Combining Bodies of Dependent Information

Robert Hummel
Courant Institute of Mathematical Sciences

Larry M. Manevitz
Courant Institute and Bar Ilan University

New York University
251 Mercer, New York, NY 10012 USA

Abstract

Recently, Hummel and Landy proposed a variation on the Dempster/Shafer theory of evidence that tracks only the first and second order statistics of the opinions of sets of experts. This extension permits the tracking of statistics of probabilistic opinions, however, as opposed to tracking merely Boolean opinions (or possibilities within the "frame of discernment"). Both the Dempster/Shafer formulation and the Hummel/Landy formulation assume that bodies of experts that are combined to form new statistics have independent information. We give a model for parameterizing degree of dependence between bodies of information, and extend the Hummel/Landy formulation for combining evidence to account for sets of experts having dependent information sources.

1. Background

Many systems using artificial intelligence concepts must combine information from disparate sources of knowledge to make a decision. Often the information that is given is incomplete: evidence is accumulated suggesting one alternative or another, but in a quantitatively inconclusive way. The use of purely Bayesian techniques sometimes encounters difficulties, due to the lack of sufficient information. There are thus many different ways that have been proposed for combining evidence. One method, called here the Dempster/Shafer theory of evidence [1], has received considerable interest and some use in experts systems.

Central to the Dempster/Shafer theory, and several other formulations of combination of evidence, is a way of handling uncertainty in propositions. Rather than assigning probabilities to possible labels (from a "frame of discernment"), these theories attempt to assign degrees of confidence to the various propositions. In Shafer's explanation of the Dempster/Shafer theory of evidence, this is done through the use of "belief functions" to assign weights to subsets of labels in their theory. In general, there is a set of possible labels, and a set of numbers representing a current state of belief. When additional information is obtained, the numbers are changed to a new state. Each state is associated with the body of evidence obtained to that point, and the updating method represents the combination of the current body of evidence with the incremental evidence.

For example, for medical diagnosis applications, a patient can have one of a set of possible diseases. Evidence is obtained in the form of symptoms and test results. Given a current set of symptoms and results, a doctor might decide to run an additional test, and update the assessment of the patient's condition based on the results, in conjunction with the information already present.

In the theory of "belief functions," a state is represented by a probability distribution over the power set of the set of possible labels. Thus a number is assigned to every subset of labels. New evidence is represented, in the Dempster/Shafer theory, by a new state, also assigning a number to every subset. The Dempster rule of combination [2] is used to combine these two states to form a new state.

Other possibilities include "Bayesian" approaches, for which a state is generally represented as a probability distribution over the set of possible labels. Each value is regarded as a "subjective" or "inferential" probability, and the use of Bayes' formula in the presence of various independence or simplifying assumptions is defended by a body of research and results, especially those developed by Good, Savage, De Finitti, and Ramsey *et al.* A survey treatment is given in [3].

In a recent work by Hummel and Landy [4,5], it is shown that the Dempster/Shafer formulation is equivalent to the tracking of statistics of sets of experts expressing Boolean opinions over the set of labels. The sets of experts update by combining, using Bayesian updating, in pairs over the product space of experts. Further, an alternate formulation is suggested. In this formulation, experts have probabilistic opinions, rather than Boolean opinions, and the state of the system is represented by the mean and covariance (of logarithmic values) of those opinions.

A drawback of both developments is that conditional independence assumptions regarding the sources of information must be imposed. These assumptions are often unrealistic. Roughly speaking, the assumptions state that the probability of a particular piece of information in the context of a single label, given some existing information, is the same as the probability without the existing information. Independence is defined in terms of probabilities taken over the set of all labeling situations. Consider, for example, the case of a medical diagnosis example. Assume that s_1 represents the set of symptoms and information obtained to date, and J_2 is the new information. What is required is that the probability of the existence of the symptoms s_1 amongst the set of all patients having a given disease λ must be the same as the probability of the existence of the same set of symptoms s_1 among the set of patients having the symptoms s_2 and the disease λ. Further, this equivalence must hold for all diseases λ. In essence, this says that information about the symptoms s_2 yield no information as to the probability of the symptoms s_1, in the presence of any given disease. Since symptoms generally have a common basis, this assumption rarely holds.

In fact, the independence assumption is not very realistic for most applications. It is required to justify the updating formulas, and is so predominant in most formulations for the combination of evidence that the limitations are generally overlooked.

In this paper, we introduce a model for measuring a "degree of independence" between sets of information. The degree of independence is measured by a single variable α, $0 \leq \alpha < \infty$, which can in turn depend upon the information values (the symptoms), i.e., $\alpha = \alpha(s_1, s_2)$. We then extend the Hummel/Landy formulation for the combination of information to the case where the information it a-independent. The case $\alpha = 1$ will correspond to the same independence assumption as before. The case $\alpha = 0$ corresponds to completely dependent information, in the sense that s_1 implies the information J_2.

Moreover, we develop formulti to that the statistics are taken over the union of the sets of experts, rather than over the product space. We find the use of the product sets of experts "less natural" than simply combining all experts into one collection. The difficulty, of course, is that when combining the experts into one collection, each expert must be required to update his opinion based on some other opinion, and it is not *a priori* specified which other opinion should be used. We suggest a solution.

2. Formulation

Let $\mathcal{E}$ be a set of experts. Each expert $e \in \mathcal{E}$ is privy to a body of information (symptoms) about the current situation. We denote by *t* the information shared by the associated experts. The goal is to label the current situation (i.e., the current patient) with a label X from the set of possible labels A. It is assumed that A is mutually exclusive and exhaustive. Each expert $e \in \mathcal{E}$ evaluates the information *t* and assigns a probability distribution to the set of possible labels A, represented by the set of values $p_s(e,\lambda)$. The average opinion, computed by taking a mean over all $e \in \mathcal{E}_s$ is denoted by $\mu_s(\lambda)$. Likewise, the covariance values are given by the formula

$$C_s(\lambda_1,\lambda_2) = \underset{e\in\mathcal{E}}{\mathrm{Avg}}\left[(p_s(e,\lambda_1)-\mu_s(\lambda_1))\cdot(p_s(e,\lambda_2)-\mu_s(\lambda_2))\right].$$

Logarithmic opinions are denoted by $y_s(e,\lambda)$, and are given by the formula

'<•))

where Prob(X) is a prior probability of label X over all situations (e.g., the probability of the given disease among all patients). The value c, is an indeterminate constant, meaning that the y_s values are defined only to within an additive constant independent of X and of *e*. Means and covariances of the *y't* are also defined, yielding means and covariances of the logarithms, and are denoted by $u^{(f)}$ and $C^{(/)}$ respectively. The use of the logarithmic opinions simplifies the formulas, and is suggested in [6].

Now, suppose that we have two collections of experts *C* and £$_2$ And thus two bodies of information s\ and s_2. We wish to combine the information μ_{s_1}, μ_{s_2}, C_{s_1}, and C_{s_2} to obtain a new mean u.#1^2 10s2 * new covariance C_{n9n}. Similarly, we should combine the means and covariances of logarithms.

In Hummel and Landy [4], the formula is given for the log's, with complete independence. The formulas are:

$$\mu^{(l)}_{s_1\oplus s_2}(\lambda) = \mu^{(l)}_{s_1}(\lambda) + \mu^{(l)}_{s_2}(\lambda),$$

$$C^{(l)}_{s_1\oplus s_2}(\lambda_1,\lambda_2) = C^{(l)}_{s_1}(\lambda_1,\lambda_2) + C^{(l)}_{s_2}(\lambda_1,\lambda_2).$$

These formulas are derived by assuming that updating takes place using the set of all committees of two consisting of one expert from $\mathcal{E}_1$ and one expert from $\mathcal{E}_2$ (i.e., the product space), and that within each committee, Bayesian updating is used with a conditional independence assumption.

3. Unions of experts and a -independence

Conditional independence, used in the formulas above, assert that $\mathrm{Prob}(s_2|s_1,\lambda) = \mathrm{Prob}(s_2|\lambda)$ for all λ, where the probabilities are taken over the set of all labeling situations (e.g., patients), and not over the experts.

We now define the information s_1 and s_2 to be $\alpha(s_1,s_2)$ independent if:

$$\mathrm{Prob}(s_2|s_1,\lambda) = \left[\mathrm{Prob}(s_2|\lambda)\right]^{\alpha(s_1,s_2)}$$

for all λ. It is important to realize that a-independence is not symmetric, that $\alpha(s_1,s_2) \neq$ in $\alpha(s_2,s_1)$ general. Note that for a - 1, the assumption reverts to conditional independence. For a=0, we have that the information s_1 implies (with probability one) the information i_2. The case $\alpha > 1$ corresponds to negative evidence. The existence of such an a constitutes an assumption, and is not a completely general measure of independence or dependence. Specifically, we are assuming that $\alpha(s_1,s_2)$ is independent of λ. This is a strong assumption, but is not as strong as the assumption of independence.

One possibility for obtaining the $\alpha(s_1,s_2)$'s would be to poll experts for their estimates. Alternatively, a might be obtained from the formula:

$$\alpha(s_1,s_2) = \underset{\lambda\in\Lambda}{\mathrm{Avg}}\left\{\frac{\log(\mathrm{Prob}(s_2|s_1,\lambda))}{\log(\mathrm{Prob}(s_2|\lambda))}\right\}.$$

Of course, these kind of joint statistics are often hard to obtain.

In the presence of a(s_1,,s_2)-indepcndence, it is not hard to show that log-probabilities now update according to the formula

$$y_{s_1\oplus s_2}((e_1,e_2),\lambda) = y_{s_1}(e_1,\lambda) + \alpha(s_1,s_2)\cdot y_{s_2}(e_2,\lambda).$$

The new updating formulas become

$$\mu^{(l)}_{s_1\oplus s_2}(\lambda) = \mu^{(l)}_{s_1}(\lambda) + \alpha(s_1,s_2)\cdot\mu^{(l)}_{s_2}(\lambda),$$

$$C^{(l)}_{s_1\oplus s_2}(\lambda_1,\lambda_2) = C^{(l)}_{s_1}(\lambda_1,\lambda_2) + \alpha^2(s_1,s_2)\cdot C^{(l)}_{s_2}(\lambda_1,\lambda_2).$$

Obviously, the formulas have changed very little: the updating term giving the new information is simply weighted by the degree of independence of the new information. While this idea is fundamentally simple, we have justified the use of this weighted updating through a precise definition of a-independence.

We now introduce a second new concept to the formulation, that of union-based combination. In the formulas from the previous section, the information in $s_1\oplus s_2$ is based on the product space of experts $\mathcal{E}_1\times\mathcal{E}_2$. We find it more desirable to base formulas on the union space of experts $\mathcal{E}_1\cup\mathcal{E}_2$. The resulting set of information, $s_1\odot s_2$, contains updated opinions for each expert $e\in\mathcal{E}_1\cup\mathcal{E}_2$. However, we must specify the manner in which each expert performs the updating, since there are no longer obvious pairs of opinions for each resultant expert. The method advocated here is to have the experts in $\mathcal{E}_1$ update in a Bayesian fashion based on the mean opinion from the set $\mathcal{E}_2$. Likewise, the experts in $\mathcal{E}_2$ update using the information obtained from the mean opinion of experts in $\mathcal{E}_1$. This changes the component formulas, so that, for $e\in\mathcal{E}_1$,

$$y_{s_1\odot s_2}(e,\lambda) = y_{s_1}(e,\lambda) + \alpha(s_1,s_2)\cdot\mu^{(l)}_{s_2}(\lambda).$$

Likewise, for $e\in\mathcal{E}_2$, we have the same formula, but with all occurrences of s_1 and s_2 exchanged. In particular, we make use of the value a(s_2,si)-

The combination formulas for the means and covariances over the union of the sets of experts can be calculated, and now depends on the number of experts in the component sets, $|\mathcal{E}_1|$ and $|\mathcal{E}_2|$. For the log-probabilities, with a-independence, the formulas are:

$$\mu^{(l)}_{s_1\odot s_2} = \frac{|\mathcal{E}_1|\mu^{(l)}_{s_1} + |\mathcal{E}_2|\mu^{(l)}_{s_2} + \alpha(s_1,s_2)|\mathcal{E}_1|\mu^{(l)}_{s_2} + \alpha(s_2,s_1)|\mathcal{E}_2|\mu^{(l)}_{s_1}}{|\mathcal{E}_1| + |\mathcal{E}_2|}$$

$$\begin{aligned} C^{(l)}_{s_1\odot s_2}(\lambda_1,\lambda_2) = {} & \rho C_{s_1}(\lambda_1,\lambda_2) + \theta C_{s_2}(\lambda_1,\lambda_2) + \\ & \rho\left[\mu^{(l)}_{s_1}(\lambda_1)+\alpha(s_1,s_2)\mu^{(l)}_{s_2}(\lambda_1)-\mu^{(l)}_{s_1\odot s_2}(\lambda_1)\right] \\ & \times\left[\mu^{(l)}_{s_1}(\lambda_2)+\alpha(s_1,s_2)\mu^{(l)}_{s_2}(\lambda_2)-\mu^{(l)}_{s_1\odot s_2}(\lambda_2)\right] \\ & + \theta\left[\mu^{(l)}_{s_2}(\lambda_1)+\alpha(s_2,s_1)\mu^{(l)}_{s_1}(\lambda_1)-\mu^{(l)}_{s_1\odot s_2}(\lambda_1)\right] \\ & \times\left[\mu^{(l)}_{s_2}(\lambda_2)+\alpha(s_2,s_1)\mu^{(l)}_{s_1}(\lambda_2)-\mu^{(l)}_{s_1\odot s_2}(\lambda_2)\right] \end{aligned}$$

where we have suppressed the A. argument in the first equation, and p = $|\mathcal{E}_1|/(|\mathcal{E}_1|+|\mathcal{E}_2|)$, and $\theta = 1-\rho$.

We see that a state now consists of the mean log-opinion $\mu_j^{(f)}(\lambda)$ for all labels $\lambda \in \Lambda$, the covariance of the opinions $C_j^{(f)}(\lambda_1,\lambda_2)$, and a weight of evidence, $|\mathcal{E}|$, corresponding to the number of experts participating in those opinions. In the case of complete independence, $\alpha(s_1,s_2) = \alpha(s_2,s_1)$ = 1, the above formulas become simple addition of the means and covariances, as in the original Hummel/Landy formulation. However, with the covariance formula, there are additional mixed terms which measure the difference in the mean opinions of the experts in €\ and the experts in $\mathcal{E}_2$.

Many other formulations are possible. For example, we can compute means and covariances of the probabilities instead of the log-probabilities. We omit the formulas here, for lack of space. We could also use only incremental evidence, so that experts in $\mathcal{E}_1$ update using the means of the opinions of $\mathcal{E}_2$, while the experts in $\mathcal{E}_2$ use their actual opinions. This leads to slightly different formulas. Finally, as we have seen, the formulations can be posed for either the product spaces of experts, or the union set of experts: many variations are possible.

4. Discussion

The entire approach of tracking statistics of sets of experts has a number of features to commend it. For example, a belief function as used in the Dempster/Shafer theory of evidence requires the specification of 2^N values, where there are N labels. If only first and second order statistics are tracked, as suggested here, then the number of values needed to specify a state is only $N + (N^2+N)/2$. For large N, this can mean a substantial savings in computational effort needed to update a state.

Further, we at least in principle have replaced the notion of subjective probabilities with objective statistics. These statistics, given sufficient resources, could be measured by, for example, "polling" methods. Our formulas are thus firmly based on objective probability theory, and thus foundationaly secure. Of course, the assumptions are still debatable in the context of any particular application, and the value of $\alpha(s_1,s_2)$ in the a-independence of two sets of evidence is most likely to be a subjective quantity. We expect that, in practice, information will be deemed to be, for example, 0.4-independent, based on subjective criteria. In essence, the subjective component of combination of information has been pushed to a meta-level, where degrees of independence of information sources are estimated, instead of estimating degrees of confidence of labels in the presence of specific information.

The use of a collection of opinions permits a distinction between uncertainty and ignorance. A state of belief consists not of a single opinion, but of a collection of opinions. The state can be represented by a mean opinion, and a measure of the spread (or distribution) of those opinions. The spread measures a degree of uncertainty, since if all opinions are identical, there is a considerable degree of certainty in the single expressed opinion. Updating is done by combining the mean opinions and combining the uncertainties. Basically, the new mean opinion becomes a compromise between the two mean opinions of the composing evidence. Uncertainties are likewise mixed, and generally accumulate. Further, in the presence of dependencies in the information sources, uncertainty increases if the opinions from the two sources of information are divergent.

Finally, we note that the theory makes explicit the dependence on the order in which information is combined. That is, if information s_1,, J_2, •••,*,, are to be combined, the various a values and the outcome of the entire system will depend upon the order in which the information is mixed. The system is neither commutative nor associative, in the presence of the a-independence formulation. This may be realistic, in the sense that decisions are often based on incrementally gaining evidence, and that the interpretation and outcome depends on the order in which information is obtained. A separate expert system could be used to decide on the order in which to combine information.

Acknowledgements

This research was supported by Office of Naval Research Grant N00014-85-K-0077, Work Unit NR 4007006. We thank Michael Landy for useful discussions. Manevitz thanks the Courant Institute for their kind hospitality during his visit to NYU.

References

[1] Shafer, G., *A mathematical theory of evidence,* Princeton University Press, Princeton, N.J. (1976).

[2] Dempster, A. P., "Upper and lower probabilities induced by a multivalued mapping/' *Annals of Mathematical Statistics* 38, pp. 325-339 (1967).

[3] Kyburg, Henry E., *The Logical Foundations of Statistical Inference,* Reidel, Dordrecht (1974).

[4] Hummel, Robert A. and Michael S. Landy, "A viewpoint on the theory of evidence," NYU Robotics Research Report #57 (November, 1985). Submitted to *IEEE Transactions on Pattern Analysis and Machine Intelligence.*

[5] Hummel, Robert and M. Landy, "Evidence as opinions of experts," *Proceedings of the "Uncertainty in AI" Workshop,* pp. 135-143 (August 8-10, 1986).

[6] Charniak, Eugene, "The Bayesian basis of common sense medical diagnosis," *Proceedings of the AAAI,* pp. 70-73 (1983).