Chemometrics and intelligent laboratory systems

An International Journal Sponsored by the Chemometrics Society

SCOPE OF THE JOURNAL

Chemometrics and Intelligent Laboratory Systems publishes articles about new developments on laboratory techniques in chemistry and related disciplines which are characterized by the application of statistical and computer methods.

Special attention is given to emerging new technologies and techniques for the building of intelligent laboratory systems, i.e. artificial intelligence and robotics.

The journal deals with the following topics:

chemometrics: The chemical discipline that uses mathematical and statistical methods to design or select optimal procedures and experiments, and to provide maximum chemical information by analyzing chemical data.

A non-exhaustive list of subjects is:

- statistical methods to evaluate performance characteristics of methods
- proficiency of laboratories and interlaboratory comparisons
- calibration models
- information theory
- correlation techniques and time series analysis
- optimization and experimental design
- regression
- transformation techniques
- deconvolution
- factor analysis
- pattern recognition and clustering
- artificial intelligence and expert systems
- process control
- graph theory
- operations research

computerized acquisition, processing and evaluation of data

- processing of instrumental data
- storage and retrieval systems
- computerized and automated analysis for industrial processes and quality control

robotics

developments in statistical theory and mathematics with application to chemistry

intelligent laboratory systems including self optimizing instruments, planned organic synthesis, data banks with interpretative facilities, and in general applications of expert systems and knowledge representation systems in analytical chemistry

application (case studies) of statistical and computational methods to chemical or related data obtained from natural (medical, geochemical, environmental, food science, pharmacological, toxicological, etc.) and industrial systems (including modelling of processes and quality control)

new software to implement the methods described above and problems associated with the use of software (validation of software for instance)

imaging techniques and graphical software applied in chemistry

Chemometrics and Intelligent Laboratory Systems, 5 (1988) 33–35
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Invited Comments

On Gray's Interpretation of the Dendral Project and Programs: Myth or Mythunderstanding?

BRUCE G. BUCHANAN and EDWARD A. FEIGENBAUM *
Knowledge Systems Laboratory, Stanford University, Stanford, CA 94804 (U.S.A.)

JOSHUA LEDERBERG
Rockefeller University, New York, NY 10021 (U.S.A.)

INTRODUCTION

Gray's paper contains a description of the DENDRAL system and some of its subprograms, and opinions of the merits of this project. Previous overviews of the DENDRAL project include Gray's first references [1] and [2]. We feel that Gray has misunderstood several of the motives and achievements of this work and briefly comment on these below.

Gray's paper goes beyond the narrow and specific difficulties he finds with the DENDRAL programs. He broadens the scope to cast a shadow on the underlying symbolic-processing technology and methodology ("AI").

Since Gray has the perspective of a chemist, not a computer technologist, we feel it necessary to address this broader question as well.

EXAMPLES OF MISUNDERSTANDINGS

The first question is whether Gray has reported accurately on the DENDRAL programs. Because the DENDRAL work has been documented widely in the literature (see references in Gray's paper and in [1], it is neither useful nor necessary to do a thorough accounting in this brief note, but we present below two examples of misunderstandings:

1. The central element and conceptual foundation of the entire set of programs is the DENDRAL generator (later renamed CONGEN). Gray describes the generator in his Fig. 1 as a "heuristic filter" that checks for "some spectral evidence for each radical when building up structure". This is fundamentally wrong: the generator systematically enumerates possible structures within constraints.
2. Gray states that the Planner is part of Meta-DENDRAL (our learning program). In fact, it is the Plan part of the basic plan–generate–test method for producing solutions for structure elucidation problems, and has nothing to do with rule learning.

GOALS OF THE DENDRAL PROJECT AND GRAY'S MISINTERPRETATION OF THEM

1. Certainly intellectual assistance to chemists in doing the difficult and often error-prone task

0169-7439/88/$03.50

of systematic structure elucidation was one of our goals. We did not regard structure elucidation as a task that was trivial and uninteresting for chemists, as Gray seems to believe it is.

2. Notwithstanding this, the DENDRAL work was motivated by the computer science question of how computer programs could be designed and built to assist with hypothesis formation in science. Chemistry and mass spectrometry were secondary to this question, although we also wanted to demonstrate the power of the design by building high performance programs. To illustrate this, when we published the major retrospective of the first decade and a half of this research in book form (Gray's first reference [1]), it was published in the McGraw-Hill Advanced Computer Science Series.
3. All work on expert systems, including our own in DENDRAL and many other projects, is motivated by the intent to provide assistance to human intellectual endeavor, not to provide replacements for skilled human problem solvers. It is unreasonable to assume, as Gray seems to, that Carl Djerassi and Joshua Lederberg believed structure elucidation chemists ignored all empirical data except mass spectra. The DENDRAL programs focused on mass spectra as exemplary of a source of information about structure, and in fact we explored uses of other sources as well.
4. University research projects almost never produce "industrial strength" products. They aim to pioneer proof-of-concept prototypes. The effort and expense to productize is an order-of-magnitude more than to do the basic research. In the usual way, the DENDRAL programs were licensed by Stanford University to an industrial firm for this productizing effort, but the firm (for whatever reason) never made the necessary efforts at product engineering, market education, or marketing. Hence market acceptance of DENDRAL was low. However, the firm used portions of DENDRAL in its other products, and indeed hired the three key young computational chemists of the DENDRAL Project (Smith, Nourse, and Carhart), one as Director of Research. The lack of market acceptance was not due, contrary to Gray's presumption, to lack of performance: DENDRAL's performance as a structure elucidation tool was demonstrated repeatedly to be high because of its systematic nature. The plan–generate–test paradigm used in DENDRAL was indeed adequate for solving complex hypothesis formation problems.
5. Gray does not appear to understand what kind of code is needed in practical expert systems. Gray remarks on how little "AI" code there is in DENDRAL (how does he count it? lines? concepts?), but apparently doesn't know that practical expert systems contain major chunks of code for "system", user interface handling, and routine symbol manipulation. The rule of thumb is: well over half.

CONCLUSION: THE CONTRIBUTIONS OF DENDRAL

The DENDRAL Project is regarded as the grandfather of knowledge based systems in AI. Today's commercially important sector of these is called expert systems. There are at least two thousand of these in use today in companies in the U.S.A., Europe, and Japan, and several thousand more prototyped or under development. The technology has become routinized, and the payoff in some cases has been in the tens of millions of dollars per year. This is not the place to tell that story; it is told elsewhere [3].

DENDRAL began at a time when the focus of AI research was on general problem solving methods (e.g. GPS, theorem-proving, etc.). DENDRAL research was the forcing function behind the so-called "shift to the knowledge-based paradigm" in AI, the paradigm that has since then dominated the field. DENDRAL experiments prior to 1968 led to the knowledge-is-power hypothesis, now called the knowledge principle by AI scientists, that the power of programs to solve complex problems derives not from the power of their reasoning methods but from what they know about their task domains.

The DENDRAL Project scientists were the first to formulate the rule-based representation of knowledge that today dominates the expert sys-

tems field. This representation method led directly to the famous MYCIN program at Stanford, whose clones fill the space of expert systems in use today.

The work on the DENDRAL generator, culminating in CONGEN, was also landmark work. It solved in a completely orderly, logical, and mathematically rigorous way the problem of the systematic generation of chemical structures, with and without constraints imposed by the user; and it offered canonical notational forms for the structures (not only important but necessary for chemistry).

The work on Meta-DENDRAL, the mass-spectral rule learning program that was our first attempt at theory formation from empirical data, revived the research area of machine learning research in the early 1970s. Its "test case" produced mass-spectral fragmentation rules interesting enough to be published by a major journal of the chemical literature. Much more important, its concepts led directly to the Version Space approach to the search problem in machine learning and through that path to the approach called Explanation-Based Learning.

Thus the DENDRAL legacy is rich in contributions to the stream of AI science and technology and to computational chemistry. Gray's paper tends to obscure that legacy and we wish to correct the misunderstandings arising from it.

REFERENCES

1. R.K. Lindsay, B.G. Buchanan, E.A. Feigenbaum and J. Lederberg, *Applications of Artificial Intelligence for Organic Chemistry: the DENDRAL Project,* McGraw-Hill, New York, 1980.
2. C. Djerassi, D.H. Smith, C.W. Crandell, N.A.B. Gray, J.G. Nourse and M.R. Lindley, The Dendral Project: computational aids to natural products structure elucidation, *Pure and Applied Chemistry*, 54 (1982) 2425.
3. E. Feigenbaum, P. McCorduck and H.P. Nii, *The Rise of the Expert Company*, Times Books, New York, 1988.

GENERAL INFORMATION

A detailed leaflet *Information for Authors* is available from the publisher, Elsevier Science Publishers B.V., P.O. Box 330, 1000 AH Amsterdam, The Netherlands, upon request. The most important items are given below.

Types of contributions

= *Chemometrics and Intelligent Laboratory Systems* publishes original research papers, short communications, software descriptions and tutorial reports.

The journal publishes papers from all areas of mathematics, including computer science, numerical methods, operations research, probability and statistics. The motivation and results of the papers must be understandable to chemists and chemometricians.

The journal also participates actively in software dissemination through articles on software developments, software descriptions and reviews of software.

Short communications are usually complete descriptions of limited investigations, and should generally not exceed four printed pages.

Editors responsible for tutorial reports are R.G. Brereton and R.E. Dessy. Tutorial reports are published by invitation of the Editors, but may also submitted. They will be refereed in the usual manner.

Submission of papers

Papers should be written in English. Manuscripts (three copies are required) should be submitted to one of the Editors whose addresses are mentioned on page 2 of the cover. Illustrations must also be submitted in triplicate. One set should be in a form ready for reproduction; the other two may be of lower quality.

Manuscript preparation

The manuscript should be typed in double spacing on consecutively numbered pages of uniform size. The title page should include the title, name(s) of author(s), and their affiliations. The author to whom correspondence should be addressed must be indicated. Research papers begin with an abstract (50–250 words) which comprises a brief factual account of the contents of the paper. As a rule, papers should be divided into sections, headed by a caption (e.g., Abstract, Introduction, Experimental, etc.).

References

References should be numbered in the order in which they are cited in the text, and listed in numerical sequence on a separate sheet at the end of the article. The numbers should appear in the text at the appropriate places in square brackets, on the line. The reference list must contain the names of all authors, full titles of articles and full journal names, and first and last pages of each cited paper.

Tables

Tables should be compiled on separate sheets. They must be numbered with Arabic numerals, and have brief descriptive headings. They should be referred to in the text.

Illustrations

All illustrations are to be numbered consecutively using Arabic numerals and should be referred to in the text. Legends must be typed together on a separate sheet. Line drawings should either be drawn in Indian ink or submitted as sharp glossy prints suitable for immediate reproduction. Any lettering must be large enough to stand photographic reduction. Half-tone illustrations should be black and white prints on glossy paper and have as much contrast as possible. Colour illustrations may occasionally be published.

Proofs

One set of proofs will be sent to the author to be checked for printer's errors.

Reprints

Fifty reprints of research papers, short communications and software descriptions will be supplied free of charge. Additional reprints (minimum 100) may be ordered by the authors. The order form containing price quotations will be sent together with the galley proof of the article.

PUBLICATION

Chemometrics and Intelligent Laboratory Systems has two volumes (eight issues) per year. The subscription price for 1988 is Dfl. 532.00 (ca. US$ 288.00) and for 1989 Dfl. 564.00 (ca. US$ 305.00), including postage. Members of the Chemometrics Society are entitled to order a personal subscription to the journal at a special rate. Advertising rates are available from the publisher.

Subscription orders may be sent to:
ELSEVIER SCIENCE PUBLISHERS B.V., P.O. Box 211, 1000 AE Amsterdam, The Netherlands.

ISSN 0169-7439

A FREE SAMPLE COPY OF THE JOURNAL IS AVAILABLE ON REQUEST