to forecasting demand for electric power to be supplied by a large electric utility over the next 24 hours (Yuan and Fine [1992,1993]). Such forecasts are of economic importance and impact unit commitment and possible purchases from, and sales of power to, other utilities. Experience with nets composed of only discrete-valued nodes led us to an architecture combining a linear predictor with a small net of ternary-valued nodes to nonlinearly forecast the residuals. However, we encountered sparsity problems with our training method and abandoned this direction in

favor of small nets (about five logistic nodes and about twenty weights). This compares with much more complex nets (several hundred weights) that have been reported in the literature on forecasting demand. We suspect that in many of the reported nets, most of the logistic nodes are actually operating over their linear region and can be combined into a single node. Nor could such designs yield reliable nonlinear forecasters for the number of weights to be trained was usually comparable to the number of training samples.

To assist our design of small networks, we developed a new feature selection criterion that enables us to reduce a large number (about fifty) of potential network inputs to between one and three inputs. In our application to load forecasting these small networks performed as well as any other method that had been used on this problem. This study will enable us to benchmark our neural net design against the working methods employed by electric utilities.

Publications and Talks Partially Supported by this Grant:

1. A Structure for Neural Network Pattern Classifiers, INNC-90-Paris, July 9-13, 1990, accepted for both oral and poster presentation.
2. Fine, T. [1991], Designing linear threshold based neural network pattern classifiers, in R. Lippmann, J. Moody, D. Touretzky, eds., *Advances in Neural Information Processing Systems -3-*, Morgan Kaufmann Pub., Palo Alto, 811-817.
3. Yuan, J.L., T.L. Fine [1992], "Forecasting Demand for Electric Power Using Autoregressive Neural Networks", Conf. on Information Sciences and Systems, Princeton, NJ, March 18-20.
4. Yuan, J., T. Fine [1993], Forecasting demand for electric power, in S. Hanson, J. Cowan, C. Giles, eds. *Advances in Neural Information Processing Systems -5-*, Morgan Kaufmann Pub., Palo Alto, to appear.