

# Using Decision Tree Analysis To Develop An Expert System

Earl Chrysler  
earlchrysler@bhsu.edu

## Abstract

The development of an expert system typically requires a two-member team: the knowledge engineer and the expert. The knowledge engineer needs to extract information from the expert to build a knowledge base that is then used with a set of logical rules to develop the expert system. While performing a review of the literature in the area of expert systems one may locate several articles that demonstrate the use and effectiveness of expert systems, there is no discussion of any methodology used to develop the expert system. Upon reviewing various methods of determining an approach to logically analyze the results of sequential decision-making, one notes that a popular and apparently efficient method frequently used in this type of situation is the decision tree method. This paper suggests that a very efficient method a knowledge engineer could use is the decision tree analysis approach.

**Keywords:** Expert system, decision tree analysis, logical flowcharting

## 1. BACKGROUND

When it appears an expert system could be of value in an organization, a knowledge engineer, that is, a person well versed in the development of an expert system, typically using an expert system "shell" software package, is assigned to work with one or more persons designated as having expert knowledge in some specific area. Examples of the justification of the development of an expert system are for use as a training device or the documentation and preservation of the logical decision-making process of someone who performs a unique function in the organization. An example of the use of an expert system as a training device is shown in "The Development of an Expert System for Managerial Evaluation of Internal Controls" by Changchit and Holsapple (2004). Another application area is optimizing a decision process. In CIO, Richard Pastore (2003) discussed the results of an expert system that analyzes 80,000 customer pickup orders and change requests per day for Con-Way Transportation Services and develops an optimum schedule of overnight shipping routes. An application that

is affecting the IT area is one called a BRMS (business rules management system). James Owen describes a system that separates the business logic of a computer system from the data validation logic. The business analysts develop the BRMS "that governs how enterprise applications behave" (2004).

Decision tree analysis has long been used when a multi-stage decision process is involved. Some recent examples appearing in the literature are "A decision tree for selecting the most cost-effective waste disposal strategy in foodservice operations" (2003) by Wie, Shanklin and Lee. Their application of the decision tree analysis methodology developed an illustration of the decision-making process that occurs when conducting cost analysis and subsequent decisions. When faced with the multitude of small business assistance programs conducted by public, private and nonprofit organizations, it was apparent an integrated approach was needed to determine which program(s) were appropriate for small businesses with specific characteristics. In their paper entitled "A decision tree approach for integrating small business assis-

tance schemes," (2004) Temtime, Chinyoka and Shunda provide empirical evidence of the need for an integrated model using a case study of small business assistance programs in the Republic of Botswana and how a decision tree analysis approach could match small businesses with existing assistance programs. Recently O'Brien, and Ellegood (2005) developed a decision tree approach to assist social service administrators in determining the validity of an ADA claim. The decision tree allows administrators "to break the decision-making process into discrete steps that can be considered separately and sequentially." It is just this latter capability that makes the decision tree analysis technique such an effective tool to assist one in the development of an expert system.

## 2. STATEMENT OF THE EXAMPLE PROBLEM

Ideally, when a researcher designs a study the hypotheses to be tested are clearly defined, the type of data that will be collected is described, the types of statistical tests that will be performed are stated and the implications of the expected findings are discussed.

In reality, however, many times a researcher may well define the questions to be addressed by the study and the general nature of the data that will be collected, but be unaware of which statistical test(s) would be appropriate. Also, many times a researcher collects data of various types and afterward realizes that perhaps the data could be subjected to additional analyses to identify relationships that were initially not considered. In addition, there are individuals who review the papers of others who may question whether the researcher applied the proper statistical technique, given the nature of the data and the hypotheses to be examined. It is suggested that it would be useful, therefore, if one could have access to an expert system that could assist one in determining the appropriate statistical technique to use in a specific situation. While there may be many approaches to developing such an expert system, the use

of decision tree analysis is proposed and its application to this example problem will be demonstrated.

## 3. METHODOLOGY

Decision tree analysis is, simply defined, examining a decision point and asking, "What are the possible outcomes or options?" Then, for each option or outcome, one must obtain the probability of that outcome or option occurring. For each one of the outcomes or options, one then asks the same question, "Given that we are here, what are the outcomes or options that could occur?" Once again, the likelihood of each of the options is estimated. This process is continued until each possible path has reached a conclusion. This process was applied to the basic question at hand, i.e., given one wishes to perform a statistical test, the general nature of the data is known and the purpose of the test is given, which statistical technique is appropriate?

## 4. BUILDING THE MODEL

The first step is to consider the purpose of the statistical test. The basic question, then, is: what is one attempting to determine? If one is attempting to determine if two groups are equal or one is significantly different than the other, then one wishes to compare two groups. If one is concerned with determining if more than two groups appear to be the same or are significantly different, then one wishes to compare more than two groups. If one suspects that two events are interacting in some way, then one wishes to determine if two events are related. If one wishes to investigate the possible impact several events are having on some outcome, then one wishes to determine if one event is related to many other events. As a consequence, the first question to be posed would list the possible options of what one wishes to determine that were developed above. The start of the decision tree analysis that shows these first stage options is shown in Figure 1. As an example, the first option will be pursued.

If one wishes to compare two groups, there are additional questions that must be an-

swered and, it is suggested, there is a specific sequence in which the questions need to be answered. Again, for the purpose of this example it will be assumed that all of the following questions will be answered "Yes". One would first be interested in the type of data available for analysis. Therefore, the question to be answered would be: is the data ratio scale or equal interval? Assuming the answer is "Yes", one would then need to know the characteristic of the distributions of the data points. Therefore, the question to be answered would be: are the parent populations approximately normal? Assuming the answer is "Yes", one would then need to know another characteristic of the distributions of data points. Therefore, the next question to be answered is: are the variances of the parent populations approximately equal? Assuming the answer is "Yes", the independent t-test is indicated.

For the "technical" questions regarding the normality or Gaussian quality of the parent populations and the equality of the variances of the parent populations, one should be given the options of not only "Yes" or "No", but also "Unknown". Where a user responds "Unknown" there should be as a conclusion advice as to how the user may be able to ascertain that the answer is either "Yes" or "No". In that manner, the user can leave the system, perform the task(s) that were stated in the conclusion, then return to the expert system and progress deeper into the rules. The reason the questions were presented in the specific sequence shown above is that the response to an answer of "Unknown" at one point assumes the answer to the previous question was not "Unknown".

Also, in order to be most useful to the user, the knowledge base and the resulting expert system should be able to advise the user on the most common parametric tests. However, if the user has selected options that have bypassed the most common parametric tests, the expert system should continue until a conclusion is either the recommendation of one of the most common non-parametric tests or, as is sometimes possible, that there is no known statistical test

that meets the nature of the test to be performed and the characteristics of the data.

## 5. THE RESULTING EXPERT SYSTEM

Using decision tree analysis resulted in a flowchart that one would follow to determine which statistical test (if an appropriate one can be identified) should be used by the researcher. Sample sections of the resulting flowchart are shown as Figures A, B and C. The expert system that was developed using this technique and the VP Expert expert system shell will be demonstrated at the conference.

There are two methods that are available for the expert system to interface with the user. One method is for the monitor to display the questions to be answered and the available answers from which the user is to select. Another method is for the above presentation to appear in only the top half of the monitor screen. The lower half of the display is divided into two displays. On the left half of the lower half, the rules being applied are shown. On the right half of the lower half of the screen is shown the result of the decision and the confidence factor (CNF) allied with the outcome. For situations where the outcome of the decision is probabilistic, the probability is known as the level of Confidence one has with outcome. Screen images for the second option are shown for the situation where the independent t test is indicated.

## 6. CONCLUSION

The decision tree analysis method was used to develop the expert system discussed and presented here. The decision tree analysis method assisted the expert system developer in the creation of the necessary knowledge base and rules section of the expert system due to the step-by-step, multi-stage decision process the developer had to follow. In addition, the developer had to consider every possible option at every step in order to assure that the expert system would not

make an erroneous recommendation to the user. It is suggested that, for those who wish to create an expert system, for whatever purpose, the decision tree analysis approach will assure that the resulting expert system will have all options considered and will have been the most efficient method the expert system developer could have used to create the resulting expert system.

## REFERENCES

1. Changchit, Chuleeporn and Holsapple, Clyde W. "The Development of an Expert System for Managerial Evaluation of Internal Controls", *Intelligent Systems in Accounting, Finance and Management*, Vol. 12, No. 2, April-June, 2004, pp. 103-120.
2. Owen, James. "Putting Rules Engines to Work", *Infoworld*, Vol. 26, No. 26, June 28, 2004, p. 34.
3. O'Brien, Gerald V. and Ellegood, Christina. "The Americans with disabilities act: A Decision Tree for Social Services Administrators", *Social Work*, Vol. 50, No. 3, July, 2005, pp. 271-279.
4. Pastore, Richard. "Cruise Control; This freight delivery company's leaders took four years to get a new expert system right. Now they're watching as the benefits roll in", *CIO*, Vol. 16, No. 8, February, 2003, p. 1.
5. Temline, Zelealem, Shinyoka, S.V. and Shunda, J.P.W. "A decision tree approach for integrating small business assistance schemes", *The Journal of Management*, Vol. 23, No. 5/6, 2004, p.563.
6. Wie, Seunghee, Shanklin, Carol W. and Lee, Kyung-Eun. "A decision tree for selecting the most cost effective waste disposal strategy in foodservice operations", Vol. 103, No. 4, April 2003, p. 475.

FIGURE A  
EXPERT SYSTEM LOGIC FLOWCHART  
DETERMINING THE PURPOSE OF A TEST

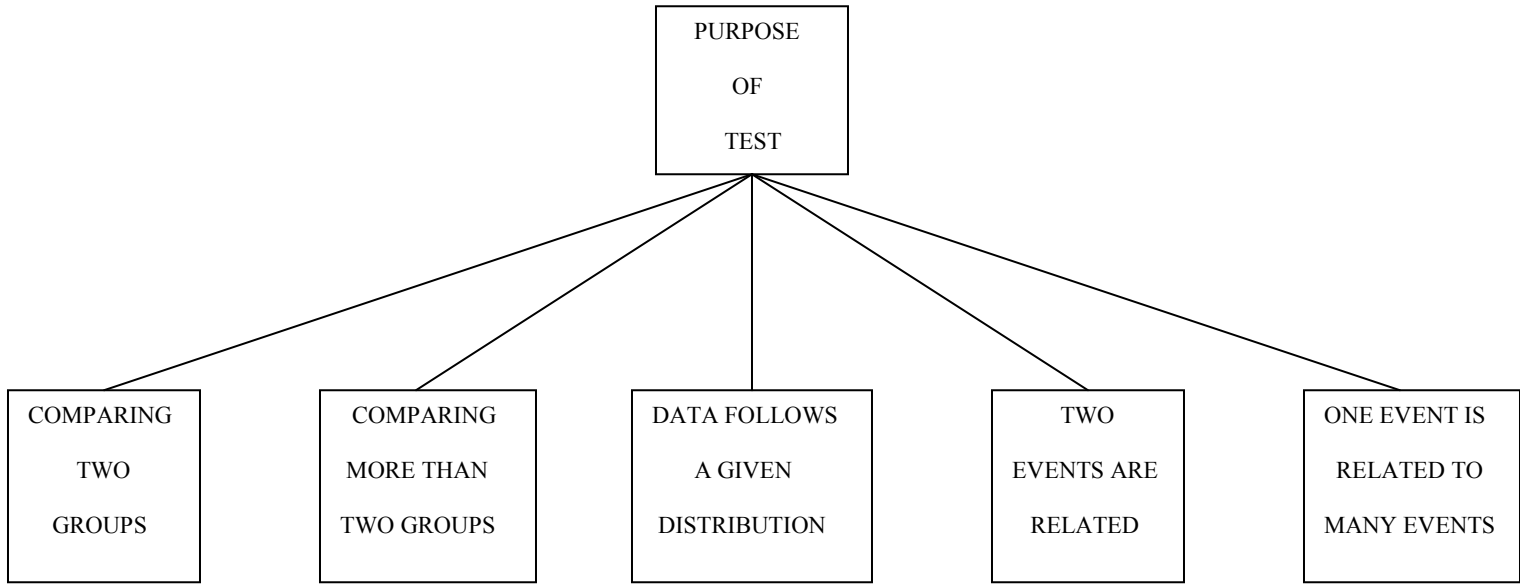


FIGURE B  
EXPERT SYSTEM LOGIC FLOWCHART  
COMPARING TWO GROUPS – INDEPENDENT T TEST INDICATED

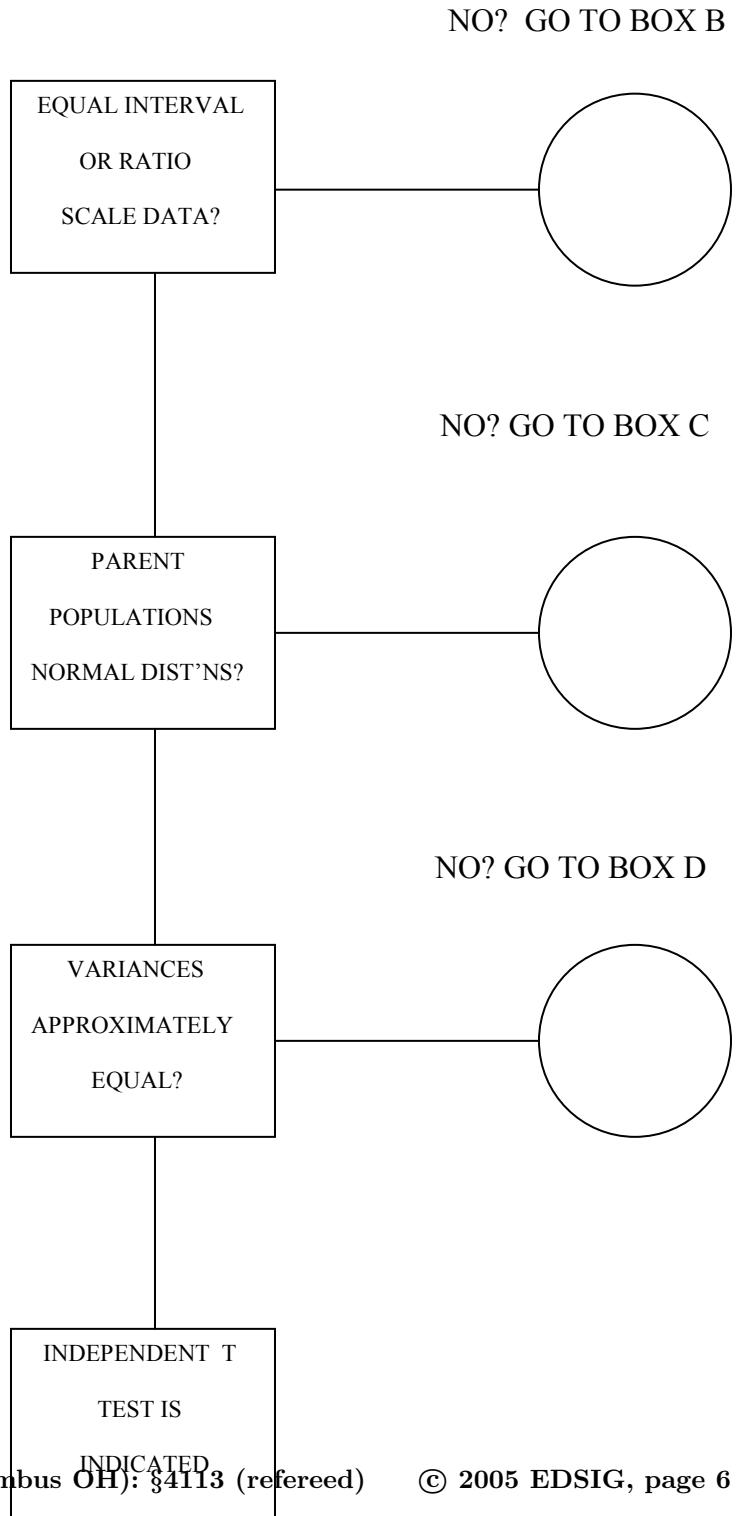


FIGURE C

EXPERT SYSTEM LOGIC FLOWCHART

COMPARING MORE THAN TWO GROUPS – TWO WAY ANOVA INDICATED

NO? GO TO BOX F

