The Dempster-Shafer Theory of Evidence as a Model of Human Decision Making

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 12(0)

Author
Mitchell, Donald H.

Publication Date
1990

Peer reviewed

The Dempster-Shafer Theory of Evidence as a Model of Human Decision Making

Donald H. Mitchell
Amoco Production Company
Tulsa Research Center
P.O. Box 3385
Tulsa, OK 74102*

Abstract

Many psychology researchers have shown that humans do not process probabilistic information in a manner consistent with Bayes' theory [9, 10, 16, 24, 23, 27]. Robinson and Hastie [24, 23] showed that humans made non-compensatory probability updates, produced super-additive distributions, and resuscitated zero probability possibilities. While most researchers have classified these behaviors as non-normative, we found that the Dempster-Shafer theory could model each of these behaviors in a normative and theoretically sound fashion. While not claiming that the theory models human processes, we claim that the similarities should aid user acceptance of Dempster-Shafer based decision systems.

1 Introduction

Due to the inherent uncertainty of evidence and conclusions in the world, decision support systems (including artificial intelligence systems) must often use methods for representing and reasoning under uncertainty. There are a number of possible methods. Each method has a different effect on the three major expert system stages: 1) acquisition, 2) inferencing, and 3) user interpretation of the results. While many products and papers downplay the importance, the choice is difficult and important. The chosen paradigm can mitigate or exacerbate errors in any of the stages thus making the system's results meaningless.

There are a number of results supporting each reasoning method. One attribute of comparison is theoretical soundness [6, 17, 18, 19, 2]. Most of these comparisons uphold the theoretical foundation of probability theory and particularly of Bayes' theorem. Another attribute is empirical performance [8, 1, 29, 22, 17, 20, 21]. These studies support a variety of conclusions. Dawes [8], for example, shows that using a simple, yet incorrect, linear model is often better than a theoretically sound probabilistic model when they are both based on the same error-prone human estimates. In this paper, we argue for a third attribute—user interpretation and acceptance.

When consulting a decision support system, a human's ability to understand the computer's beliefs and decisions is important. Early research on automated tools showed that users more readily accept systems if they understand the systems' behaviors [4, 5, 11, 14, 25]. This understanding can result from any of three processes: 1) training the human to understand the theoretical correctness of the reasoning processes, 2) using a reasoning process that directly corresponds to the human's or 3) using a process whose observable behavior corresponds to the human's. The first process is apt to meet with resistance and makes general distribution and acceptance difficult. The second process, while of great potential, is difficult to accomplish due to the hidden nature of human decision-making processes. The third process corresponds directly to the way that most collaborative human decision-making works: when humans defend their reasoning, they refer to the evidence that caused them to increase or decrease their belief and not to their reasoning mechanisms.

This paper uses experimentally observed similarities between the Dempster-Shafer theory of evidence [26] and humans solving a probabilistic updating task to argue that humans may more readily understand Dempster-Shafer based systems. The paper does not directly address user acceptance in that it does not involve actual users of a system, but it indirectly addresses acceptance through the ability of humans to empathize with the behavior.

*(918) 660–4270, dmitchell@trc.amoco.com

2 The task and human data

This paper reports an experiment that compared the behavior of an automated Dempster-Shafer evidence accumulation system with the behavior of humans performing the same evidence accumulation task. The task and human experimental results come from a study by Robinson and Hastie [24, 23].

In an effort to find out whether humans followed Bayesian probabilistic reasoning principles, Robinson and Hastie asked human subjects to solve a murder mystery. The subjects saw a series of clues. After each clue, the subjects stated their beliefs about the guilt of each suspect in terms of probability. Robinson and Hastie found that the humans did not follow probabilistic principles. We will explain the exact form of the discrepancy when we describe the behavior of the Dempster-Shafer system.

Before describing the Dempster-Shafer system, we will address two concerns with the Robinson and Hastie data. Some may argue that Robinson and Hastie's subjects did not have adequate training in the probabilistic concepts. To test this hypothesis, Robinson and Hastie explicitly taught the fundamentals of probability to some of the subjects. Depending on the subject, the training led to either the same behavior as those without training or a behavior that did not reflect any evidence accumulation. Robinson and Hastie conjectured that the cognitive overhead prevented the subjects from applying the learning.

Another objection may be that one study makes an insufficient basis for concluding that humans are not Bayesian decision makers. Robinson and Hastie, however, are not the only researchers to show that humans make poor Bayesian probabilistic information processors. Many other psychology researchers have shown the non-Bayesian character of human information processing [9, 10, 16, 24, 23, 27].

3 Dempster-Shafer predictions

To test the predictions of the Dempster-Shafer [26] theory, we developed a straight-forward implementation of the theory [20] and then submitted the clues to it as a series of consonant belief functions. Although many artificial intelligence researchers [3, 7, 13, 12] have restricted their Dempster-Shafer representations to simple and dichotomous belief functions, we chose consonant belief functions because they are the form that Shafer says most naturally represent inferential evidence [26, pp. 223–229]. The reason other artificial intelligence researchers have ignored this representation is that it can result in exponential computational time requirements. For the sake of this study, computational time was not of direct concern.[1]

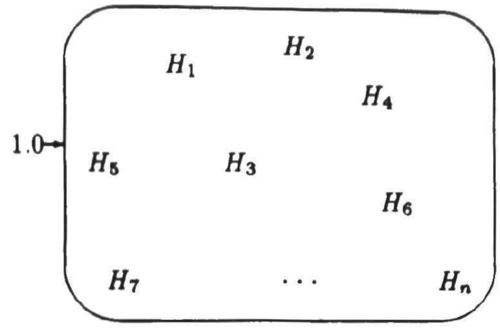

Figure 1: Venn graph of the vacuous function: there has been no evidence.

While it is not the intent of this paper to fully introduce the Dempster-Shafer theory, we will briefly describe belief functions. A belief function is the assignment of probabilities to sets of conclusions. This assignment differs from standard probabilistic theories in that it uses sets rather than single hypotheses. All probability theories can use sets to represent multiple simultaneously true hypotheses; however, the Dempster-Shafer system uses sets to indicate lack of differentiation in the evidence for mutually exclusive hypotheses. The interpretation of this assignment is that some element of the set is true but the evidence does not provide fine enough granularity to directly point to one hypothesis.

A consequence of this representation for belief is that there is a clear distinction between the inability to decide due to lack of evidence and the inability to decide due to too much conflicting evidence. In the Dempster-Shafer theory, a believer represents the lack of evidence as the assignment of all probability to the undifferentiated set of all possible hypotheses (e.g. figure 1), whereas the representation for conflicting evidence is the assignment of roughly equal amounts of probability to many separate singleton sets of hypotheses (e.g. figure 2). For example, in a well-matched musical competition, the judge's initial belief should be no one has evidence in their favor and all contestants can fight for the prize like a pie ready to be divided: there is no conflicting beliefs concerning the outcome. After listening to everyone perform

[1] Note, however, that [20, 15] both show linear complexity when applying the theory to naturally constrained problems.

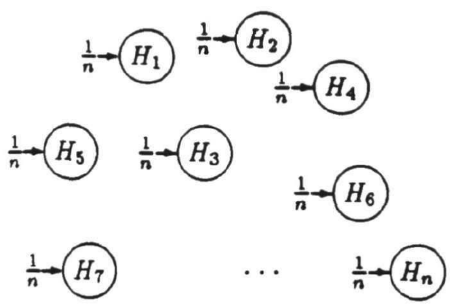

Figure 2: Venn graph of a maximally conflicting belief function: there has been evidence supporting each hypothesis equally.

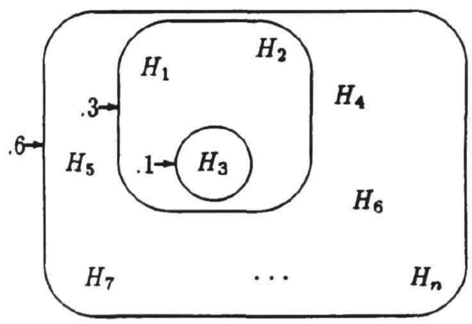

Figure 3: Venn graph of the consonant belief function: $\Pr(H_3) = .1$, $\Pr(H_1 \vee H_2 \vee H_3) = .3$, and $\Pr(H_1 \vee \cdots H_n) = .6$.

well, the judge should have belief in each contestant individually with little or no residual belief: each contestant has claims to more pie than is available. The evidence causes conflicting beliefs.

More generally, assigning non-zero probabilities to sets of hypotheses that don't subsume one-another represents conflicting belief. The common Dempster-Shafer representation used in artificial intelligence of assigning belief to a hypothesis and its negation, thus, directly encodes conflict. Individual pieces of evidence, however, should not show any conflict with themselves, and, therefore, this dichotomous representation is usually inaccurate. The consonant belief function is the non-conflicting alternative.

Figure 3 depicts a consonant belief function. In this example, some evidence supports hypothesis H_3. The same evidence has less direct support for hypotheses H_1 and H_2 and also has some residual uncertainty. This belief function is represented by assigning non-zero probabilities to the sets H_3, H_1, H_2, H_3, and finally to $H_1 \ldots H_n$. This gradual focusing of probabilities on progressive subsets is the definition of a consonant belief function. A consonant belief function is consonant with itself: that is, it shows no conflict with itself.

The consonance of an individual piece of evidence with itself implies nothing about consonance between pieces of evidence. Different pieces of evidence can conflict with each other. The detective story used in this experiment, for example, shows considerable conflict between clues. While each clue might be self-consistent and therefore consonant, that does not imply that all clues agree. The result of combining these disagreeing but self-consonant belief functions will not be consonant.

To make a choice among the hypotheses requires comparing the beliefs assigned to each hypothesis. In the Dempster-Shafer theory, there is not a single measure of belief for individual hypotheses. Shafer provides several measures. The most important ones are the Bel function that indicates the lower bound of belief and the Pl plausibility function that indicates the upper bound. This experiment uses both of these to compare the Dempster-Shafer system's results with the human subjects' guilt estimates.

4 Comparisons of behavior

One way in which the humans did not follow probabilistic principles was that they usually changed only the probability of the suspect directly impugned by the clue without making compensatory changes to the other suspects. Because probability requires that the sum of the probabilities over suspects equals 1.0, each change must be balanced with an equal change in the opposite direction for the other suspects. Bayesian probability requires proportionately equal changes in the non-impugned hypotheses. Robinson and Hastie termed their subjects' omission of this required compensation "non-compensatory probability updating."

Although subjects in general did not compensate, there were two conditions under which they did, at least partially: 1) when a clue had an extreme impact on the guilt of one suspect, and 2) after a large number of clues had already been processed. In the first case—extreme impact—the clue often contradicted prior belief: that is, it indicated that a subject's favorite suspect was actually innocent or that

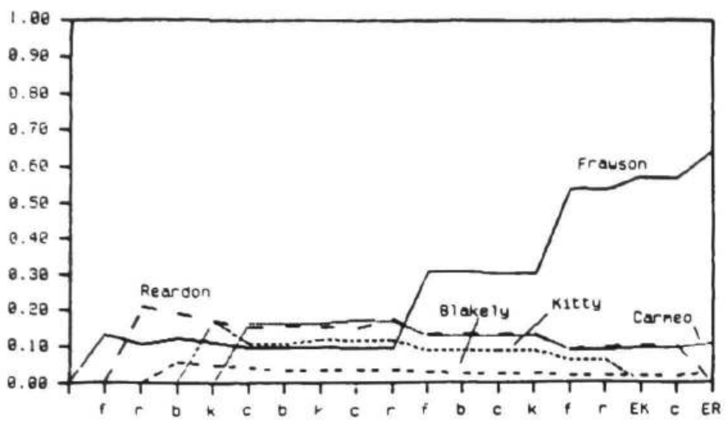

Figure 4: Dempster-Shafer Bel score guilt ratings for each suspect after each clue.

a long-shot was guilty.

Figure 4 shows the Dempster-Shafer system's Bel function assignment of guilt to each suspect after each clue. Each point on the abscissa is a separate clue. The first initial of the suspect mentioned by the clue marks each point. Some of the clues indicate guilt, some indicate innocence, and some are neutral. Like the human subjects, most of the Dempster-Shafer system's belief changes were non-compensatory. The only strong demonstration of compensation occurred in reaction to the two strong clues regarding Frawson and the clues eliminating Reardon and Kitty. This behavior corresponds exactly with Robinson and Hastie's descriptions of their subjects.

As an explanation for the humans' behavior, this result suggests that the humans may not fully partition their belief before collecting evidence. The humans may take an approach that is analogous to the Dempster-Shafer system's approach: that is, slowly portion out belief and only when there is a preponderance of evidence for one hypothesis do they take back belief from other hypotheses. We are not, however, claiming that the Dempster-Shafer system literally models the individual subjects. There was far too much variance between subjects to even attempt to analyze the system's ability to model the individuals.

Robinson and Hastie found two other aspects of the human data that conflicted with probability: probabilities usually added to over one—"super-additivity"—and some humans sometimes gave non-zero probability ratings to suspects after giving them zero ratings—"resuscitation."

If the Dempster-Shafer upper-bound probability measure Pl is used, then most of the probabilities add to over one thus qualitatively modeling the super-additivity. The interpretation in this case is that the subjects were sensitive to their residual uncertainty about the suspects and felt that the lower-bound estimates made them look overly convinced of the suspect's innocence. Assigning probabilities as low as the Bel scores in figure 4 might look like an admission of implausibility that did not correspond to the subjects' beliefs. The scoring mechanism did not give the subjects any way to indicate the suspects' potential guilt, and, therefore, the subjects may have blended the plausibility score with the actual belief.

Because the Bel measure naturally increases from zero to some non-zero value as evidence is collected, the Bel score could model the resuscitation. To use this explanation in conjunction with the super-additivity explanation requires the assumption that the subjects were somehow sensitive to both the Bel and Pl and chose to respond in some hybrid manner that sometimes allowed the Bel score to override the Pl score.

Simultaneously using the Bel and Pl measures to explain the human behavior is not adequately convincing especially because there are other possible explanations for the super-additivity and resuscitation. Explanations based on the input scale and human understanding seem more appealing than an explanation based on the Dempster-Shafer theory. For example, there are some problems with Robinson and Hastie's method of soliciting probability ratings. They used a scale marked into 0.05 probability intervals. The subjects may not have realized that position within an interval was significant. This explanation could explain the resuscitation effect because no suspect resuscitated to more than a probability of 0.1.

While the Dempster-Shafer system does provide an explanation of these anomalies, we feel that the major contribution of this work is to propose a behavior with which decision system users could empathize. The similarity is especially strong for the non-compensatory behavior. Because the Dempster-Shafer system is theoretically sound, system developers can feel secure using it.

5 Conclusion

If, as previous work has suggested [4, 11, 14, 28, 25], user acceptance depends on the ability of the user to empathize with system behavior, and if humans are particularly poor at understanding Bayesian probabilistic notions, then this result showing the similarity between human and Dempster-Shafer updating behaviors encourages further exploration of the use of the Dempster-Shafer theory in automated reasoning. These results combined with the Dempster-Shafer theory's theoretical soundness and Mitchell's results [20, 21] concerning acquisition and computational requirements are a strong argument in favor of the Dempster-Shafer theory.

References

[1] Tversky A. and D. Kahneman. Causal schemas in judgments under uncertainty. In M. Fishbein, editor, *Progress in Social Psychology*, volume 1, pages 49–72. Lawrence Erlebaum, Hillsdale, NJ, 1980.

[2] AAAI. *Uncertainty in Artificial Intelligence*, 1987.

[3] J. A. Barnett. Computational methods for a mathematical theory of evidence. In *International Joint Conference on Artificial Intelligence*, pages 868–875, Vancouver, British Columbia, Canada, 1981.

[4] H. L. Bleich. The computer as a consultant. *New England Journal of Medicine*, 284(3):141–147, 1971.

[5] B. G. Buchanan and E. H. Shortliffe. *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*. Addison-Wesley, Reading, MA, 1984.

[6] P. Cheeseman. In defense of probability. In *International Joint Conference on Artificial Intelligence*, pages 1002–1009, 1985.

[7] Bruce D'Ambrosio. A hybrid approach to reasoning under uncertainty. *International Journal of Approximate Reasoning*, 2:29–45, 1988.

[8] R. Dawes. The robust beauty of improper linear models in decision making. *American Psychologist*, 34:571–582, 1979.

[9] Ward Edwards. Conservatism in human information processing. *Nature*, 32:414–416, 1971.

[10] B. Fischhoff and R. Beyth-Marom. Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, 90(3):239–260, 1983.

[11] R. B. Friedman and D. H. Gustafson. Computers in clinical medicine, a critical review. *Computers and Biomedical Research*, 10:199–204, 1977.

[12] J. Gordon and E. Shortliffe. A method for managing evidential reasoning in a hierarchical hypothesis space. *Artificial Intelligence*, 26:323–357, 1985.

[13] J. Gordon and E. H. Shortliffe. The Dempster-Shafer theory of evidence. In B. G. Buchanan and E. H. Shortliffe, editors, *Rule-Based Expert Systems: The MYCIN Experiments of the Stanford Heuristic Programming Project*, pages 272–292. Addison-Wesley, Reading, MA, 1984.

[14] G. A. Gorry. Computer-assisted clinical decision-making. *Meth. Inform. Med.*, 12(1):45–51, 1973.

[15] R. C. Hughes and J. N. Maksym. Acoustic signal interpretations: Reasoning with non-specific and uncertain information. *Pattern Recognition*, 18(6):475–483, 1985.

[16] Daniel Kahneman, Paul Slovic, and Amos Tversky, editors. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge University Press, New York, 1982.

[17] Jin H. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference systems. In *International Joint Conference on Artificial Intelligence*, pages 190–193, 1983.

[18] Lemmer and Kanal, editors. *Uncertainty in Artificial Intelligence: the 1985 workshop*. AAAI, North-Holland, 1985.

[19] Lemmer and Kanal, editors. *Uncertainty in Artificial Intelligence: the 1986 workshop*. AAAI, North-Holland, 1986.

[20] Donald H. Mitchell. *Automated Decision Support Using Variations on the Dempster-Shafer Theory*. PhD thesis, Northwestern University, 1987.

[21] Donald H. Mitchell, Steven A. Harp, and David K. Simkin. A knowledge-engineer's comparison of three evidence aggregation methods. In *Second Workshop on Uncertainty in AI*, pages 297-304. AAAI, Morgan Kauffman, 1987.

[22] J. Pearl. On evidential reasoning in a hierarchy of hypotheses. *Artificial Intelligence*, 28:9–15, 1986.

[23] L. B. Robinson. *In the Footsteps of Sherlock Holmes: Information Search and Hypothesis Testing*. PhD thesis, Northwestern University, 1986.

[24] L. B. Robinson and R. Hastie. Revision of beliefs when a hypothesis is eliminated from consideration. *Journal of Experimental Psychology: Human Perception and Performance*, 11(4):443–456, 1985.

[25] William B. Rouse and Nancy M. Morris. Understanding and enhancing user acceptance o computer technology. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(6):965–973, 1986.

[26] Glenn Shafer. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[27] P. Slovic and S. Lichtenstein. Comparison of Bayesian and regression approaches to the study of information processing judgement. *Organizational Behavior and Human Performance*, 6:649–744, 1971.

[28] T. S. Startsman and R. E. Robinson. The attitudes of medical and paramedical personnel toward computers. *Computers and Biomedical Research*, 5:218–227, 1972.

[29] A. Tversky and D. Kahneman. Causal schemas in judgments under uncertainty. In Kahneman, Slovic, and Tversky, editors, *Judgment Under Uncertainty: Heuristics and Biases*, pages 117–128. Cambridge University Press, 1982.