76-3

A MULTI-STEP FORMATION OF VARIABLE VALUED LOGIC HYPOTHESES

James Larson
Department of Computer Science
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801

76-3

A multi-step method for forming variable valued logic hypotheses is given. An experiment applying this method to a problem in medical diagnosis is described. The results are compared to those of a direct (one-step) method and to a classical Bayesian approach.

Introduction

The understanding of many phenomena depends on a continuing process of forming hypotheses which explain various observations and correcting or expanding them to be consistent with new observations. Although the actual process which human beings use to formulate and modify hypotheses and to make decisions is not well understood, it can be approximated by applying certain formal rules of inference to specific facts and observations in order to form generalized descriptions of these facts and observations. This paper contains a brief description of a computer implementation of such a process using the variable valued logic system VL_1. Some results of the application of this method to a pattern recognition problem in medical diagnosis are included to demonstrate the feasibility of the method.

The VL_1 system has been applied to other pattern recognition problems [Michalski[5]] with promising results. It has been shown that simple, easily understood descriptions can be inferred from sets of objects and used to classify new objects accurately. A computer system which uses such descriptions as a basis for giving advice must be able to generate and update its descriptions as new observations are made. One of the methods for maintaining these descriptions is described here (see [Cuneo[4]] for another approach). This method uses the hypotheses which are consistent with observations and facts which are available whenever possible and returns to the observations and facts to correct hypotheses which are inconsistent.

Hypothesis Generation and Updating

All facts and events (observations) are expressed in the variable valued logic system VL_1. The system is described in detail in [Michalski[1]]; here we will only briefly review some of the most basic concepts. The language provides a formal and flexible medium for expression which can be easily applied to real phenomena. Descriptions in this language are formal to make application of rules of inference possible. The descriptions use a very natural set of operators to relate expressions in the language to real situations. Each variable x_i and formula $\vee$ is assigned its own domain (sizes d_i and d, respectively) from which it can draw its values. The domain for different variables may vary in size and structure to reflect the properties of the associated variables. Expressions in the language may be composed of variables, symbols and a set of operators ($\&$, $\vee$, $\curlyvee$). An example of an expression in the language is the following:

$$3[x_1 = 1{:}3]\ [x_2 = 2,6] \vee 2[x_4 = 1] \curlyvee [x_3 = 0]$$

An expression in the VL_1 can assume values from a set $\{0,1,2,\ldots d\}$ where each value may be interpreted as a degree of certainty that the formula is true or degree of certainty that objects belong to a certain object set. The above expression takes the value or degree of certainty 3 when the variable x_1 has the values 1 through 3 and x_2 has the values 2 or 6 or if this is not true, then the expression assumes the value 2 when x_4 equals 1 except when x_3 has the value 0. If neither of the above expressions is satisfied, the entire expression has the value 0. The operator '$\vee$' takes the maximum of the two operands so if both terms are satisfied, the expression taks the value 3.

Descriptions can be inferred from sets of events and facts as follows [Michalski[2]]: Suppose events (i.e., representations of objects) are represented in the event space $E = (2,3,3)$ (i.e., E is a cartesian product of three variables where variable x_1 may assume 2 values, and variables x_2 and x_3 may assume three values). Let e_0 and e_1 be sample events from the sets of events F^0 and F^1, respectively.

$$e_0 = (0,1,1) \qquad e_1 = (0,2,0)$$

Then the expressions $(x_3 = 1)$ and $(x_2 = 2)$ are plausible discriminant descriptions of the sets F^0 and F^1 based on the given events e_0 and e_1. Suppose now that a second event e_0' of the set F^0 is given:

$$e_0' = (0,2,1)$$

Then the above descriptions are not consistent with the available facts, namely the event e_0' of set F^0 satisfies the description for the set F^1. The expression $[x_3 = 1]$ describing F^0 is correct, but the description for F^1 must be modified to form $[x_2 = 2] \curlyvee [x_3 = 1]$ or a simpler expression $[x_3 = 0]$.

A few programs have been developed which infer VL/1 expressions of various types from facts or from other VL_1 expressions (denoted AQVAL/1 programs). Among these programs is UNICLASS which generates the description of a set of events against its complement [Yuen[8]]. AQ7 is the original and currently the most flexible program which infers discriminant expressions from sets of events which are near optimal with respect to a user specified criteria list [Larson, Michalski[7]]. AQ9 infers discriminant descriptions from sets of events or formulas or both [Cuneo[4]]. SYM1 infers VL_1 expressions which are composed of symmetric selectors (i.e. selectors of the form $(x_1+x_2=3)$ or $(x_1+x_4'=2,3)$ where + is the arithmetic sum and x_4' means x_4 complement defined as d_4-x_4). A new program denoted AQ11 will be described here.

The program AQ11 accepts disjoint sets of events and facts along with sets of hypotheses where each hypothesis is a description in VL_1 of one event set (at least in part). The hypotheses are first corrected in order to resolve all inconcistencies (i.e. to insure that a hypothesis for one set doesn't explain or 'cover' an event from another set). Then new formulas are formed based on the given events and the new corrected hypotheses. Finally, an estimate is made of the degree to which each hypothesis is expected to describe the associated event set.

The process of correcting generalized VL_1 expressions requires that some expressions be reduced to cover or describe fewer facts and that some of the expressions be expanded to cover new facts not covered before. Each of these operations can be carried out by extending the expression using the operators already introduced (namely &, $\vee$, $\curlyvee$). The first type of correction can be realized by appending to each formula a description of the events which are inconsistent with the formula. That is, if the events e_1 and e_2 of the set F^1 are covered by a formula V^0 which is supposed to cover only known events of F^0, then one can form a description EX^0 of the events e_1 and e_2 against the events of F^0. Appending this description of the exception events to the expression V^0 gives $V^0 \curlyvee EX^0$. (A cover or description of an event set E^1 against a set of events or formulas E^0 denoted $C(E^1|E^0)$ is an expression which describes all the events of E^1 but none of the events or formulas in E^0.) Application of this process to each formula and each set of events gives a consistent set of formulas (i.e. each formula $V^i \curlyvee EX^i$ covers events of only one event set F^i and no others).

Each of these consistent expressions can be expanded to cover events which were not covered earlier (but should now be) by appending a description of events which were not covered (using the $\vee$ operators). Let us denote these new expressions $\hat{V} = (\hat{V}^i)$ where $\hat{V}^i$ covers completely the event set F^i but no other event sets. This set of formulas is a complete description of the available facts. We now search for a simpler set of complete expressions.

Since each step of the correction process has thus far increased the complexity of the expressions, some refinement of the formulas is now in order. It is evident that the area of generalization of the original formulas was in error (or else no correction would have been necessary). The formulas have been corrected at the expense of increased complexity of the descriptions. Because the set of formulas is now consistent, some refinement of the formulas can be performed at little expense. We can form new descriptions which are once again simple and correct by covering each complete set of events F^i against the corrected set of formulas (except the formulas for F^i) quite economically.

This completes the formula correction phases of the hypothesis correction procedure. It may be helpful to now study a simple example of this process. Consider the event space E = (3,4,2) (a cartesian product of the domains of 3 variables x_1, x_2, x_3 having 3, 4, and 2 elements respectively, where the elements of each domain are not ordered). Assume that two sets of facts (F^1 and F^2) are known and for each set (i=1,2), an hypothesis V^i is given.

Set 1	Set 2
F^1: $e_{1,1} = (1,0,0)$	F^2: $e_{2,1} = (2,0,0)$
$e_{1,2} = (1,2,0)$	$e_{2,2} = (2,0,1)$
$e_{1,3} = (1,3,0)$	$e_{2,3} = (2,2,1)$
$e_{1,4} = (0,0,0)$	$e_{2,4} = (0,0,1)$
$e_{1,5} = (0,2,0)$	$e_{2,5} = (1,3,1)$
V^1: $[x_1 = 0,1]$	V^2: $[x_1 = 2]$

The formulas V^1 and V^2 may have been formed at a time when only the first three facts of each set F^1 and F^2 were known. Now, two new facts have been added to each set and the formulas are no longer correct. The formula V^1 still describes all events in the set F^1 but also covers two events in the set F^2. The formula V^2 describes only three events in the set F^2. The formulas could be corrected in several ways. The original hypotheses could be discarded and new formulas created based on the facts alone. A procedure for doing this is given by Michalski in [2]. This could be called a direct method (or one step method) and can be quite time consuming if the sets of facts are large (e.g. 100 to 200 events per set). A second approach is to simply modify each hypothesis by appending facts using the appropriate operators ($\curlyvee$ or $\vee$). For this example, the resulting formulas would be:

$$V^1 \curlyvee (e_{2,4} \vee e_{2,5}) \quad \text{and} \quad V^2 \vee e_{2,4} \vee e_{2,5}$$

These expressions are quite complex and fail to generalize the new information in the facts $e_{2,4}$ and $e_{2,5}$.

The approach taken here is to first generalize the exceptions to V^1. Using a VL form of de Morgan's laws, an expression can be formed which covers $e_{2,4}$ but none of the events of F^1. This expression is:

$$([x_1 = 0,2] \vee [x_3 = 1])([x_1 = 0,2] \vee [x_2 = 0,1,3] \vee [x_3 = 1])([x_1 = 0,2] \vee [x_2 = 0,1,2] \vee [x_3 = 1])[x_3 = 1]([x_2 = 0,1] \vee [x_3 = 1])$$

Performing the indicated operations and applying absorption laws, this expression reduces to

$$[x_3 = 1]$$

Since this term also covers $e_{2,5}$, it is a description of all exceptions to V^1. Simpler expressions for the correction to the formulas V^1 and V^2 are:

$$[x_1 = 0,1] \curlyvee [x_3 = 1] \quad \text{and} \quad [x_1 = 2] \vee e_{2,4} \vee e_{2,5}$$

Now, using a similar method, the formulas V^1 and V^2 and be generated

$$\hat{V}^1 = C(F^1|[x_1 = 2] \vee e_{2,4} \vee e_{2,5})$$

$$\hat{V}^2 = C(F^2 | [x_1 = 0,1] \vee [x_3 = 1])$$

or

$$\hat{V}^1 = [x_1 = 0,1][x_3 = 0]$$

$$\hat{V}^2 = [x_1 = 2] \vee [x_3 = 1]$$

These expressions are simple and consistent with respect to the sets to facts F^1 and F^2. It is interesting to now investigate the discriminant capabilities of such formulas.

Formulas which are generated by inductive inference to describe event sets are actually generalizations of specific examples of each set which are simple with respect to the available operators and the specified criteria. These formulas may indeed be very far from an accurate description of the actual events set. There may be, in fact, no simple description using the available set of operators and cost criteria, and no matter how much effort is spent in the formation of simple descriptions, they will not describe the associated event set well. An estimate of the degree to which formulas describe the associated events set is, therefore, crucial to the credibility of the resulting hypothesis. This estimate can be a description of the past performance of the formulas for each event set from previous correction steps, or, as described here, a description of the performance of the set of hypothesis when applied to the problem of classifying a separate set of events (a testing set) of known set membership.

Two criteria describing the degree to which a set of hypotheses is able to classify a set of events are presented here. The first criterion, the percent correction classification for each hypothesis is calculated as follows: For each event in the testing set, the degree to which a formula describes that event is computed [Larson, Michalski[7]]. The event is assigned to the set(s) whose formula(s) comes closest to the event. If one of assigned decisions is the set from which the event was selected, the decision is correct. The percent correction decision is then the percentage of events in the testing set which were given the correct assignment.

An event in the testing set may come close to more than one formula resulting in multiple decisions. This ability to give multiple decisions is most desirable; however, if the testing events are known to reside in only one event set, this behavior should be minimized. Therefore, the second criterion which is used here is the average number of decisions for an event. The combination of a high percent correct and a low number of decisions per event is then the measure of the discriminant capabilities of the formula.

Another measure of the quality of VL_1 expressions is their complexity (the simpler the descriptions are, the easier they are to comprehend). Two such measures of complexity are used here: The number of terms in a formula, and the average number of selectors in each term of the formula. The experiment described below evaluates each of the above criteria to give an indication of the quality of the expressions which are inferred by the program.

Application to Medical Diagnosis

The program described above has been applied to a problem of distinguishing cases of different liver diseases. The data used in the experiment was supplied by Dr. James Croft from the University of Utah [Croft[6]]. This data was originally collected by Professor Klatskin at the Yale University Medical School. Three diseases were selected from the data for this experiment:

Symbol	Diseases Name
(P)	Post Necrotic Cirrhosis
(G)	Granulomata
(F)	Fatty Liver

The total set of cases for each disease was separated into a training set and a testing set as given in Figure 2 (Cols. 3,4). Each case is represented by values of 50 signs and symptoms in Figure 1. A simulation of many applications of this hypothesis correction has been performed in this experiment and the decision making capabilities of the resulting expressions are compared to those of a set of expressions which were inferred from the complete set of training facts directly [Michalski[3]].

The complete training set was partitioned into five sets of events each consisting of 20% of the complete training set. Five sets of hypotheses were inferred in five stages as follows: An initial set of hypotheses was obtained from one of these sets and the decision making capabilities evaluated with respect to the testing set. Then, each new set of training data was merged in turn with the previous training set and the previous hypotheses to form new expressions. Each new formula set was applied to the testing set to evaluate the performance of the formulas (Figures 3, 4) and statistics about the complexity of each new set of expressions were calculated (Figures 5, 6). (At each stage, the hypotheses from the previous stage are used in addition to the set of training events. As an alternative method, new formulas could be generated using only the hypotheses from the previous stage and the events not yet classified by the old hypotheses [Cuneo[4]].) In these figures, the abcissa represents the total amount of training data which is available at one stage. The value for each disease is indicated by the point on the graph with the corresponding symbol (P,F,G). To show the change from stage to stage, the points for each disease are connected with dashed lines. An example of an expression describing Post Necrotic Cirrhosis is the following:

Description of P using 20% of the training data --

$$[x_{17} = 0][x_{34} = 0][x_{42} = 0][x_{48} = 1] \vee$$

$$[x_{22} = 0][x_{24} = 0][x_{25} = 1,2][x_{31} = 0][x_{37} = 0,1]$$
$$[x_{42} = 1] \vee$$

$$[x_{20} = 1][x_{33} = 0,1][x_{41} = 0,1][x_{43} = 0][x_{43} = 1,2]$$

The results were then compared to those formulas which were obtained directly from the complete training set (Figure 2, Cols. 5-8). It is clear from these results that as events are added to the training set, though the complexity of the expressions increases somewhat, the decision making capabilities also increase to form a more unique correct description of each disease. Figure 7 gives the computer time required to form each new set of expressions. Although the time increases slightly as more events are added to the training set, the time for the last stage (15 sec.) is significantly less than the time required by the same program to infer formulas directly

from the training set (without the aid of the previously generated hypotheses -- 1 min., 20 sec.). Therefore, if some hypotheses are available, new corrected expressions can be generated efficiently using this method.

To make a comparison of this new method and a classical pattern recognition method, a parallel experiment was performed using the same data as before but now using Bayesian statistics to distinguish one event set from another. (The Bayesian method has been shown to be as accurate as several common pattern recognition methods [Croft[6]].) The Bayes description for each event set (assuming that all symptoms are independent) is a function of conditional probabilities generated from the training set as follows:

$$G_k(e) = P(d = k|x = e) = P(k) \prod_{i=1}^{n} \frac{P(x_i = e_i|d = k)}{P(e)}$$

where

$e = (e_1, e_2, \ldots, e_n)$ is a particular event or case which is to be classified (e_i value of i^{th} variable or symptom).

$P(d = k|x = e)$ is the probability that an event e will be of the event set k (in this experiment, e is a particular case of a disease k).

$P(k), P(e)$ a priori probability of an occurrence of disease k and event e (assumed here to be 1).

$P(x_i = e_i|d = k)$ probability of observing variable value e_i in an event of set k.

A decision is made by calculating $G_P(e)$, $G_F(e)$ and $G_G(e)$ and selecting the disease with maximum probability. The conditional probability matrix $P(x_i = e_i|d = P)$ using 20% of the training data (the Bayesian counterpart to the VL_1 description above) is given below:

$$P(x_i = j|d = P)$$

Val. (j) — Variable (x_i)

VAL	x 1	x 2	x 3	x 4	x 5	x 6	x 7	x 8	x 9	x10
0	0.19	0.50	0.69	0.38	0.88	0.69	0.50	0.57	0.63	0.69
1	0.44	0.50	0.32	0.63	0.13	0.32	0.50	0.44	0.38	0.32
2	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

VAL	x11	x12	x13	x14	x15	x16	x17	x18	x19	x20
0	0.63	0.88	0.50	1.00	1.00	0.75	0.94	0.38	0.44	0.32
1	0.38	0.13	0.50	0.00	0.00	0.25	0.07	0.63	0.57	0.69
2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

VAL	x21	x22	x23	x24	x25	x26	x27	x28	x29	x30
0	0.25	0.19	0.38	0.94	0.00	0.07	0.50	0.94	0.82	0.19
1	0.75	0.82	0.38	0.07	0.25	0.75	0.32	0.00	0.19	0.82
2	0.00	0.00	0.25	0.00	0.75	0.19	0.19	0.07	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

VAL	x31	x32	x33	x34	x35	x36	x37	x38	x39	x40
0	0.75	0.69	0.13	0.94	0.00	0.19	0.57	0.32	0.25	0.75
1	0.25	0.32	0.32	0.00	0.88	0.82	0.44	0.69	0.57	0.25
2	0.00	0.00	0.57	0.07	0.13	0.00	0.00	0.00	0.19	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

VAL	x41	x42	x43	x44	x45	x46	x47	x48	x49	x50
0	0.75	0.88	0.75	0.75	0.75	0.88	1.00	0.00	0.57	0.69
1	0.19	0.13	0.25	0.13	0.19	0.07	0.00	0.88	0.44	0.13
2	0.07	0.00	0.00	0.13	0.07	0.07	0.00	0.13	0.00	0.00
3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.19

For comparison, this decision process was applied to the same sets of training events used in the five stages above (Figure 8 and Figure 2, Col. 9). The accuracy seems to be equivalent although much more testing is necessary to arrive at a conclusive statement. Some observation about the differences between the Bayesian and VL_1 methods can be made [Michalski[3]].

1. VL_1 expressions are easily translated into English descriptions. The Bayesian description is a matrix of condition probabilities containing an entry for each value of each variable in the event space. Such a matrix is very difficult to interpret.

2. VL_1 expressions involve only a subset of the complete set of variables. This subset contains only the variables necessary to distinguish one object set from all the others. In the expression given for P above, only 13 variables are used in the description the Bayesian analysis requires values for all 50 variables.

3. Evaluation of VL_1 expressions requires only comparison operations where the Bayesian method requires extensive multiplication of floating point numbers.

Conclusion

We have briefly described here one method for maintaining a body of knowledge by using an inferential learning process. The results of the small experiment presented here gives hope to the application of these techniques to a larger scale knowledge acquisition system.

Acknowledgments

The author expresses his deepest and most sincere thanks to Professor R. S. Michalski for his unending patience and valuable advice. Thanks also go to Mr. R. Chilausky for valuable discussions and comments.

The author gratefully acknowledge the National Science Foundation for partial support of this work under Grant No. DCR 74-03514.

References

1. Michalski, R. S., VARIABLE-VALUED LOGIC: System VL_1, Proceedings of the 1974 International Symposium on Multiple-Valued Logic, May 29-31, 1974, Morgantown, West Virginia.

2. Michalski, R. S., Synthesis of Optimal and Quasi-Optimal Variable-Valued Logic Formulas, Proceedings of the 5th International Symposium on Multiple Valued Logic, Bloomington, Indiana, May 13-16, 1975.

3. Michalski, R. S., Problems of Designing an Inferential Medical Consulting System, First Illinois Conference on Medical Information Systems, Urbana, Illinois, Oct. 17-18, 197-.

4. Cuneo, R. P., Selected Problems of Minimization of Variable-Valued Logic Formulas, Master's Thesis, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, May 1975.

5. Michalski, R. S., AQVAL/1--Computer Implementation of a Variable-Valued Logic System VL_1 and Examples of Its Application to Pattern Recognition, Proceedings of the First International Joint Conference on Pattern Recognition, Washington, D.C., October 30 - November 1, 1973.

6. Croft, D. James, Is Computerized Diagnosis Possible?, Computers and Biomedical Research 5, 351-367 (1972).

7. Larson, J., Michalski, R. S., AQVAL/1: User's Guide and Program Description, Department of Computer Science, University of Illinois, Urbana, May 1975.

8. Yuen H., Uniclass Cover Synthesis: User's Guide, Internal Report.

9. Jensen, G., SYM-1: A Program that Detects Symmetry of Variable-Valued Logic Functions, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, 1975.

Symptom number	Symptom	Symptom number	Symptom
1	Age	21	Thymol turbidity
	Under 10	22	SGOT
	10–19	23	Protein
	20–29	24	Albumin
	30–39	25	Globulin
	40–49	26	Alkaline phosphatase
	50–59	27	Necrosis: diff. or focal
	60–69	28	Necrosis: cent. or portal
	70–79	29	Bile thrombi
	80 and over	30	Regeneration: bile ducts
2	Sex	31	Regeneration: retic. endo.
3	Heavy alcoholic intake	32	Regeneration: paren. or mitoses
4	Significant weight loss	33	Degeneration: diff. or focal
5	Nausea and/or vomiting	34	Degeneration: central or portal
6	Abdominal pain	35	Cells: diff. or focal
7	Malnutrition	36	Cells: central or portal
8	Jaundice	37	Cells: polys
9	Ascites	38	Cells: lymphs
10	Edema	40	Cells: eos
11	Abdominal collaterals	41	Cells: plasma or giant
12	Spider nevi	42	Fat: diff. or zonal
13	Splenomegaly	43	Fat: 1–2+ or 3–4+
14	Body temperature	44	Pigment: iron or bile
15	Liver size	45	Pigment: paren. or gen.
16	Liver tenderness	46	Pigment: kupff. or portal
17	Liver nodules	47	Mall. B. or culture
18	One-minute bilirubin	48	Fibrosis: diff. or focal
19	Total bilirubin	49	Fibrosis: portal or cent.
20	Cephalin flocculation	50	Abnormal stool color

Figure 1 Variables used in experiments

1	2	3	4	5	6	7	8	9
Disease	Total # Cases	# Learning Cases	# Test Cases	Correct Classification (% correct, # dec./event)		Complexity (# terms, avg # sel)		Bayesian Model % Correct Decisions
				Direct	Multi-Step	Direct*	Multi-Step**	
P	111	83	28	(82, 1.1)	(86, 1.1)	(4 , 6)	(4 , 6)	82
G	146	112	34	(88, 1.0)	(91, 1.0)	(11 , 9)	(17 , 7)	88
V	123	92	31	(87, 1.1)	(97, 1.1)	(9 , 8)	(9 , 9)	97

*Learning time = 1 min., 20 sec.
**Learning time in last step = 15 sec.

Figure 2 Comparison of multi-step hypothesis formation, one-step hypothesis formation and Bayesian model

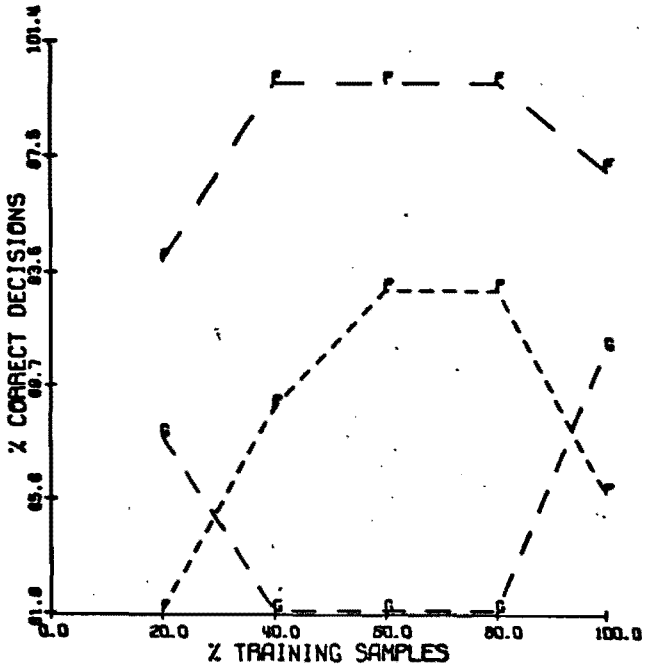

Figure 3 Correctness of expressions in multi-step hypothesis formation

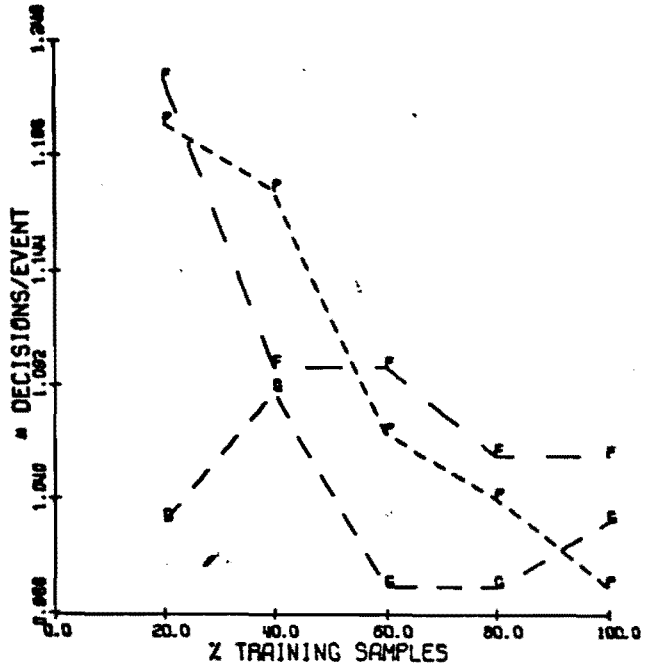

Figure 4 Indecision ratio

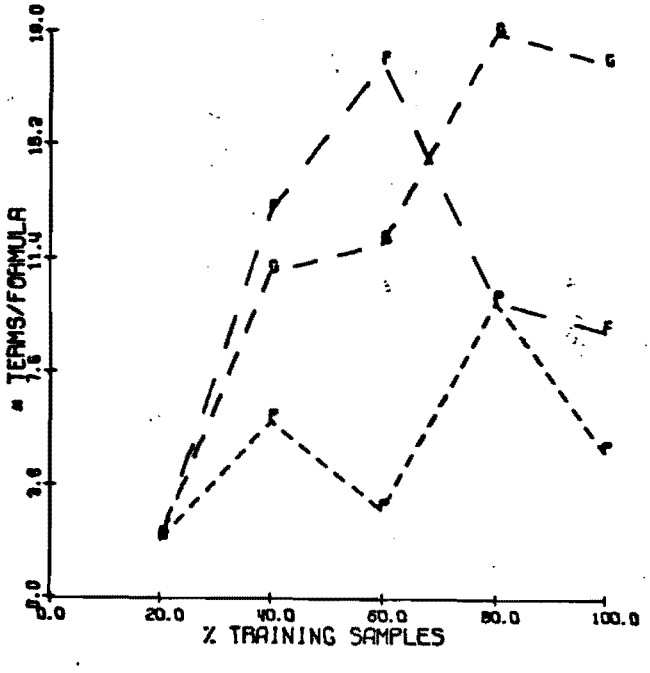

Figure 5 Complexity: # terms/formula

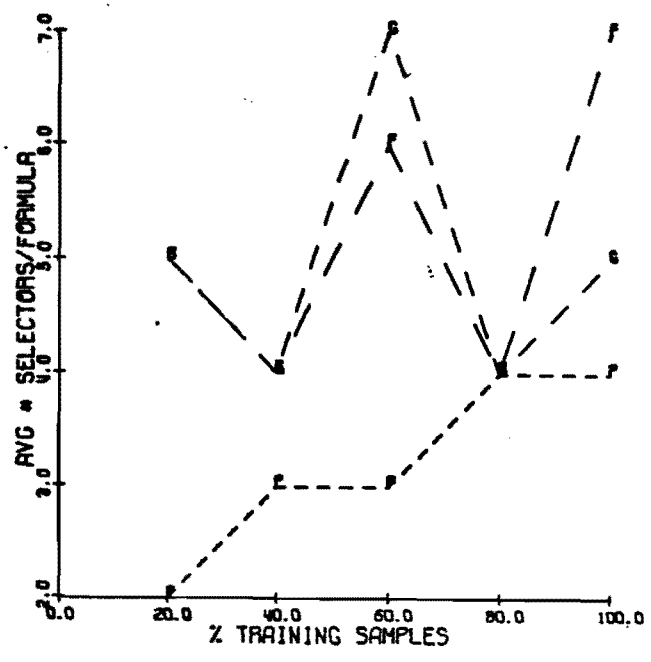

Figure 6 Complexity: Avg. # selectors/formula

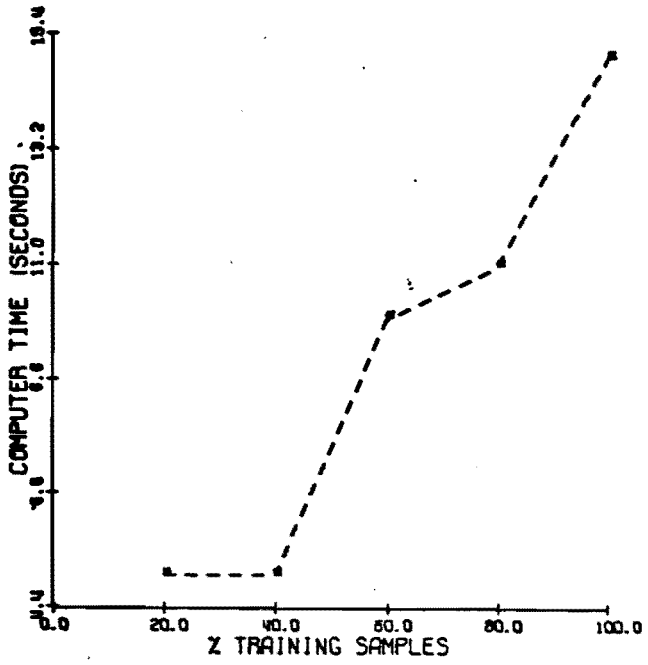

Figure 7 Computer time for each stage of learning

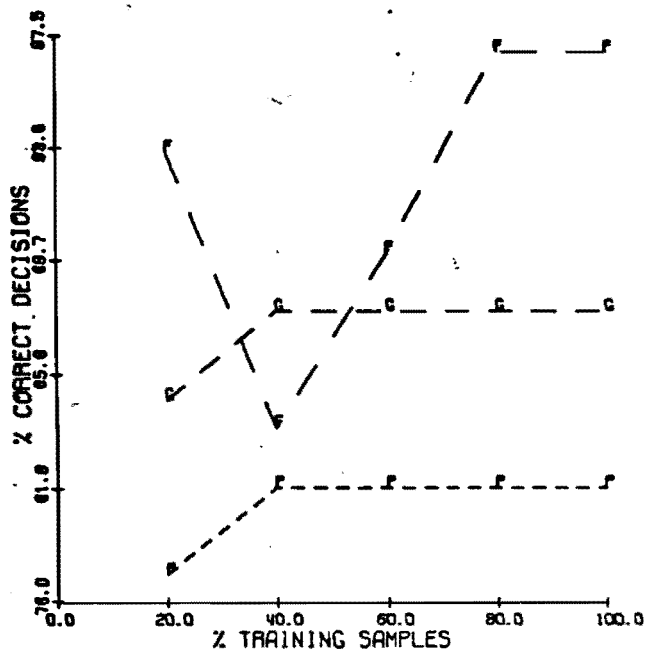

Figure 8 Bayesian analysis