PREDICTOR: AN ALTERNATIVE APPROACH TO UNCERTAIN INFERENCE IN EXPERT SYSTEMS

Allan P. White

Centre for Computing and Computer Science, University of Birmingham,
P.O. Box 363, Birmingham B15 2TT, United Kingdom.

ABSTRACT

An alternative approach to uncertain inference in expert systems is described which might be regarded as a synthesis of techniques from automatic induction and mathematical statistics. It utilises a type of pattern matching in which comparisons are made between new cases (as yet unclassified) and a database of past cases (in which the outcome is known). The method uses a stepwise approach in which, at each stage, that evidence variable providing the greatest additional discriminating power between classes (above that already obtained) is utilised. It avoids relying on the assumption of conditional independence.

A rudimentary system (PREDICTOR) which operates according to these principles has been written. Various adaptations to deal with missing and uncertain evidence are described, as are additional features such as a window, a facility for focusing discrimination on a subset of classes and a modification to deal with subjective data.

I INTRODUCTION

Traditional rule-based systems suffer from a number of serious deficiencies when attempts are made to incorporate mechanisms for dealing with uncertainty. The present author has attempted to illustrate these shortcomings elsewhere (White, 1984). Perhaps the most serious flaw is the assumption of conditional independence in circumstances where it is not justified. This particular topic has been discussed by Szolovits and Pauker (1978). It has also been admitted by members of the PROSPECTOR team in their final report (Duda et al, 1979). Other difficulties relate to the peculiarities arising from the use of fuzzy logic for combining probabilities. This has been criticised both by myself (White, 1984) and by Quinlan (1983). Further problems are caused by inconsistencies in the parameters built into such systems as a consequence of their being subjective estimates. Attempts to avoid these problems by not using formal probability theory have fallen victim to exactly the same difficulties, as shown by Adams (1976) in his discussion of the MYCIN system.

II AN ALTERNATIVE APPROACH

Another approach to the problem of uncertain inference in expert systems might loosely be described as a synthesis of techniques from automatic induction and mathematical statistics. It utilises data on past cases in order to make predictions or classification judgements about new ones. In this respect, it is similar to ideas on automatic induction in the AI field (e.g. Michalski, 1983) and also to a large family of statistical techniques. However, it differs from automatic induction in that the latter is really a logical procedure which, by its very nature, cannot handle uncertainty. It differs from orthodox statistical techniques in that no model is constructed for predictive purposes - instead each new case is matched against past data on a subset of variables, by a process about to be described. Other differences are that the system is capable of providing more of the customary features of expert systems than are usually found in traditional statistical approaches.

Suppose that we are concerned with some problem domain in which each case possesses values on a number of binary attributes, or evidence variables. Let us further suppose that each case falls into one of k classes that are mutually exclusive and jointly exhaustive. The first thing to do is to form a frequency table from a sample of past cases in which their class membership is known as well as their values on the various attributes. (As the number of possible patterns of evidence becomes large, the table would be expected to become increasingly sparse, i.e. would have an increasingly large proportion of evidence patterns with all class frequencies zero. Of course, such patterns need not be stored).

If we now consider a new case, in which the pattern of evidence is known but the class is not, then the computation of probabilities for classifying the case proceeds along the following lines. The technique yields probabilities which are conditional upon some appropriate subset of the variables.

The first step is to find that evidence variable which is the most powerful discriminator between the k classes. The way that this is done in PREDICTOR is by forming a k x 2 contingency table for each of the evidence variables (cross-tabulating class against the evidence variable) and choosing as the best discriminator that particular variable showing the strongest interaction with class, as measured by the x^2 test statistic for the k x 2 table. The entire frequency table is then conditioned upon that value of the variable found in the new case, thereby forming a sub-table.

This process is repeated, discriminating upon successive variables in turn, until no further steps can be taken. This occurs when none of the remaining independent variables yields a significant X^2 value when crossed with class.

The final step consists of collapsing the remaining sub-table into k cells (one for each class) and forming the probabilities for class membership from these in the obvious manner (along with their confidence intervals, if required).

It should be noted that when each new variable is conditioned upon, its interactions with those variables that have already been used are automatically taken into account by the *Very* nature of the process employed. Thus there is no need to make the customary assumption of conditional independence between variables, which is so often violated in conventional expert systems.

III DYNAMIC PATH GENERATION

The approach as outlined so far has statistical similarities with the discriminant function for categorical data described by Sturt (1981) and also with the technique outlined by Mabbett et al (1980). It resembles most closely the "probabilised" automatic induction procedure described by Hart (1984) but differs from all these procedures as follows.

The techniques just mentioned all generate decision trees for classification purposes (as does Quintan's 103 algorithm). The approach used in PREDICTOR does not do this. Although the process of branching on selected variables is involved, only that path necessary to classify the case under consideration is generated. Although all the information necessary to create the full decision tree is available in the database, this is not done. Thus the tree remains a virtual one, apart from the path through it that is actually generated. This neatly circumvents the worst aspects of the combinatorial explosion which could otherwise be problematic, even with quite small numbers of variables.

The interactive version of PREDICTOR interrogates the user in order to obtain the values of variables that it is going to condition (i.e. branch) on. The user has the option of giving a "don't know" response to any question. The algorithm used ensures that in such cases, these variables will not be conditioned upon but will be collapsed over at the final stage. This feature means that the information in the frequency table corresponds to a <u>number</u> of virtual trees, one of which will have a path actualised if it is required to classify a particular case.

IV THE STRATEGY FOR UNCERTAIN EVIDENCE

If the user is completely uncertain concerning the value of a particular piece of evidence, then it should assume its prior value and be dealt with as a missing value, as described previously.

However, for intermediate levels of uncertainty, some other strategy is required. If a variable that has uncertainty associated with it is not conditioned upon, then no further special action need be taken because the final sub-table will be collapsed over this variable (i.e. it will not feature as a node in the dynamically generated path). On the other hand, if the variable is one that is branched on, then further steps are required.

The modification is straightforward. Essentially, it uses <u>both</u> sub-tables that result from conditioning upon an uncertain variable, followed (at the final stage) by the collapse of both sub-tables and the computation of <u>two</u> sets of conditional probabilities. Each final conditional probability is then computed as the weighted sum of the two appropriate conditional probabilities, where the weights are the subjective probabilities for the different values of the uncertain variable. (In terms of the tree structure, a forked path is generated, with the uncertain variable located at the point of bifurcation).

A useful heuristic to save computation time is not to condition on uncertain variables when they are thrown up by the selection algorithm but to flag them for postponed conditioning at the penultimate stage (immediately before collapsing the sub-tables).

V FURTHER FEATURES

One of the features judged to be important in conventional expert systems is the presence of a "window", i.e. the user should be able to see how the system got to its present state. The approach outlined earlier can be supplied with a "window" in much the same way as any other stepwise statistical procedure. It would be quite easy to program the system to report at each stage what variables have been conditioned upon and, if required, what the k conditional probabilities would be if the current sub-table was collapsed without further conditioning.

Another idea concerns the focus of discrimination. At any given stage in the conditioning process, the user might wish to shift the focus of discrimination away from a global consideration of all k classes to a particular subset - perhaps two or three - which he judges more likely to be relevant. This feature could be implemented simply by altering the x^2 test used so that it would be based on the approriate subset of classes.

Finally, the issue of subjective data should be tackled. As PREDICTOR was conceived, it was intended to operate on objective (i.e. actual) data from past cases. However, it has recently occurred to the author that it might be feasible to use the same scheme to operate on subjective data. These data might be elicited from the domain expert, as follows. He would be asked to imagine some large number of cases (say fifty thousand) and would then be asked to specify how he would expect these cases to be distributed over the various evidence vectors and classes in the imaginary database. The intention is that the expert should pick out those combinations of class and evidence which are relatively common and also those which are relatively rare and attempt to estimate their frequencies. Having estimated these "peaks" and "troughs" in the frequency distribution, the remaining cases would be spread evenly over the database. Such a method of eliciting subjective data neatly circumvents problems arising from getting the expert to estimate impossibly complicated likelihood ratios and ensures consistency in the estimates - thereby avoiding the problem of inconsistent priors (White, 1984). However, one problem with this approach remains. The statistical sensitivity (or "power" in statistical terminology) depends on the number of imaginary cases that are used. If this number is too small, the system will be of little use because it will make few inference steps (or perhaps none at all) before coming to an automatic halt. If the number is too large then the system will possess a spurious degree of precision. Some further thought on this particular topic is obviously required.

REFERENCES

[1] Adams, J. B. "A Probability Model of Medical Reasoning and the MYCIN Model." Mathematical Biosciences 32 (1976) 177-186.

[23 Duda, R. 0., P. E. Hart, K. Konolige and R. Reboh. "A Computer Based Consultant for Mineral Exploration." Final Report, SRI Projects 5821 and 6415, SRI International, Menlo Park, California, September, 1979.

[33 Hart, A. E. "Experience in the Use of an Inductive System in Knowledge Engineering." In Bramer, M. A. (ed.) Research and Development in Expert Systems. Cambridge University Press, 1984.

[43 Mabbett, A., M. Stone, and J. Washbrook. "Cross-validatory Selection of Binary Variables in Differential Diagnosis." Applied Statistics 29:2 (1980) 198-204.

[53 Michalski, R. S. "A Theory and Methodology of Inductive Learning." Artificial Intelligence 20 (1983) 111-161.

[63 Quinlan, J. R. "Inferno: a Cautious Approach to Uncertain Inference." The Computer Journal 26:3 (1983) 255-269.

[7] Sturt, E. "Computerized Construction in Fortran of a Discriminant Function for Categorical Data." Applied Statistics 30:3 (1981) 213-222.

[83 Szolovits, P. and S. G. Pauker. "Categorical and Probabilistic Reasoning in Medical Diagnosis." Artificial Intelligence 11 (1978) 115-144.

[93 White, A. P. "Inference Deficiencies in Rule-based Expert Systems." In Bramer, M. A. (ed.) Research and Development in Expert Systems. Cambridge University Press, 1984.