LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

CONF-901130--1

LA-UR -90-1548

LA-UR--90-1548

DE90 012097

TITLE A CLINICAL FLOW CYTOMETRY DATA ANALYSIS ASSISTANT

AUTHOR(S) Gary C. Salzman, Ph.D., C. C Stewart, Ph.D., R. E. Duque, M.D., R. C. Braylan, M.D.

SUBMITTED TO IEEE Engineering in Medicine and Biology Society Meeting Philadelphia, PA

DISCLAIMER

MASTER

Los Alamos Los Alamos National Laboratory
Los Alamos, New Mexico 87545

A CLINICAL FLOW CYTOMETRY DATA ANALYSIS ASSISTANT

G.C. Salzman, Ph.D.[1] C.C. Stewart, Ph.D.[2] R E. Duque, M.D.[3] R.C. Braylan, M.D.[4]

Los Alamos Nat. Lab., Los Alamos, NM[1], Roswell Park Cancer Ctr., Buffalo, NY[2]

Norwood Clinic, Birmingham, AL[3]. Univ. Florida Medical Sch., Gainesville, FL[4]

Abstract

A rule-based expert system is being developed to assist clinicians in the analysis of multivariate flow cytometry data for patients with leukemias or lymphomas. The cells are stained with fluorescently labeled monoclonal antibodies and the cell fluorescence is measured with a flow cytometer. Cluster analysis is used to isolate subpopulations in the data on which the clinical decisions are made. Symbolic facts for the expert system are instantiated using these numerical data and the knowledge of the clinicians and experts in flow cytometry. The first prototype used a decision tree and rigid rules. It successfully classified only nine of eleven leukemia cases. A second prototype incorporating certainty factors into the rules is now being developed that should remove the need for a rigid decision tree.

INTRODUCTION

Flow cytometry has become an important research tool in the clinical laboratory [1-3]. It is being used a a variety of applications ranging from immunophenotyping to the quantitation of DNA on a single cell basis. In a flow cytometer, cells in suspension pass in single file through a focused laser beam where fluorescent tags on the cells are excited. For each cell the amount of fluorescence is measured along with the amount of light scattered at several angles. The set of data for each cell is a vector in a multidimensional space and is referred to as an event. The elements of each vector are called variates. These data are typically displayed as a series of bivariate dot plots, where each dot represents the measurements for one cell. The dots for each cell subpopulation are usually grouped closely together on each bivariate plot. An experienced clinician who uses flow cytometry regularly can easily interpret these patterns and arrive at a diagnosis. Less-experienced clinicians or those who uses flow cytometry infrequently may find interpreting the data a more difficult task. The expert system being developed is intended to serve as an intelligent assistant to these clinicians.

Immunophenotyping

Immunophenotyping enables an investigator to use fluorescence conjugated antibody reagents to discriminate among subsets of lymphocytes. The antibody reagents bind with high specificity to membrane-associated molecules that serve to identify the cells. The reagent anti-CD3, for example, labels all T cells. One or several of these reagents can be incubated with a cell sample and then analyzed with a flow cytometer.

Cluster Analysis

Cluster analysis [4-5] is a mathematical technique for grouping together events that are close to each other in a multidimensional space. K-means clustering is used for the flow cytometry data in this work. Each event is assigned to one of K different means, where the number of means is specified by the investigator. The clusters are identified by means and standard deviations for each variate.

Symbolic facts needed for the expert system rules are extracted from the numerical output of the cluster analysis. The clinicians typically interpret a cluster as being "negative", "dim" or "bright" for some immunophenotyping marker. The cluster means are translated into these terms to create a set of symbolic facts that form part of the expert system knowledge base for each sample analyzed. For example, (panel B1 cluster 3 CD10 negative) is a fact that says "for the monoclonal antibody panel B1, which measures the marker CD10, cluster number 3 is negative for CD10."

Expert System Tools

The CLIPS (C Language Integrated Production System)[6] developed by NASA is used as the expert system development tool. CLIPS compares the conditions of a rule against facts in the knowledge base. If all of the required conditions match for a rule, it "fires" and places appropriate new facts in the knowledge base. An example is

```
(defrule rule4 "Evaluate CD10 fluorescence"
   ?node <- (current-node node 6)
   (panel B1 cluster ?clus FSC low)
   (panel B1 cluster ?clus SSC low)
   (panel B1 cluster ?clus CD10 negative)
=>
   (retract ?node)
   (assert (current-node node 9)))
```

The variable identifying a node in a decision tree is called ?node. Similarly, ?clus is a variable identifying some cluster number. If the current

position in a decision tree is node 6, there will be a fact in the knowledge base (current-node node 6). If there are three other facts in the knowledge base (panel B1 cluster 3 FSC low), (panel B1 cluster 3 SSC low), and (panel B1 cluster 3 CD10 negative) then the rule above will fire. Its action will be to remove the fact from the knowledge base (current-node node 6) and replace it with the fact (current-node node 9).

Problems with Decision Trees

The first prototype used a decision tree based on the methods used by the clinicians in analyzing the data by hand. Rules such as the one above were used to direct the branching at each node in the decision tree. The results from this prototype expert system were very sensitive to the numerical values used to set the thresholds between fluorescence values called "negative", "dim" and "bright." As a result classification errors were unacceptable. Clinicians familiar with flow cytometry data analysis had little difficulty in defining these boundaries because they could rely on their past experience.

Discussion

Because of the problems encountered with decision trees, a second prototype now under development incorporates uncertainty into the rules in the form of certainty factors[7-8]. A strict decision tree is then no longer needed. The rules can fire opportunistically and place facts in the knowledge base that lead toward a recommended diagnosis. This approach has been used by other expert system developers [9]. The certainty factors for the rule conditions and conclusions are determined based on case histories.

References

1. Howard M. Shapiro. Practical Flow Cytometry, 2nd Ed., Alan Liss, NY, 1988.
2. M.A. Van Dilla, P.N. Dean, O.D. Laerum, M.R. Melamed (eds). Flow Cytometry: Instrumentation and Data Analysis. Academic Press, NY, 1985.
3. Andrew Yen. Flow Cytometry: Advanced Research and Clinical Applications. Vols I and II. CRC Press, Boca Raton, FL, 1989.
4. H. Spath, Cluster dissection and analysis, theory, FORTRAN programs and examples. Halsted Press, John Wiley and Sons, NY (1985).
5. J.A. Hartigan, Cluster algorithms. John Wiley and Sons, NY, 1975.
6. J. Giarratano and G. Riley, Expert Systems, Principles and Programming. PWS-KENT Publishing Co., Boston, 1989.
7. E.H. Shortliffe and B.G. Buchanan, "A model of inexact reasoning in medicine," Mathematical Biosciences 23, 1975.
8. E. Rich, Artificial Intelligence, McGraw-Hill, NY, 1983.
9. J.E. Weber, P.H. Bartels, W. Griswold, W. Kuhn, S.H. Puplanus, A.R. Graham. Colonic Lesion Expert System. Performance Evaluation. Analytical and Quantitative Cytology 10, 150-159, 1988.