AD-A043 464 BROWN UNIV PROVIDENCE R I DIV OF APPLIED MATHEMATICS F/G 9/2
STATISTICAL COMPUTER PERFORMANCE EVALUATION.(U)
MAR 77 W FREIBERGER N00014-75-C-0461
UNCLASSIFIED NL

1 OF 1
AD A043464

END
DATE FILMED
9-77
DDC

DEPARTMENT OF THE NAVY
Office of Naval Research

FINAL REPORT
ON
Contracts N00014-67-A-0191-0026
and N00014-75-C-0461
entitled

STATISTICAL COMPUTER PERFORMANCE EVALUATION

by

Walter Freiberger

Division of Applied Mathematics
Brown University
Providence, R.I. 02912

March 1977

Brown University

Division of Applied Mathematics

Final Summary Report

Contracts N00014-67-A-0191-0026 and N00014-75-C-0461

with the Office of Naval Research

Jan. 1, 1973-Dec. 31, 1976, Total: $105,000

This contract had its origin in a seminal conference, sponsored by the Office of Naval Research and held at Brown University on November 22-23, 1971, entitled "Statistical Methods for Computer Systems Performance Evaluation", the Proceedings of which were published by Academic Press in 1972 under the title "Statistical Computer Performance Evaluation". It was the purpose of the conference to investigate the application of modern statistical methods to the study of computer performance - a subject which came to be known as compumetrics - and the research contract which resulted from its deliberations addressed itself to a number of specific problems in the field. We shall attempt to summarize the work done in the course of the contract by listing and annotating the reports published under its auspices. Copies of these reports are still available from the Division of Applied Mathematics.

1. "An Introduction to Compumetrics" by Ulf Grenander, July 1973, 141 pp. Chapter headings: 1. Introduction; 2. Methodology; 3. System Traffic; 4. The CPU; 5. Main Memory, 6. Secondary Memory; 7. Program behavior, program organization and program reliability; 8. Global systems studies; 9. Conclusions.

This report summarized the state-of-the-art at the time when there existed no other systematic introduction to the subject. Although it laid no claim to all-inclusiveness, it reported on the analytic methods available for the analysis of the various components of a computing system.

2. "The Optimal Choice of Window Sizes for Working Set Dispatching" by Gregory Henderson and Juan Rodriguez-Rosell, October 1973, 35 pp. (presented at the International Symposium on Operating Systems Theory and Practice, Paris, April 1974). In this report, the concept of varying window size in a working set dispatcher to control working set size and number of page faults was examined. A space-time cost equation was developed and used to compare fixed window size to variable window size for different types of secondary storage, based on the simulated execution of real programs.

3. "Some Observations on the Stochastic Behavior of Page References in Computer Programs" by Gregory Henderson, September 1974, 19 pp.. In this report, the motivation for studying the page reference behavior of computer programs was discussed, and one such study presented.

4. "Regime Behavior in Page Referencing Patterns of Computer Programs" by Paul Sampson, July 1974, 41 pp. + appendix (31 pp.). This paper described in detail a study of some of the characteristics of program referencing patterns. It applied statistical methods for the purpose of increasing our understanding of the

execution of programs in multiprogrammed environments by inferring the stochastic structure of the mechanisms underlying the generation of such patterns from real-life data. It obtained a mathematical description of these phenomena in terms of stochastic processes of a certain type.

5. "An Interactive Software Engineering Tool for Memory Management and User Program Evaluation" by Wolfgang W. Millbrandt and Juan Rodriquez-Rosell, Proc. National Computer Conference 1974, pp. 153-158.

This paper described a user-oriented tool which was developed to enable a programmer to monitor interactively the memory referencing behavior of his modules. It was used to help programmers write better localized code and thus reduce necessary memory resources, and as a tool for investigating program behavior experimentally in connection with various research projects.

6. "Patterns in Program Reference" by Walter Freiberger, Ulf Grenander and Paul D. Sampson, October 1974, 49pp. An improved version of this paper was published in the I.B.M. Journal of Research and Development, Vol. 19, No. 3, May 1975, pp. 230-243 (special issue on performance evaluation of computer systems). This paper described a study of some of the characteristics of program referencing patters. Program behavior was investigated by constructing stochastic models for the page reference mechanism and evaluating the validity of the assumptions made through comparison with empirical results. The notion of a

regime stochastic process is shown to play a useful role in describing the observed phenomena mathematically.

A discussion of this problem from the point of view of general pattern theory is included in Chapter 3 (Analysis of certain temporal patterns), Section 4 (Image analysis for other regime patterns) of the book "Lectures in Pattern Theory, Volume II: Pattern Analysis" by Ulf Grenander, to be published by Springer-Verlag early in 1978 (Volume I: Pattern Synthesis, was published in 1976).

7. "Heuristic-Adaptive Search for Regimes in a String of Data" by Beng-Tung Ang, July 1974, 166 pp. A condensed version of this report was published as "A heuristic-adaptive procedure for segmentation of time patterns" by Beng-Tung Ang in the International Journal of Computer and Information Science, volume 4, number 4, 1975, pp. 329-348.

This report treated segmentation of time patterns as a heuristic search problem. Segmentation is formulated in terms of image restoration. An observed pattern, which is the stochastically deformed image of a pure image consisting of a number of regimes, is to be segmented to recover the regime structure. Standard statistical decision methods are not very useful here because of the computational difficulties involved. The search process described here consists of application of a sequence of heuristic-adaptive operators. Each operator is designed to detect certain flaws in previous segmentations and make modifications accordingly.

The search path thus generated ends in a loop from among which the final solution is chosen by an evaluating function. Results of experiments with simulated data are presented and discussed. One of the motivations for this study was the discovery of regimes in reference strings, as discussed in reports 4 and 6 above.

8. "A Statistical Study of an Operating System" by Wen-Te Kobe Lin, July 1974, 191 pp. This study was commenced as an investigation of the CP-67/CMS virtual machine operating system for the IBM 360/67 computer, and developed into a more general investigation into the application of modern statistical techniques to the evaluation of computer systems performance. It investigated empirically the stochastic structure of random variables such as 'percentage of time spent in problem-mode', 'percentage of time spent in idle-mode', 'queue length of executable users', 'number of page I/O per second', 'number of privileged operations per second', and several others, to fit a regime stochastic process model to the performance of the operating system. It should be looked upon as a pilot study, giving an indication of the statistical methodologies appropriate for such investigations, rather than for the actual results achieved.

Note: At this point in time, the thrust of the research effort turned from the investigation of program behavior to an investigation of the structure of the input stream. It was felt that in order to be able to model the behavior of programs under execution

in a more detailed manner than had been possible so far, it would be necessary to have more information on the structure of the probabilistic grammar underlying the language in which the executing program was expressed. This grammar is in general not known to the processor, and our attention therefore turned to the problem of grammatical inference or, as it came to be called, linguistic abduction. A number of reports, some written earlier, were communicated in connection with this research contract, in order to define the new problem area and put it into perspective.

9. "Syntax-Controlled Probabilities" by Ulf Grenander, 29 pp.

This report, although unpublished, has become widely known and influential. Its first version was issued in 1967 and the syntax-controlled probability model there expounded developed out of work first reported in the paper "Can we look inside an unreliable automaton?" by Ulf Grenander, in the volume "Research Papers in Statistics - Festschrift for Jerzy Neymann", John Wiley and Sons, 1966, pp. 107-124. In this latter paper, certain probabilistic grammarswere studied from the point of view of inference. In the report, finite-state probabilistic languages are investigated in detail. The model defines a probability distribution over the set of phrase markers, and an induced probability distribution over the set of grammatical sentences. This made it possible to discuss syntactic analysis when the grammar was ambiguous, from the point of view of maximizing the probability of a correct analysis. The fundamental problem how to organize inference algorithms, given empirical data and

assuming some logical structure, in order to determine the remaining structure was posed, and this led to the work on linguistic abduction to which we now turn.

10. "Abduction Machines that Learn Syntactic Patterns" by Ulf Grenander, 27 pp., 1976.

This report - which followed a working paper entitled 'Linguistic Abduction Machines" - used the syntax-controlled probability model (report #9 above) for analysis and computational experimentation. The model can be described as follows: with a lexicon of words a,b,c,... context free rules of the form "v → string" are postulated, where v is one of the syntactic variables and 'string' is made up of words and variables. For a given variable v, a probability distribution over the rules is assumed which rewrites that symbol. It was shown what properties those probability distributions must have in order that the probability measure induced over the set of all finite word-strings be indeed a proper one in the sense that this set have total probability one. The grammars discussed were assumed to possess these properties and called syntax-controlled probability grammars.

In this report, then, various "natural" algorithms were discussed for learning the grammar of a language from its realizations. This was called the abduction problem, following C.S. Pierce's terminology: abduction is distinguished from inference in that it creates successive plausible hypotheses, whereas inference leads to a single decision on acceptance or

rejection of a given hypothesis. It was conjectured that abduction mirrored more closely the process a child goes through when learning the grammar of its mother-tongue, and we wished our algorithms to be "natural" in this sense. A full account of the work on linguistic abduction will be published as a technical report during 1977.

11. "Network Pattern Processors" by Ulf Grenander, 36 pp, 1976.

This report represented a preliminary version of Chapter 7 (entitled "Pattern Processors for Language Abduction") of the book "Lectures in Pattern Theory, Volume II: Pattern Analysis" by Ulf Grenander, to be published early in 1978 by Springer-Verlag. The general problem studied is that of a network which can modify itself in order to learn some of the patterns appearing in its environment; in particular, the emphasis is on the explicit generation of plausible hypotheses (i.e., abduction) when the patterns come from some formal language. The problem is posed in terms of Grenander's pattern theory formalism: the speaker Ω lives in an environment (a "microworld") characterized by some image algebra $\text{env}(\Omega)$; a given image $I \in \text{env}(\Omega)$ can give rise to many different sentences belonging to a language $L(g)$ described by a grammar g. This means that an image processor maps the image algebra $\text{env}(\Omega)$ into another image algebra, so that an image operator takes microworld images into language images. The report then describes the abduction process, in which sentences from $L(g)$ are subjected to "deformation", the deformed sentences are presented to a teacher (or "oracle") who rules on their

grammaticality, and an algorithm is developed which converges to a limiting grammar weakly equivalent to g and with performance parameters which characterize the probability distribution over L(g).

In this report, abduction of the language L(g) is studied when no semantic input from the image algebra env(Ω) is available; the problem of abduction when such input is available is challenging and will be studied.

12. "Application of pattern theory to problems in computer science" Technical Report, October 1976, 47 pp.

This report summarizes Grenander's pattern theory framework, and expresses several problems in computer science in terms of that framework. In particular, the following problems are discussed in this context: syntax and semantics of programming languages, analysis of algorithms, the specification problem, program verification, program synthesis, semantics and language definition, abstraction. Also, the applications to linguistics and abduction are summarized. The essence of this report has been submitted for publication.

DISTRIBUTION LIST

Defense Documentation Center Cameron Station Alexandria, VA 22314	12 copies
Office of Naval Research Information Systems Program Code 437 Arlington, VA 22217	2 copies
Office of Naval Research Code 1021P Arlington, VA 22217	6 copies
Office of Naval Research Code 200 Arlington, VA 22217	1 copy
Office of Naval Research Code 455 Arlington, VA 22217	1 copy
Office of Naval Research Code 458 Arlington, VA 22217	1 copy
Office of Naval Research Branch Office, Boston 495 Summer Street Boston, MA 02210	1 copy

Office of Naval Research
Branch Office, Chicago
536 South Clark Street
Chicago, IL 60605 — 1 copy

Office of Naval Research
Branch Office, Pasadena
1030 East Green Street
Pasadena, CA 91106 — 1 copy

Office of Naval Research
New York Area Office
715 Broadway - 5th Floor
New York, NY 10003 — 1 copy

Naval Research Laboratory
Technical Information Division, Code 2627
Washington, D.C. 20375 — 6 copies

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps (Code RD-1)
Washington, D.C. 20380 — 1 copy

Naval Ocean Systems Center
Advanced Software Technology Division
Code 5200
San Diego, CA 92152 — 1 copy

Mr. E. H. Gleissner
Naval Ship Research & Development Center
Computation and Mathematics Department
Bethesda, MD 20084 — 1 copy

Captain Grace M. Hopper
NAICOM/MIS Planning Branch (OP-916D)
Office of Chief of Naval Operations
Washington, D.C. 20350 — 1 copy

Mr. Kin B. Thompson
Technical Director
Information Systems Division (OP-942)
Office of Chief of Naval Operations
Washington, D.C. 20350 — 1 copy

REPORT DOCUMENTATION PA		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Unnumbered	2. GOVT ACCESSION NO	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Statistical Computer Performance Evaluation		5. TYPE OF REPORT & PERIOD COVERED Final, 1/1/73-12/31/76 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Walter Freiberger		8. CONTRACT OR GRANT NUMBER(s) N14-67-A-0191-0026 N14-75-C-0461
9. PERFORMING ORGANIZATION NAME AND ADDRESS Division of Applied Mathematics Brown University Providence, R. I. 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Mathematical & Computer Sciences Div. Arlington, VA 22221		12. REPORT DATE 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research 495 Summer Street Boston, Mass. 02210		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Reports Distribution List DISTRIBUTION STATEMENT A Approved for public release; Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Computer Performance Evaluation, Statistics, Grammatical Inference, Pattern Theory		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) An annotated bibliography of all reports published under these contracts, with supplementary remarks.		

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)