Department of Computer Science Newsletter

no

10 May 1976 ?

yes

PRINTOUT

Volume 2

Number 3

PRINTOUT, the newsletter of the Department of Computer Science of the University of Maryland at College Park, is published sporadically and distributed to faculty, staff, and students in the Department. Opinions expressed in signed articles may be those of the author, but no opinions represent the policy of the Department, or of the College Park Campus, or of the University.

Contributions may be submitted to the editor, and unless they are obscene or seditious they will probably be used, but minor editing may be done. Complaints directed to the newsletter will be investigated and publicized when possible. It is well to keep in mind however that the Department is subordinate to higher levels of administration, not the other way around; and, the Department does not provide computing service to the campus. Complaints in these areas are best directed to other publications.

STAFF

EDITOR

DICK HAMLET

PRINTOUT

Volume 2
Number 3

CONTENTS

LABORATORY FOR PATTERN ANALYSIS

1. INTRODUCTION

In 1974, with the blessing of Dr. Joseph Marchello, MPSE Division Provost, and grants from the Control and Automation Program, Engineering Division, National Science Foundation, and the Mathematics and Information Sciences Branch of the Air Force Office of Scientific Research, Professor Laveen Kanal and a group of students, with the assistance of Professor Mills, began the job of transforming a portion of the fourth floor laboratory room into the Laboratory for Pattern Analysis.

A research equipment grant from NSF led to the acquisition of a PDP 11/45 computer and a GT40 refresh display terminal. This month a new grant from the National Institutes of Health was awarded which will permit the purchase of additional peripheral equipment for the PDP 11/45. Together with peripherals to be purchased by the Department of Computer Science, and the UNIX operating system from Bell Telephone Laboratories, the research equipment for the Pattern Analysis Laboratory is finally taking the shape envisioned for it. This report describes the Laboratory, and provides some of the motivation for research in pattern analysis.

Because the pattern recognition courses are not part of the required set of courses, the first question might well be, "What is pattern analysis?" The term may be explained in the context of society's willingness to pump ever-increasing amounts of money into developing methods of data collection without first figuring out what to do with all the masses of raw data gathered. Substantial resources are being devoted to data gathering techniques with the hope of understanding complex environments in various fields of commerce, industry, medicine, defense, space, and the physical and social sciences. Computerized acquisition of data allows more data to be acquired faster than ever before. Timely use of the increasing quantities and types of data being gathered can only be made if meaningful patterns are inferred and extracted from the data, the data are categorized, and the underlying processes generating the data are modeled with succinct descriptive or generative models.

2. PATTERN ANALYSIS AND RECOGNITION

Pattern analysis consists of using whatever is known about the specific problem at hand to guide the gathering of data about the patterns and pattern classes which may exist in the environment, and then subjecting the data to a variety of procedures for inferring deterministic and probabilistic structures that are present in the data analysis. Histogram plots, scatter plots, cluster analysis routines, regression analysis, analysis of variance, discriminant analysis, nonlinear mappings, are examples of procedures used to detect and identify structures and substructures in the data. The purpose is to understand the regularities and peculiarities of a data base.

Pattern recognition covers a growing body of theories and techniques for the development of automatic and interactive systems. An early movivation for work on automatic pattern recognition was to model recognition and intelligence as found in living systems. This led to work on biologically motivated automata, neural models and "adaptive," "self-organizing," and "learning" networks. Much of the work on machine recognition of patterns has not been biologically motivated but has adopted one of the other of two models, the feature extraction-classification model, or the linguistic model. Figure 1 shows the general feature-extraction-classification model. Table 1 lists some of the many applications of pattern recognition which have been tried using the feature extraction-classification model.

Some computer scientists criticized the feature-extraction-classification model for performing too severe data compression, since it provided only the class designation of a pattern rather than a description that would allow one to generate patterns belonging to a class. They put forth proposals for a linguistic model for pattern description whereby patterns are viewed as sentences in a language defined by a formal grammar. By 1968 these proposals together with the success of syntax-directed compilers had attracted many to research in pattern grammars. The linguistic or syntactic model for pattern recognitions uses a "primitive extractor," which transforms the input data into a string of symbols or some general relational structure. The primitive extractor may itself be a feature extractor classifier. Then a structural pattern analyzer uses a formal grammar to parse the string and thus constructs a description of the pattern.

In the past, much has beem made of the apparent difference between the two models. The stress on the distinction between the two models hides many similarities: in practice, in the syntactic model, the extraction of "primitives" can involve statistical classification procedures, and the association of patterns with generative grammars is equivalent to the classification of patterns into categories.

Much of the literature on structural pattern recognition deals with formal string grammars and their multidimensional generalizations. While one of the reasons for introducing linguistic methods was the limited relationships handled in statistical pattern classification, phrase-structure string grammars are also severely limited in the relationships they model. Basically, they deal only with concatenation of primitives and immediate constitutent structure.

The definition of the formal linguistic model can be enlarged to include other relationships and other familiar generative mechanisms, such as differential equations and finite-state Markov chains. When a formal model is not explicitly present, the terms "ad-hoc" or "heuristic" are used. The phase "structural pattern recognition" refers to all pattern recognition approaches based on defining primitives and identifying allowable structures in terms of relationships among primitives and substructures that combine primitives. This term represents less a specific set of procedures than an attitude, i.e., that pattern recognition algorithms should be based on the mechanisms that generate and deform patterns.

3. RESEARCH AT THE LABORATORY FOR PATTERN ANALYSIS

As is evident from Table 1, the methodologies of pattern analysis and recognition provide general approaches and specific tools which if appropriately modified and extended, can be applied to a large variety of practical problems. At the Laboratory the areas of application currently being investigated are: error control in data communications, design procedures for automatic and interactive classification of sensor data, such as waveforms and images, computer aided medical diagnosis, pattern analysis of nuclear magnetic resonance spectroscopy patterns, a general waveform parsing system and its application to biomedical wave-

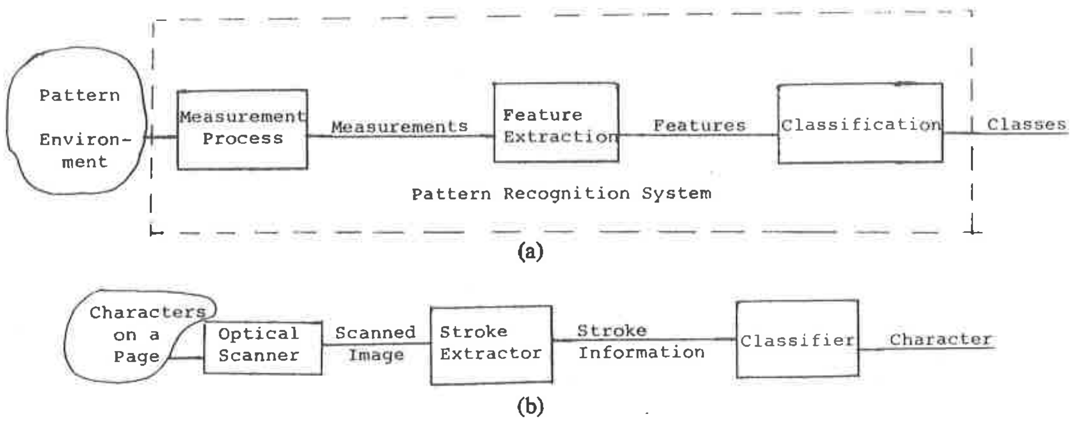


Fig. 1. (a) Operational system. (b) An example.

TABLE I
SOME APPLICATIONS OF PATTERN RECOGNITION

Problem	Input to Pattern Recognition System	Output of Pattern Recognition System
Medical Applications		
Identification and counting of cells	slides of blood samples, microsections of tissue	types of cells
Detection and diagnosis of disease	electrocardiogram waveforms	types of cardiac conditions
	electroencephalogram waveforms	classes of brain conditions
	slides of blood samples	various types and proportions of normal and abnormal cells
Prosthetic control devices	mypotentials	categories of movements of limbs
X-ray diagnosis	X-ray photograph	presence or absence of specific conditions
Military Applications		
Interpretation of aerial reconnaissance imagery	visual, infra-red, radar, multispectral imagery	tanks, personnel carriers, weapons, missile launchers, airfields, campsites
Detection of enemy navy vessels	passive and active sonar waveforms	surface vessels, submarines, whales, fish
Detection of underground nuclear explosions	seismic waveforms	nuclear explosions, conventional explosions, earthquakes
Commercial and Government Applications		
Automatic detection of flaws— impurities in sheet glass, bottles, paper, textiles, printed circuit boards, integrated circuit masks	scanned image (visible on infra-red, etc.)	acceptable vs. unacceptable, markings, bubbles, flaws, radiation patterns, etc.
Classification and identification of fingerprints	scanned image	fingerprint descriptions based on Henry system of classification
Traffic pattern study	aerial photographs of highways, intersections, bridges, road sensors	automobiles, trucks, motorcycles, etc., to determine the characteristics of the traffic flow
Natural resource identification	multispectral imagery	terrain forms, agricultural land, bodies of water, forests
Identification of crop diseases	multispectral imagery	normal and diseased crops
Economic prediction	time series of economic indicators	economic conditions
Speech recognition— remote manipulation of processes, parcel post sorting, management information systems, voice input to computers	speech waveform	spoken words, phonemes
Weather forecasting	weather data from various land-based, airborne, ocean, and satellite sensors	categories of weather
Object recognition— parts handling, inspection of parts, assembly	scanned image	object types
Character Recognition		
Bank checks	magnetic response waveform, optical scanned image	numeric characters, special symbols
Automatic processing of documents— utility bills, credit card charges, sale and inventory documents	optical scanned image	alphanumeric characters, special symbols
Journal tape reading	optical scanned image	numeric characters, special symbols
Page readers— automatic type setting, input to computers, reading for the blind	optical scanned image	alphanumeric characters, special symbols
Label readers	optical scanned image	alphanumeric characters, special symbols
Address readers	optical scanned image	letters and numerals combined into zip codes, city and state names, and street addresses
Other readers— licence plate readers, telephone traffic counter readers	optical scanned image	alphanumeric characters, special symbols

forms, and descriptions of chemical structure diagrams and displayed mathematical expressions to enable their input into computers in a natural form. Other work supported by the laboratory has included studies by Professor A. K. Agrawala and his students of the performance evaluation of computer systems using techniques of pattern analysis, viz, clustering and unsupervised learning of mixtures of probability density functions. Professor Agrawala was recently awarded a grant in this area by NASA.

Part of the laboratory's efforts are directed to the development of experimental tools for pattern analysis research. MIPACS - The Maryland Interactive Pattern Analysis Classification System, and GITSAP - The Graphical Interactive Time Series Analysis Package are two systems which have been partially implemented and used.

The motivation for MIPACS is brought out by Figure 2, which illustrates the complex nature of the process of designing a pattern recognition system. Jo Ann Parikh, a current Ph.D. student of Professor A. Rosenfeld, recently found MIPACS most useful for her studies of cloud classification.

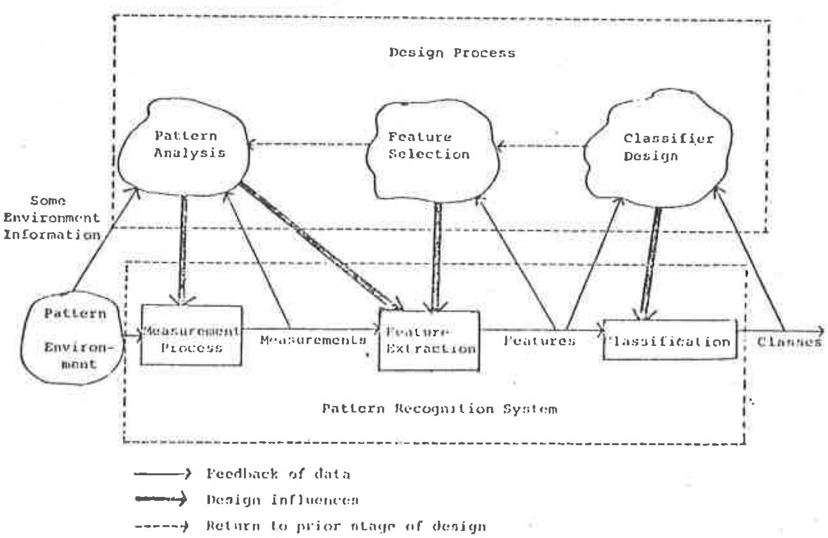


Figure 2. Development of Pattern Recognition Systems

The strategy being followed in the Laboratory for Pattern Analysis is to capitalize on past research and facilitate future research by systematizing techniques in a Laboratory centered around an interactive computer facility. The research on error control and channel modeling described next illustrates this approach.

Transmission through real communication channels does not produce statistically independent errors; rather the errors occur in clusters with the clusters often being separated by long sequences of error-free data. Real channels are therefore said to have "memory". The analysis and modeling of error patterns of digital data communication channels with memory continues to be important, since only a fraction of the performance predicted by information theory is presently relizable on real channels. The Laboratory's research on modeling channels with memory resulted in a systematic approach for establishing transmission policy by a sampling → modeling → evaluation → code prescription sequence, and demonstrated the feasibility of doing this on an interactive system.

The past and current research in the Laboratory can be summarized under the following categories:

1. Statistical pattern analysis, modeling and classification.
2. Structural pattern analysis.
3. Experimental tools for pattern analysis research.

4. RECENTLY COMPLETED PH.D. DISSERTATIONS

Two Ph.D. dissertations were completely at the Laboratory this semester. The dissertation by A.V.

(TURN to page 13 -------→

On the SEX of processes

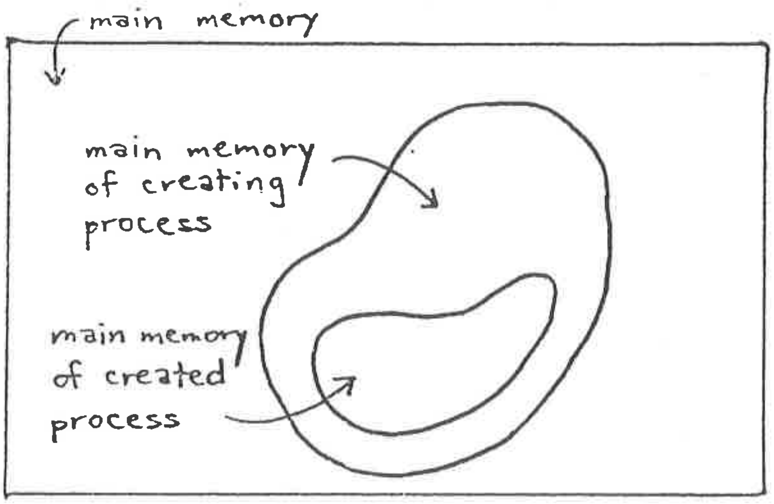
In the RC 4000 Multiprogramming System [1], processes are arranged in a tree along lines of authority (even over life and death) and responsibility. Struck by the resemblance with family trees, many researchers have described the relationships in terms of "father" and "son". The nonsexist terms would be "parent" and "child".

A biological metaphor is certainly useful in talking about processes. As my scientific instincts overcame my political ones, I set out to discover which metaphor is the most accurate -- in other words, what sex, if any, is a process? The following is report on my research.

The Reproductive Characteristics of Processes

Many processes can be active simultaneously (although their apparent parallel activity may only be simulated by timesharing a single processor). Each active process is like a living person, and can call on monitor procedures to extend its capabilities, just as a tool extends the capabilities of the person using it.

When a process wishes to reproduce, it must use a monitor procedure to do so. The new process is given a subset of the resources of the process which created it. The figure below shows what this means for the resource of primary memory.

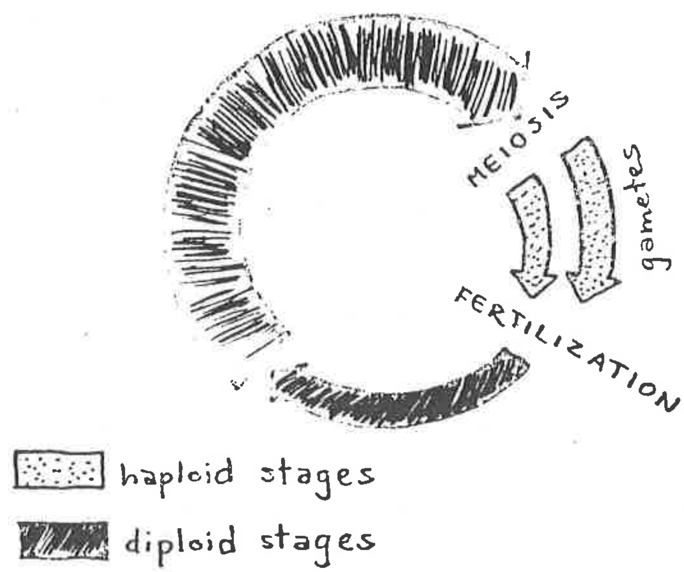


At some later time the process which created the new process activates it. It is now a full-fledged active process, and shares the attention of the physical processor(s) equally with all other processes. However, the creating process has the power to de-activate and destroy its creation.

The Reproductive Characteristics of Multicellular Organisms [2]

Mitosis is the form of cell division in which all genetic material is reproduced, and is thus the sole mechanism of asexual reproduction.

Meiosis is the form of cell division in which a diploid cell (having two sets of chromosomes) divides into two haploid cells (each with one set), and is thus the basis for sexual reproduction. The sexual reproduction cycle as observed in animals is illustrated below*.



* Many plants exhibit a more complex sexual reproductive cycle, which seems to have little to do with processes.

fertilized once by a drone who is disembowled by the act (the system programmer), and then creates progeny of both sexes, through a pupa stage. This model has the particularly attractive feature that the daughters of the operating system are capable of haploid reproduction, but with progeny only useless systems programme rs. The second flaw is that the article is grinding an axe, but achieves only a poor edge. In its haste to conclude that the female terms should be used for processes, it pictures a situation in which the all-powerful male user of the operating system controls his stable of female slaves, occasionally deigning to allow one of them to reproduce when he tires of his existing harem. Thus the real effect is to promulgate the sexist position rather than to correct it.

Besides, as you are well aware, we don't accept papers by women.

MCP

Reference

[1] Root, A. E. The ABC of Bee Culture. Medina, Ohio, 1877.

News

The Science Research Associates Mark Ellson Award to an outstanding undergraduate in the United States has been given to BARBARA TROMBKA, who received her B.S. from the Department in January.

The following Computer Science undergraduates were placed on the MPSE Provost's List for Fall, 1975:

Richard Blankenship	Johnnye Lopes
Deborah Clark	James Menke
Michael Corbett	George Paprotny
Yi Doo	Nancy Pradel
Byorgy Fekete	Jonathan Rosenberg
Daniel Field	William Ruvinsky
William Kennedy	Hsiao-wei Shian
William Krieger	Russel Smith
Howard Larsen	Mark Stega
Marc Libowitz	John Tennyson
Florence Ling	Charles Wolf

Professional News

The Society for Industrial and Applied Mathematics (SIAM) has elected WERNER C. RHEINBOLDT president. His term begins in January, 1977. He has also been appointed Acting Vice President for Publications, and Chairman of the Publications Committee of SIAM.

ROBERT NOONAN will join the faculty of the College of William and Mary in Williamsburg, Virginia in Fall, 1976. VIRGIL GLIGOR from the University of California, Berkeley, will come to Maryland in the Fall.

JACK MINKER has been appointed to the Computer Science Board, an organization of department chairpersons which coordinates university and college computer science activities, including the Computer Science Conference.

The Graduate Record Examinatons Board has announced that a computer science advanced test will be added to the GRE, offered the first time in October, 1976. RICHARD AUSTING is Chairman of the committee developing the examination.

MARVIN ZELKOWITZ has been appointed Newsletter editor of the IEEE Computer Society Technical Committee on Software Engineering. He will be an invited speaker at the Second International Software Engineering Conference in October.

LAVEEN KANAL has been elected to the Board of Governors of the IEEE Information Theory Group for 1976-78. He delivered an invited address, "Hierarchical Classifier Design using MIPACS" at the meeting of the Classification Society of North America at the University of Rochester in May.

Publications, Etc.

Atchison, W.F. "Computer Science Curriculum in Universities " PROCEEDINGS OF THE INTERNATIONAL CONFERENCE ON INFORMATION PROCESSING, Warsaw, Poland, March 22-27, 1976.

Agrawala, A. "On Models for Line Detection" (with L.S. Davis and A. Rosenfeld) IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, Vol. SMC-6, No. 2 (Feb 1976) 127-133.

Agrawala, A. FOUNDATIONS OF MICROPROGRAMMING, Academic Press, ACM Monograph Series, January, 1976, (with T.G. Rauscher).

Agrawala, A. "Developing Application Oriented Computer Architectures on General Purpose Microprogrammable Machines," PROCEEDINGS OF NATIONAL COMPUTER CONFERENCE, May 1976, (with T.G. Rauscher).

Basili, V.R. "Structured Programming: Foundations, Structured Programming Tutorial". INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC., IEEE Catalog No. 75CH1Ø49-6,3-17.

Basili, V.R. "An Example of Program Development, Structured Programming Tutorial." INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, INC., IEEE Catalog No. 75CH1Ø49-6, 176-177.

Basili, V.R. "Language as a Tool for Scientific Programming," PROCEEDINGS THIRD ICASE CONFERENCE ON SCIENTIFIC COMPUTING, April 1976, Academic Press.

Chu, Y. HIGH LEVEL LANGUAGE COMPUTER ARCHITECTURE, Academic Press, Inc. 1975. (Editor and an author)

Gannon, J. D., "An Experiment for the Evaluation of Language Features,", INTERNATIONAL JOURNAL OF MAN-MACHINE STUDIES, Vol. 8, No. 1 (January 1976) 61-73.

Gannon, J. D., "Data types and Programming Relia ility: Some Preliminary Evidence," MRI SYMPOSIUM ON COMPUTER SOFTWARE ENGINEERING, New York, N.Y., April 2Ø-22, 1976.

Hecht, M. "A Simple Algorithm for Global Data Flow Analysis Problems," SIAM J. COMPUTING, Vol. 4, No. 4 (Dec. 1975) 519-532 (with J.D. Ullman).

Kanal, L. "Hybrid Error Control Using Retransmission and Generalized Burst-Trapping Codes," IEEE TRANS. ON COMMUNICATIONS, April 1976 (with A.R.K. Sastry).

Kim, C. "On 3-Head Versus 2-Head Finite Automata," ACTA INFORMATICA 4 (1975) 193-200. (With O.H. Ibarra).

Rheinboldt, W. "On Piecewise Affine Mappings in R^n"(with J. Vandergraft) SIAM J. APPL. MATH. 29 (Dec. 1975) 680-689.

Rheinboldt, W. "On the Solution of Large Sparse Sets of Nonlinear Equations," in COMPUTATIONAL MECHANICS, ed. by J.T. Oden, Lecture Notes in Mathematics, Vol. 461, Springer Verlag, Heidelberg/New York, 1975, 169-194.

Rheinboldt, W. "A Combinatorial Search Process for M-functions (with C. Mesztenyi) BEITR. z. NUM. MATH. 4 (Dec. 1975) 171-178.

Rheinboldt, W. "On Measures of Ill-Conditioning for Nonlinear Equations," MATH. COMP. 30 (Jan. 1976) 104-111.

Rieger, C. "An Organization of Knowledge for a Problem Solver and Language Comprehender," ARTIFICIAL INTELLIGENCE, Vol 7, No. 2 (1976) 125-163.

Rosenfeld, A. "Models for Line Detection" (with L.S. Davis and A.K. Agrawala), IEEE TRANS. ON SYSTEMS, MAN AND CYBERNETICS SMC-6 (1976) 127-133.

Rosenfeld, A. "A Note on Thinning" (with L.S. Davis), IEEE TRANS. ON SYSTEMS, MAN, AND CYBERNETICS SMC-6 (1976) 226-228.

Rosenfeld, A. DIGITAL PICTURE PROCESSING (with A.C. Kak), Academic Press, New York, 1976.

Rosenfeld, A. DIGITAL PICTURE ANALYSIS, Springer Verlag, Berlin, 1976.

VanderBrug, G. J. "Line Detection in Satellite Imagery," IEEE TRANS. ON GEOSCIENCE ELECTRONICS, Vol. GE-14, No. 1 (January 1976).

Vandergraft, J. "On Piecewise Affine Mappings in R^n"(with W.R. Rheinboldt), J. APPLIED MATH., Vol. 29, 680-689.

Vandergraft, J. "A Note on Irreducibility for Linear Operators on Partially Ordered Finite Dimensional Vector Spaces," LIN. ALGEBRA AND APPLICATIONS, Vol. 13, 139-146.

LABORATORY FOR PATTERN ANALYSIS from page 6

Kulkarni, "Optimal and Heuristic Synthesis of Hierarchical Classifiers," developed a systematic approach to the design of probalistic decision trees for multiclass classification, using combinatorial optimization, branch and bound and state-space methods of mathematical and heuristic programming. Most of the theory of statistical classification is devoted to partitioning type classifiers. Kulkarni's research led to new results and insights on the combination of decision tree structures and statistical classification procedures found useful in practice. This work is finding direct application in industry in the design of machines to recognize white blood cells.

John Lemmer's dissertation, "Algorithms for Incompletely Specified Distributions in a General Graph Model for Medical Diagnosis," unifies and extends earlier models proposed in the area of pattern recognition for medical diagnosis. It gives new algorithms for updating probabilistic information about symptoms and diseases on the basis of partially observed and incomplete information and shows how to do this even when the causing events may not be mutually exclusive. John Lemmer's work is being taken up by a group at Rutgers, which is going to exercise the model and the algorithms on a large medical data base. The procedures have many other applications, such as urban planning, and electrical fault diagnosis.

Forthcoming dissertations by Michael Lefler, George Stockman and William Underwood are also expected to have wide applicability. The philosophy guiding the selection of topics for research in the Laboratory is that the best theoretical research often comes out of practical applications and any techniques for practical application should be systematically developed on a sound theoretical foundation.

FedUp

Q: What do you get when you put 8-1/2 piles of PRINTOUT and 3-3/8 piles of PRINTOUT together?

A: True.

(Unsigned) A Friend

Sir:

In regard to your sequence Fishman, Milgram, Cook, Yeh,...,the next five members are Kulkarni, Wilson, McSkimin, Turner, Lemmer. Beyond that no one can say at present.

JM

To the computer people:

Why is it that half the time I can't get any timesharing service on the 1106?

Disgruntled User

Because there is no timesharing service on Univac 1100 machines!-Ed.

Dearly beloved,

What are the guitar chords for "The B220 Doesn't Give a Damn" in PRINTOUT 2, #182?

Peter, Paul, & Moses

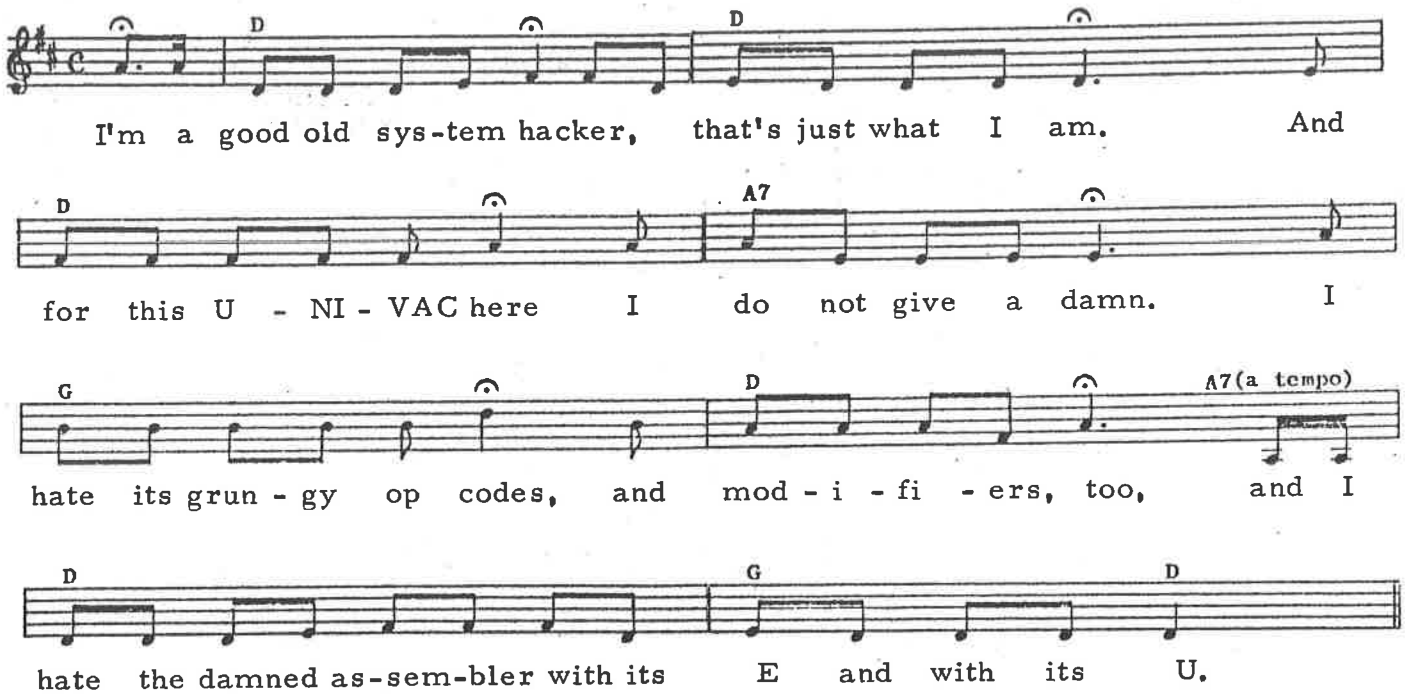
D, A7, and G; sorry about their omission. - Ed.

REFERENCES

Laveen Kanal, "Patterns in Pattern Recognition: 1968 - 1974," IEEE Trans. on Info. Theory, Vol. IT-20, No. 6, pp. 697-722, Nov. 1974.

G.C. Stockman, and L.N. Kanal, "Design Choices in the Construction of the Maryland Interactive Pattern Analysis and Classification System (MIPACS)," Proc. 5th Annual AIPR Symposium on Imagery Pattern Recognition, Univ. of Assoc. AIPR Committee, April 1975.

THE REBEL SYSTEM HACKER



I hate this UNIVAC here, with its software built of trash.
And I hate the damn executive, 'cause all it does is crash.
I hate half-duplex "service," "timesharing" with a catch:
The "commands" are just control cards: leftovers from batch.

I hate the Center system staff; they never fix a glitch.
And I hate the campus users' group, 'cause all they do is bitch.
And I don't want explanations about why it failed again:
It's obvious what's wrong with it, it's not a SYSTEM10.

I hate this UNIVAC here, each library and PROC.
Code is sprawlling everywhere, it's nothing but a crock.
All written in assembler, unfit for human eyes,
And the documents describing it are riddled through with lies.

When things go past a certain point, fixes come too late.
But system hackers never die; we lie low and wait.
A new machine has got to come, and we will hack again...
Unless...Oh no! It's UNIVAC, the damned 1110!

-- Dick Hamlet