Interactive Learning Tool for Statistical Reasoning with Uncertainty

Deborah A. Vastola and Ellen L. Walker
Computer Science Department
Rensselaer Polytechnic Institute
Troy, NY 12180-3590
vastod@rpi.edu, walkere@cs.rpi.edu

Abstract

We describe a tool which was developed as a Masters project to help students learn the concepts of reasoning with uncertainty. While the initial implementation has focused on Dempster-Shafer theory, the tool is designed so that it can be expanded to other models of uncertain reasoning. The tool uses interactive graphical representations, together with animation and color, to convey the principles of reasoning with uncertainty. It runs on the Unix platform under the X-window system and uses the SUIT toolkit (Conway & Pausch & Passarella 1992). A demonstration of the tool will be included with the presentation of this paper.

Introduction

In teaching AI, it is clear that many of the students, particularly those without a strong background in probability, are confused by the concepts of statistical reasoning. Many students have little or no intuition about how measures of uncertainty should work, much less the mathematical sophistication to understand why. Thus, we have developed a framework, in the form of an interactive tool, to help students learn about the power and limitations of statistical reasoning. Since our intention was to develop a framework for learning about statistical reasoning, rather than teaching every method, our initial implementation has focused on one method, Dempster-Shafer theory (Rich & Knight 1991; Shafer 1976). Our objective is to create a tool design that can accommodate additional statistical reasoning models (Rich & Knight 1991).

The tool we have developed is based on the following pedagogical principles:

- Learning should be fun.
- Examples make concepts clearer.
- Students learn best by investigation and exploration.

It includes the following components, each with its own set of windows that can be selectively shown or hidden by the students.

1. The expert (domain knowledge specification)
2. Graphical representation of the rules
3. Animation of evolving beliefs
4. A walk-through of the mechanics of the model
5. Help panels that reference the current example

Components of the Tool

In this section we describe the five main components of the tool.

Component 1: The Expert

This component allows the student to textually specify the rules of environment on which the statistical reasoning model will act. An environment includes a set of measurable attributes, evidence, and potential hypotheses. The rules establish certainty values between evidence and conclusions which complete the domain knowledge of the expert. In the Dempster-Shafer model, these values are the mass functions, m. In other models, these values would be certainty factors or fuzzy membership distributions.

Students can create new rules from scratch or modify existing ones. Each set of rules in an environment is stored in a unique file called a *case* file. Hereafter we just use the term case to refer to these rules. An example of a very simple case, What's-in-the-package?, is given in Table 1. Here, specific characteristics of a package are linked to conclusions about the contents of that package.

The expert component makes the tool flexible and extensible. It can be adapted to the domain of the student's choice, and the student can explore various sets of connections between evidence and conclusions. This makes the tool more fun to use, and provides a chance for the student to be creative. Additionally, the *availability* of this component makes a key point: a statistical reasoning system is only as good as the expertise that it is based upon. Dempster-Shafer evidence combination, for example, does not help decide what the complete set of evidence and conclusions are; nor does it aid in selecting the correct m values.

Since developing a usable set of domain knowledge requires an understanding of how it will be used, the

students should be able to use the tool without initially developing their own domain knowledge. Therefore, the initial tool includes at least two predefined environments and associated cases.

Component 2: Graphical Representation of the Rules

This component displays the domain knowledge graphically, so that the student can make educated guesses about which pieces of evidence to query. The rules described by the expert are visually presented as a causal network, with evidence and hypotheses as nodes, and links from evidence to hypotheses labeled with their relative contribution or weight of causality. Evidence nodes are clustered by attribute, so that mutually exclusive nodes are readily visible. An example of the causal network is given in Figure 1 (left window) for the What's-in-the-package? case.

The graphical interface also serves as a user-friendly interactive vehicle for querying an underlying simulation. This simulation generates a particular situation from the domain, and allows the student to "learn" the simulated evidence by clicking on the evidence nodes. The tool provides two separate modes for interacting with the causal network: (1) experiment mode and (2) game mode. When in experiment mode, if a user clicks on a piece of evidence they are in effect asking the tool to "show me the results when I observe this piece of evidence". Experiment mode was designed to allow students to explore the possibilities. Clicking on a piece of evidence has a slightly different meaning in game mode. We discuss game mode in a later section.

Component 3: Animation of Evolving Beliefs

This component shows the progress of the evolving certainty of the various hypotheses. The concept that we want to convey to the student is that no matter which model is used, the purpose of statistical analysis is to move from a position of uncertainty toward a position of certainty (or at least to become more certain of how uncertain the beliefs are).

As each piece of evidence is "observed" (by clicking in the causal network window), the animation shows the result of the evidence on the certainty of the current set of hypotheses. This animation is intended to give the student an intuitive feel of the statistical model - not the details or the mechanics of the model. Our goal was to present a common-sense pictorial view of what is happening. So it was a conscious decision to exclude any reference to numbers or specific values.

Specifically, the Belief and Plausibility of each subset is represented as a bar (upper and lower bounds on a scale of 0 to 100%). Figure 1 (top and right windows) illustrate belief and plausibility results when the user has "observed" a small package. We should mention at this point that color plays a very important role in this tool in helping the user to make visual correlations and associations of results. Of course, our use of color cannot be captured in the monochrome figures we include in this paper. In the included figures, the height of the "blackened" area of the bar chart represents belief; the height of the "greyish" area of the bar chart represents plausibility. The bar charts for the singleton hypothesis sets are displayed in the top window to emphasize their importance. In a color system, their color is chosen to match the color of the corresponding hypothesis in the causal network.

The mass distribution (m-values) of the Dempster-Shafer model is represented as a pie chart. See Figure 2 (left window). We include a "before" and "after" snapshot of the m-values since it is instructive for users to see how the pieces of the pie change when evidence is added. It is particularly instructive to see the effect of new evidence on the m value of *theta*.

Our decision to use bar and pie charts to show results is based upon the following points:

- Belief and Plausibility measurements are the lower and upper bounds on the probability of a hypothesis. Thus, it makes sense to view the current probability of each hypothesis as a sub-range of the complete 0-100% range bounded by the Belief and Plausibility numbers. The location of the range shows the probability of the hypothesis, and the size of the range shows how certain we are of our estimate. This model makes the distinction between "50% chance of rain because I have no idea whether it will rain or not" vs. "50% chance of rain because the current weather conditions produce rain 50% of the time" very clear (the first band is 0-100%, the second is 50-50%).
- In order to effectively see how a piece of evidence contributes to the belief and plausibility, the eye needs a stable focal point from which the animated change occurs. The full bar (0-100%) provides this focal point; only the area delineated by the upper and lower bounds moves. For this reason, once a bar for a particular hypothesis subset is drawn, its position on the screen remains constant.
- To help the student intuitively relate m-values to Belief and Plausibility, they should be shown on the same screen. Since the m-values add up to 1 (100%), a pie chart is a natural display vehicle. Since we did not have enough real estate to include the causal network, bar charts, and and pie charts on the same screen, we opted to include a pop-up window for the pie charts so that the user can still easily compare the belief and plausibility results with the m-value outcome (by simply pressing the "More" button). We again remind the reader that the use of color for the regions of the pie chart is heavily correlated with the colors used for the bar charts and the causal network.

Filtering the Bar Charts. Belief and Plausibility values are associated with each subset of the complete

set of hypotheses. When dealing with only a handful of hypotheses, it is valuable to display all possible bar charts since the student can begin to learn which results are significant and which results are perhaps redundant. However, as the number of hypotheses increases, it becomes more and more impractical to read all 2^N possible bars (when N=10 hypotheses, this would yield 93 pages of bar charts!). Under these conditions, it is desirable to be able to selectively restrict the output generated. So, the tool includes a filter option which allows the user to specify thresholds for displaying results. The Dempster-Shafer filter uses three thresholds, one each for m value, plausibility, and belief. When a subset satisfies all three thresholds then, and only then, a bar chart will be created for that subset. Setting all 3 thresholds to 0 will in effect turn filtering off.

Which filter setting is most appropriate depends upon the complexity and nature of the case and depends upon which points about the model one wishes that case to convey. Setting the right thresholds also requires a more solid understanding of the model. For these reasons, we leave it to the expert to determine the appropriate settings when the case is created.

Component 4: A walk-through of the Mechanics of the Model

In this component, the precise algorithm of the model (i.e., Dempster-Shafer evidence combination, for our first implementation) is explained for the specific case environment. By pressing the "#'s" button, a Ghostview window pops up which shows the the actual stepwise computations for the pencil and paper calculations that the student would be expected to carry out, as well as comprehensive tables of all calculations.

Component 5: Extensive/Case-Specific Help

Good help panels of course are important to any learning tool. In addition to help options which convey background and "how to get started" information, the user can click on any object that shows a result to get tips on how to analyze that result and what they should focus on.

Also, feedback is often most helpful when it can reference the current example. All help panels and feedback are message-file driven. In most cases, each message has a common part and a variable part, the latter to be filled in at the time of execution of the case before the message for help/feedback is displayed. We belief we have struck a good balance between generic information and feedback tailored to the specific example.

Game Mode

While the students can experiment with specific cases of Dempster-Shafer reasoning (via the experiment mode), there is an advantage to introducing statistical reasoning as a game. This will provide a sense of fun to the tool, as well as the potential for some friendly competition. Also, since the game mode is more restricted in that the observations and objective are fixed and known in advance, we can provide more relevant and timely feedback in this situation.

The game is similar to "twenty questions" and can be played in any environment created by the expert. Thus, the initial tool will have two games (from the two predefined cases). The player's goal is to select the most reasonable conclusion (from the set of potential hypotheses) by asking the fewest questions (i.e. using the smallest set of evidence). A conclusion and set of evidence leading to that conclusion is chosen at random and revealed to the player upon clicking on the evidence nodes in the causality graph. So in our simple What's-in-the-package? case, when the user clicks on the *small* evidence button when in game mode, they are asking "Is the package small?" to which the tool will reply either "yes" or "no". If the package is small, the tool will display the Dempster-Shafer results associated with a *small* package being observed. The goal of the game is for the player to correctly guess the conclusion in as few moves (i.e. with as little evidence) as possible. The player is awarded extra points for being able to provide the correct answer in fewer moves.

It is anticipated that each player will use both the causal network and evolving certainty ranges as the evidence is revealed to determine which evidence to consider next. As the students begin to understand the relationship between the causal network and the evolving ranges, they will learn that some evidence can add more information than other evidence, and will improve (decrease) their scores in the game.

User as Active Participant

One final point. How this tool is introduced to students and used by students will be critical to its effectiveness. Minimal value will be gained if the user is a passive observer (in fact, because of the tool's simplicity, it would get pretty boring just watching the results). Time should be reserved not only for interaction and observing the results, but, more critically, to sit and ponder the "why" and "what if" of the results. For this reason, it is highly recommended that the instructor orchestrate a few questions to accompany the student's session with the tool. Questions such as the following reinforce basic concepts and uncover some of the deeper issues associated with the model:

- Looking at the experiment results for What's-in-the-package? (Figure 2) Why is it that "game" has a non-zero belief yet the belief in "clothing" is zero? Both hypotheses were stimulated by observation of evidence that had similar m values. So why such a difference in belief? Hint: look at the causal network.

- Is Dempster-Shafer commutative? How would you use the tool to try to answer this question? How would you prove your answer?
- What filter would you set to eliminate redundant information yet still ensure that significant results are will not be hidden? Hint: look at the relationship between m, belief, and plausibility.
- Subset A has belief=30% and plausibility=70%. Subset B has belief=40% and plausibility=60%. Which set do you feel most confident contains the correct answer? Why?

It is the coupling of the student's interaction and concentrated think time that will optimize the tool's effectiveness.

Conclusion

We have described a tool that uses interactive graphical representation, together with animation and color, to convey the principles of reasoning with uncertainty. We have also highlighted the pedagogical principles behind some of our design decisions. These include simplicity, the facility to phase-in concepts and details, help and feedback that focus the student's attention, and, most importantly, viewing the student as an active explorer.

Before the tool is placed in the field, it will undergo a usability test which will include participants with varied backgrounds in statistical reasoning and probability (from none to some) so that we may begin to assess the design choices and tradeoffs made and to identify any follow-on work necessary to improve the tool.

References

Rich, E., and Knight, K., 1991. *Artificial Intelligence.* New York, NY: McGraw-Hill.

Shafer, G., 1976. *A Mathematical Theory of Evidence.* Princeton, NJ: Princeton University Press.

Conway, M., Pausch, R., and Passarella, K., *A Tutorial for SUIT, the Simple User Interface Toolkit* 1992. Computer Science Department, University of Virginia.

Table 1: Case: What's-in-the-Package?

Attribute	Evidence	Hypotheses	mValue
sound	If the package does not rattle	it contains clothing, a book, or CD	0.5
sound	If the package does rattle	it contains a game	0.7
size	If the package is small	it contains a book or CD	0.8
size	If the package is large	it contains clothing or a game	0.6
weight	If the package is heavy	it contains a book	0.8
weight	If the package is light	it contains clothing or a CD	0.6

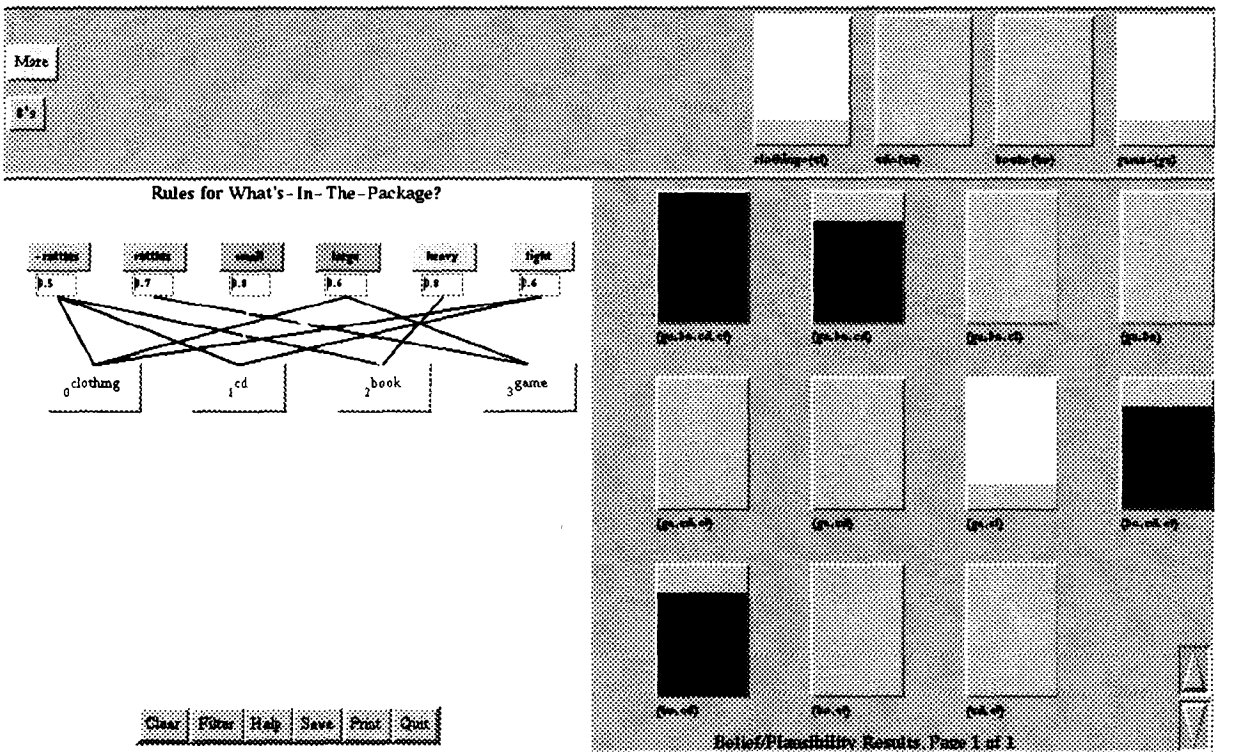

Figure 1: *When evidence "small" has been observed*

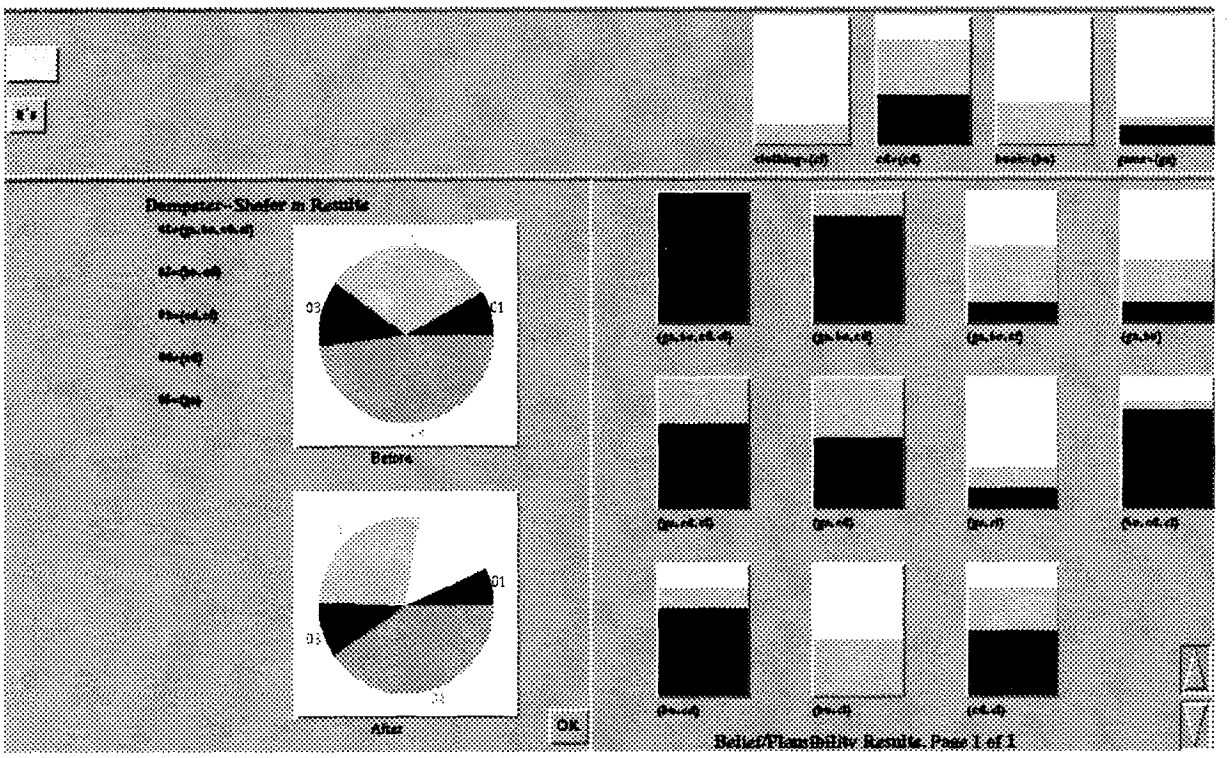

Figure 2: *When "small/light/rattles" have been observed*