1987

A Comparison of Two Evidential Reasoning Schemes

Sunny Lee

Report Number:

87-672

A COMPARISON OF TWO EVIDENTIAL REASONING SCHEMES

Sunny Lee

CSD-TR-672
April 1987

A COMPARISON OF TWO EVIDENTIAL REASONING SCHEMES

Sunny Lee

Department of Computer Science

Purdue University

West Lafayette, IN 47907

ABSTRACT

[1] advocate the use of Dempster-Shafer (D-S) theory in evidence-gathering process. It is stated that they are unaware of any formal model which could allow inexact reasoning at whatever level of abstraction. [2] later shows how evidential reasoning can be conducted in the same hypothesis space using a Bayesian model. The purpose of this note is to examine the difference of theses two schemes, and to point out certain inconsistencies of this Bayesian model with the motives behind the use of D-S model.

1. INTRODUCTION

Providing a framework for evidential reasoning in an AI system has become an important task. In a recent publication, "A Method of Managing Evidential Reasoning in a Hierarchical Hypothesis Space" [1], Gordon and Shortliffe (G-S) advocate the use of the Dempster-Shafer (D-S) theory in evidence-gathering process. They adapt the D-S theory to evidential reasoning in a tree-structured hierarchy of hypotheses. It is stated that they are unaware of any formal model which could allow inexact reasoning at whatever level of abstraction. Pearl [2] later shows how evidential reasoning can be conducted in the same hypothesis space using a Bayesian model. The purpose of this note is to examine the difference of these two approaches, and to point out certain inconsistencies of this Bayesian model with the motives behind the use of D-S model.

2. THE DOMAIN

Suppose an expert is diagnosing a case X. The diagnostic hypothesis set $H = \{h_1, \ldots, h_n\}$ are known to be mutually exclusive and exhaustive. Any subset of hypothesis in H gives rise to a new hypothesis which is assumed to be of semantic interest. The diagram can be viewed as a n-layer graph, with H as the root on the top layer and the individual hypotheses as the leaves in the nth layer and each intermediate nodes stands for the disjunction of its immediate successors. For illustrative purposes, Figure 1 shows the graph corresponding $H = \{h_1, h_2, h_3, h_4\}$.

Although [1] [2] both constrain the graph into a tree, it is clear later that this setting would be more convenient for the purpose of comparison without the necessity of introducing additional approximation step. In [2], the 3-step evidence aggregation technique is apparently not restricted to a tree-structured hierarchy of hypothesis. In [1], discussions of the Dempster-Shafer theory are in fact based on a layer graph instead of a tree. The purpose of restricting a graph into a tree is for computational purposes where G-S has to introduce approximation steps for the D-S theory to overcome the combinational computation.

Several motives which have led G-S to advocate the use of the Dempster-Shafer theory are listed below:

(i) Evidence partially in favor of a hypothesis should not be construed as evidence partially against the same hypothesis (i.e. in favor of its negation).

(ii) Model allows an inexact reasoning at whatever level of abstraction is appropriate for the evidence that has been gathered.

(iii) Ability to distinguish between uncertainty, or lack of knowledge and equal certainty.

3. BRIEF DESCRIPTIONS OF BOTH SCHEMES

3.1 D-S Model

There are two functions in the D-S model: one is called bpa - basic probability assignment, another is called belief function. A bpa, denoted m, assigns a number in [0,1] to every node in the layer graph such that these numbers sum to 1. A belief function, denoted Bel, corresponding to a specific bpa, m, assigns to every node of the layer graph the sum of m-values of every descendant (including itself).

Example 1.

An expert may have a rule knowledge:

R1: Evidence $e1$, confirms h_1 to the degree 0.2

$m(h_1) = 0.2, \;\; m(H) = 0.8$

$\mathrm{Bel}(A) = 0.2$ if $h_1 \in A \neq H$, $\mathrm{Bel}(H) = 1$

R2: Evidence $e2$, disconfirms $\{h_2, h_3\}$ to the degree 0.8

$m(\{h_2, h_3\}^c) = m(\{h_1, h_4\}) = 0.8, m(H) = 0.2$

$\mathrm{Bel}\,(A) = 0.8$ if $\{h_1, h_4\} \subset A \neq 0$, $\mathrm{Bel}\,(H) = 1$

These two examples allow one to address whatever level of abstraction which is appropriate for the gathered evidence, and illustrate (ii) effectively. Examining R1, it says nothing about the complement (or negation) of h_1 which is $\{h_2, h_3, h_4\}$. It is clear that $m(\{h_2, h_3, h_4\}) = 0$, Bel $(\{h_2, h_3, h_4\}) = 0$. This captures (i) very well. Occasionally, it is hard to think why 0.2 in R1 gives no hint on the negation of h_1. The following argument will be useful: Suppose a textbook says there are five pieces of evidence to characterize h_1. However, the relationship of these five pieces of evidence with other h_i $i \neq 1$ is complex and not clear. The expert who has read this textbook may give the knowledge engineer a rule that if a piece of evidence observed then confirms h_1 to the degree 0.2. Under this circumstance, he would be reluctant to tell the knowledge engineer anything more of other associations.

3.2 A Baysian Model

Initially, each singleton hypothesis h_i is assigned a probability in [0,1], denoting p, such that $\sum_{i=1}^{n} p(h_i) = 1$. These $p(h_i)$ are suggested to reflect the probability that h_i is true given all previous evidence. The probability in each intermediate node is the sum of the probability of its constituents i.e. $P(A) = \sum_{h_i \in A} p(h_i)$.

The most important identity, a definition, introduced in [2] is

$$\lambda_s = \frac{p(e\,|S)}{p(e\,|S^c)}$$

where S^c is the complement of S with respect to H, e is a new piece of evidence. The expert is required to estimate λ_s where the hypothesis S is of his interest. λ_s is the degree to which the evidence confirms or disconfirms S. If $\lambda_s > 1$ then it is called confirmation, while if $\lambda_s < 1$ it is called a disconfirmation.

Example 2.

An expert may have a rule of knowledge as

R3: Evidence $e1$ confirms h_1 to the degree 4

$$S = \{h_1\};\ \ \lambda_s = 4$$

R4: Evidence $e2$ disconfirms $\{h_2, h_3\}$ to the degree 0.25

$$S = \{h_2, h_3\};\ \ \lambda_s = 0.25$$

R5: Evidence $e3$ disconfirms h_1 to the degree 0.2

$$S = \{h_1\};\ \ \lambda_s = 0.2$$

The above examples can address whatever level of abstraction (i.e. any hypothesis set, not necessarily restricted to singleton hypothesis) which is appropriate for the gathered evidence, and also illustrates (ii) well.

4. THE COMPARISON

From above, one can see that the proposed Bayesian model illustrates motive (ii) effectively due to the introduction of λ_s. However we will point out that the framework is inconsistent with other motives. Next we will examine the effect of combining two rules applying to a singleton hypothesis.

4.1 Motives Requirement

"Confirmation" or "disconfirmation" used in [1] [2] mean different things. In [1], the usage has a logical semantics, while the usage in [2] is only for convenience, which might occasionally be misleading. As an example, one may have the following rule:

R6: Evidence *e*6 *confirms that Dr. Shortliffe is in his office* to the degree 0.6.

Equivalently, one could say

Evidence *e*6, *disconfirms that Dr. Shortliffe is not in his office* to the degree 0.6.

When one performs computation, statements involved "disconfirms" are always converted to "confirms". In contrast to this, the rule in [2] may be written as:

R7: Evidence *e*6 *confirms that Dr. Shortliffe is in his office* to the degree 4.

However one could not translate it into

R7': Evidence *e*6, *disconfirms that Dr. Shortliffe is not in his office* to the degree 4.

Because one could figure out the degree of belief that Dr. Shortliffe is not in his office based on the degree that Dr. Shortliffe is in his office. Recall that

$$\lambda_S = \frac{p(e \mid S)}{p(e \mid S^c)}.$$

It is easy to see that

$$\lambda_{S^c} = \frac{p(e \mid S^c)}{p(e \mid S)} = \frac{p(e \mid S^c)}{p(e \mid S^c)}.$$

Therefore, a correct way would be

Evidence $e6$ *disconfirms that Dr. Shortliffe is not in his office* to the degree 0.25.

It is clear that "confirm" or "disconfirm" is used for convenience. One might as well use the term below

Based on evidence $e6$, the degree of the belief that Dr. Shortliffe is in his office is 4.

It is apparent that evidence in favor of h is used to partially support the negation of h as opposed to what author [2] claims - that it would not be construed for other purposes. Although [2] claims that the expert is not required to apply any conscious effort whatsoever regarding other propositions in the system, the formula about λ_s is actually used in the computational process to derive the updating scheme. If λ_s were not equal to the inverse of λ_{s^c} then there will be inconsistency in the framework.

As for the motive (iii), this is one of the properties that a Bayesian model generally does not have. For instance, the situation that a piece of evidence provides equal certainty on S and S^c, i.e $\lambda_S = 1$ will be treated as if no evidence (or lack of evidence) at all in view of the updating scheme.

4.2 The Effect of Combining Two Rules

It is now for us to write down the belief updating scheme [2] as follows:

$$p(h_i \mid e_1 \cdots e_n) = \alpha(S_1 \cdots S_n) W_i(e_1, \ldots, e_n) p(h_i)$$

where e_i bears directly on S_i and

$$W_i(e_1 \cdots e_n) = W_i^1 \, W_i^2 \cdots W_i^n,$$

where

$$W_i^k = \begin{cases} \lambda_{S_k}, & \text{if} \quad h_i \in S_k \\ 1, & \text{if} \quad h_i \in S_k^c \end{cases}$$

and $\alpha(S_1 \cdots S_n)$ is a normalizing factor.

The formula could be understood by examining one piece of evidence e_1, which bears directly on S_1 and the degree of belief is λ_1. To update the probability, first one must multiply λ_1 to probability of every hypothesis in S_1 and keep the probabilities of other hypothesis the same. Next, one has to normalize the sum of the probabilities to 1. The formula also says that one could delay the normalization.

We now discuss the effect of combining two rules. One notes that there is no need to obtain any *apriori* belief in each singleton hypothesis in D-S model as required in Baysian Model. As said in [1], there are three categories when rules are applying to singleton hypothesis.

(1) they may both confirm or both disconfirm the same hypothesis,

(2) one may confirm and the other may disconfirm the same hypothesis,

(3) each may bring evidence to bear on different competing hypothesis.

Note that T1,T2 below stand for rules in D-S model while K1, K2 stand for rules in this Bayesian model

Category 1

T1: confirms h_1 to the degree s_1
T2: confirms h_1 to the degree s_2

K1: belief in h_1 to the degree λ_1
K1: belief in h_1 to the degree λ_2

For D-S model, the combined effect is that one can replace T1, T2 by a rule "confirms h_1 to the degree $s_1 + s_2 - s_1 s_2$". One observes that $s_1 + s_2 - s_1 s_2 \geq \max(s_1 \, s_2)$, therefore it always increases the belief in h_1.

For Baysian Model, through simple computation, the combined effect is that K1, K2 can be replaced by a rule "belief in h_1 to the degree $\lambda_1\ \lambda_2$". In this category, λ_1, λ_2 are both necessarily greater than 1, therefore the confirmation in h_1 is also strengthened as that in D-S model. However, the belief in other singleton hypothesis is necessarily reduced.

Category 2

T1: confirms h_1 to the degree s_1
T2: disconfirms h_1 to the degree s_2

K1: confirms h_1 to the degree λ_1
K2: disconfirms h_1 to the degree λ_2

For Baysian Model, K1 and K2 can be replaced by "belief in h_1 to the degree $\lambda_1\ \lambda_2$" as the same in Category 1. It is evident that the strength of belief in h_1 depends on the magnitude of $\lambda_1\ \lambda_2$. It confirms h_1 and reduces the support for $\{h_2\ h_3\ h_4\}$ when $\lambda_1\ \lambda_2 > 1$. It disconfirms h_1 and increases the support for $\{h_2\ h_3\ h_4\}$ when $\lambda_1\ \lambda_2 < 1$. If $\lambda_1\ \lambda_2 = 1$ then nothing changes as if there were no evidence at all. It is noted that this behaves like CF model used in MYCIN system.

To compute the net effect based on T1 and T2, one has to convert T2 to T2′ - confirms h_2, h_3, h_4 to the degree s_2. No single rule in this format can replace T1, T2. However, the result is that the supports in h_1 and $\{h_2, h_3, h_4\}$ provided by each are both reduced after combination (see [1]).

Category 3

T1: confirms h_1 to the degree s_1
T2: disconfirms h_2 to the degree s_2

K1: confirms h_1 to the degree λ_1
K1: disconfirms h_2 to the degree λ_2

For D-S model, the belief in h_1 entails the belief in $\{h_1, h_3, h_4\}$ which in turn strengthens T2. However, the belief in $\{h_1, h_3, h_4\}$ does not say anything about individual elements which therefore does not affect the confirmation in h_1.

For Baysian Model, the updates is as follows:

$$
\begin{aligned}
p(h_1|e) &= \alpha\, \lambda_1 p(h_1) \\
p(h_2|e) &= \alpha\, \lambda_2 p(h_2) \\
p(h_3|e) &= \alpha\ \ p(h_3) \\
p(h_4|e) &= \alpha\ \ p(h_4)
\end{aligned}
$$

where $\alpha = [\lambda_1 p(h_1) + \lambda_2 p(h_2) + 1 - p(h_1 \cup h_2)]^{-1}$

One sees that

$$
\begin{aligned}
\alpha\lambda_1 &= \frac{\lambda_1}{\lambda_1 p(h_1) + \lambda_2 p(h_2) + 1 - p(h_1 \cup h_2)} \\
&> \frac{\lambda_1}{\lambda_1 p(h_1) + \lambda_1 p(h_2) + 1 - p(h_1 \cup h_2)} \ \text{(since } \lambda_2 < 1) \\
&= \frac{\lambda_1}{\lambda_1 p(h_1 \cup h_2) + 1 - p(h_1 \cup h_2)} \\
&> 1 \ \text{(since } \lambda_1 > 1)
\end{aligned}
$$

and

$$
\begin{aligned}
\alpha\lambda_2 &= \frac{\lambda_2}{\lambda_1 p(h_1) + \lambda_2 p(h_2) + 1 - p(h_1 \cup h_2)} \\
&< \frac{\lambda_2}{\lambda_2 p(h_1) + \lambda_2 p(h_2) + 1 - p(h_1 \cup h_2)} \\
&= \frac{\lambda_2}{\lambda_2 p(h_1 \cup h_2) + 1 - p(h_1 \cup h_2)} \\
&< 1 \ \text{(since } \lambda_2 < 1)
\end{aligned}
$$

Now, α could be any value depending on λ_1, λ_2, $p(h_1)$ $p(h_2)$. Thus, one can summarize the result as below.

Confirmation in h_1 and disconfirmation in h_2 still holds. However, the degree of "confirm and disconfirm" may get strengthened or weakened. The effect on other hypothesis may get increased or decreased.

5. CONCLUSION

With regard to computational aspect, this Bayesian model [2] has obviously provided a simple and convenient framework. However, it is inconsistent with some of the motives which have led G-S to advocate the use of D-S model. It is evident that each technique could find its proper domains but not used in the same domain.

REFERENCES

1. Gordon, J. and Shortliffe, E.H., A method of Managing evidential reasoning in a hierarchical hypothesis space, *Artificial Intelligence* **26** (1985) 323-357.

2. Pearl, J., On evidential reasoning in a hierarchy of hypotheses, *Artificial Intelligence* **28** (1986) 9-15.

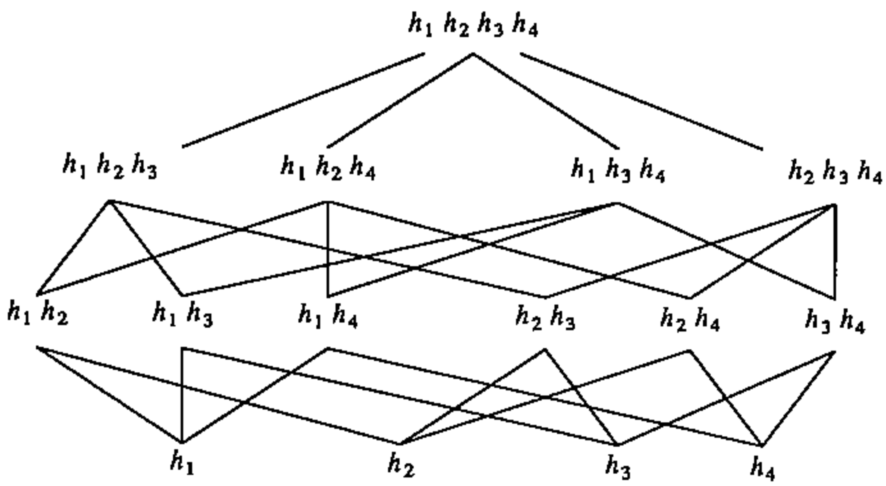

Figure 1: Each node stands for disjunction of h_i's. For instance $h_1\ h_2 = \{h_1\ ,\ h_2\}$.