Combining Bodies of Dependent Information

by

R. A. Hummel† *and L. M. Manevitz*‡
Courant Institute and Bar Ilan University

Appears in the Proceedings of the Tenth International Conference on Artificial Intelligence

Topic: Reasoning

Track: Science

Keywords: Theory of Evidence, combining evidence, independence

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.

†New York University
251 Mercer Street
New York, NY 10012 USA
hummel@nyu.arpa

‡Bar Ilan University
manevitz@nyu-acf8.arpa

Combining Bodies of Dependent Information

Robert Hummel and Larry Manevitz

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 (pathologies). 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 partly 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 completely 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. To make these points concrete, let us review these notions for the medical diagnosis situation.

The current state is represented by a collection of doctors, each expressing an opinion as to the subset of possible diseases that a particular patient might have. A doctor simply gives a list of possibilities, based on the symptoms and information available to him at that point. All doctors have the same information, but not all doctors have the same opinion (i.e., the same list of possibilities). New information is represented by another collection of doctors (perhaps specialists), each with his own list of opinions. A new state is formed as follows. The new set of "doctors" is the set of all committees of two, with one doctor from the original set, and the other doctor from the new set, the set of specialists. Each of these composite "doctors" forms a new opinion by combining the opinions of the individual doctors within the committee. The new opinion is formed by using a Bayesian updating with the individual opinions, with an independence assumption concerning the source of the information. This means that the new list of possible diseases is formed by intersecting the two component lists; that is, the committee rules out pathologies if either the pathology was originally ruled out or if the specialist rules it out. If there were n doctors in the original set, and m specialists, the new state consists of $n \cdot m$ opinions of the new committees.

The correspondence between the belief function for a given state and the opinions of the collection of experts can be stated easily. Let S be a specified subset of labels. Then the belief on the set S is associated with the percentage of experts that have ruled out labels not in S. That is, we consider the fraction of experts that specify that the subset of possible labels is S or contained in S, and call this the belief on the subset S. It turns out that the resulting updating formula when sets of experts combine as indicated above is precisely the Dempster rule of combination. This equivalence is not surprising, since the original introduction of the Dempster rule of combination was based on notions of statistics of certain measures over measure spaces (the spaces of experts).

From this viewpoint, we see that the Dempster/Shafer theory of evidence requires independence of the sources of knowledge, and that further, there is no distinction made between labels being possible and labels being probable. The former objection was known [6], whereas the latter objection is egregious, in that one of the desired properties of the formulation is that it should represent labels with "fuzzy" degrees of probability.

To handle the latter objection, in [4] an alternate formulation is given. In this formulation, each expert maintains a probability distribution over the set of labels, rather than simply a list of possible labels. The set of numbers used to represent the current state associated with a collection of experts and their opinions are related to the mean probability distribution and the multivariate covariance (standard deviation) of those opinions. In fact, in the Hummel/Landy presentation, the logarithms of the probabilities are used, and so the means and covariances of the log's of the probabilities become the numbers for the state of the system. To combine two collections of experts, the set of all committees of two is again formed. When committees update, Bayes' formula is used to compute a new probability distribution over the set of labels. The mean and covariance of the set of logarithms of the new (probabilistic) opinions are then formed of the set of composite experts (namely, the product set of the two former sets of experts). Precise formulas are given in Section 3.

However, the formulation still requires independence of the information. Roughly speaking, this means that the source of new information, to be combined with the body of existing information, must be unrelated. Further, the independence assumption that is needed is *conditional independence*, for all labels. Independence is defined in terms of probabilities taken over the set of all labeling situations, e.g., the set of all patients. More precisely, assume that s_1 represents the set of symptoms and information obtained to date, and s_2 is the new information. What is required is that the probability of the existence of the symptoms s_1 among 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.

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, 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 is α-independent.

Further, we develop formulas so that the statistics are taken over the union of the sets of experts, rather than over the set of all committees of two. 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.

2. Advantages and Disadvantages

The entire approach of tracking statistics of sets of experts has a number of features to commend it. Although the best formulation for combining evidence will typically depend upon the application, the methods to be described in this paper, and the related but earlier formulas of Hummel and Landy, have a number of advantages over, for example, the Dempster/Shafer formulation.

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. For example, to measure the state that some new information in a medical diagnosis situation should produce, we could actually poll a collection of doctors. Our formulas are thus firmly based on objective probability theory, and thus foundationally 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 α-independence of two sets of evidence is most likely to be a subjective quantity (although we suggest a method for measuring α).

Finally, the notion of tracking statistics, while initially complicated by the presence of many different sample spaces (the doctors, the patients, committees of doctors, etc.), is fundamentally simple and appealing. Specifically, we are saying that a state of belief consists not of a single opinion, but of a collection of opinions, and that the collection of opinions can be measured 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.

On the other hand, our introduction of α-independence brings with it new complications. As mentioned above, the degree of independence between two bodies of information is likely to be decided subjectively. The case $\alpha = 1$ corresponds to complete independence, as is assumed in the Hummel/Landy formulation. The case $\alpha = 0$ corresponds to complete dependence, so that the information s_1 implies the information s_2. In this case, the new evidence can be said to be redundant, so that no updating should take place. Between these two situations, there can be any level of dependence. In fact, it is possible to have $\alpha > 1$, corresponding to negative correlation between the information sources. We expect that, in practice, information will be deemed to be, for example, 0.5-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 and information values are estimated, instead of estimating degrees of confidence and likelihood of the various labels in the presence of specific information.

Finally, we note that the theory makes explicit the dependence on the order in which information is combined. That is, if information $s_1, s_2, \cdots, s_n$ are to be combined, the various α 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 α-independence formulation. While this might be deemed to be a considerable disadvantage of the formulation, it 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.

3. Formulation and Formulas

Let E be a set of experts. Each expert $e \in E$ is privy to a body of information (symptoms) about the current situation. We denote by s the information together with the associated experts. The goal is to label the current situation (i.e., the current patient) with a label λ from the set of possible labels Λ. It is assumed that Λ is mutually exclusive and exhaustive. The expert e's opinion is represented by the set of values $p_s(e,\lambda)$, given information s. The average opinion, computed by taking a mean over all $e \in E$, is denoted by $\mu_s(\lambda)$. Likewise, the covariance values are given by the formula

$$C_s(\lambda_1,\lambda_2) = \operatorname*{Avg}_{e \in E}\left[(p_s(e,\lambda_1) - \mu_s(\lambda_1)) \cdot (p_s(e,\lambda_2) - \mu_s(\lambda_2)) \right]$$

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

$$y_s(e,\lambda) = \log\left[\frac{p_s(e,\lambda)}{\mathrm{Prob}(\lambda)} \right] + c_s$$

where $\mathrm{Prob}(\lambda)$ is a prior probability of label λ over all situations (i.e., the probability of the given disease among all patients). The value c_s is an indeterminate constant, meaning that the y_s values are defined only to within an additive constant independent of λ and of e. Means and covariances of the y's are also defined, yieling means and covariances of log's, and are denoted by $\mu^{(l)}$ and $C^{(l)}$ respectively. The use of the logarithmic opinions simplifies the formulas, and is suggest in [7].

Now, suppose that we have two collections of experts E_1 and E_2 and thus two bodies of information s_1 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 $\mu_{s_1 s_2}$ and a new covariance $C_{s_1 s_2}$. Similarly, we should combine the means and covariances of log's.

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

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

These formulas are derived assuming that updating takes place using the set of all committees of two, and that within each committee, Bayesian updating is used with a conditional independence assumption. Specifically, it is assumed that

$$\mathrm{Prob}(s_2 \mid s_1,\lambda) = \mathrm{Prob}(s_2 \mid \lambda)$$

for all λ, where the probabilities are taken over the set of all "patients," (and not over the experts). The crucial point in the derivation is that the logarithmic probabilities update by adding:

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

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

$$\mathrm{Prob}(s_2 \mid s_1,\lambda) = \mathrm{Prob}(s_2 \mid \lambda)^{\alpha(s_1,s_2)}$$

for all λ. It is important to realize that α-independence is not symmetric, that $\alpha(s_1,s_2) \neq \alpha(s_2,s_1)$ in general. Note that for α=1, the assumption reverts to conditional independence. For α=0, we have that the information s_1 implies (with probability one) the information s_2. The existence of such an α constitutes an assumption, and is not in any way a completely general measure of independence. 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.

We note that α might be measured by polling among many situations, observing when information s_1 and s_2 cooccurs, and measuring

$$\alpha(s_1,s_2) = \operatorname*{Avg}_{\lambda\in\Lambda}\left[\frac{\log(\mathrm{Prob}(s_2 \mid s_1,\lambda))}{\log(\mathrm{Prob}(s_2 \mid \lambda))}\right].$$

Of course, these are precisely the kind of joint statistics that Bayesian's are often reminded are hard to obtain.

In the presence of $\alpha(s_1,s_2)$-independence, it is not hard to show that log-probabilities now update according to the formula

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

Trivially, then, the new updating formulas become

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

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 α-independence.

We now introduce a second new concept to the formulation, that of union-based combination. In the formulas above, the information in $s_1 s_2$ is based on the product space of experts $E_1 \times E_2$. We find it more desirable to base formulas on the union space of experts $E_1 \cup E_2$. The resulting set of information, $s_1 s_2$, contains updated opinions for each expert $e \in E_1 \cup E_2$. 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 E_1 update in a Bayesian fashion based on the mean opinion from the set E_2. Likewise, the experts in E_2 update using the information obtained from the mean opinion of experts in E_1. This changes the component formulas, so that, for $e \in E_1$,

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

Likewise, for $e \in E_2$, we have the same formula, but with all occurrances of s_1 and s_2 exchanged. In particular, we make use of the value $\alpha(s_2,s_1)$.

The result is easy to obtain, but now depends on the number of experts in the component sets, $|E_1|$ and $|E_2|$. For the log-probabilities, with α-independence, the formulas are:

$$\mu_{s_1 s_2}^{(l)} = \frac{|E_1|\mu_{s_1}^{(l)} + |E_2|\mu_{s_2}^{(l)} + \alpha(s_1,s_2)|E_1|\mu_{s_2}^{(l)} + \alpha(s_2,s_1)|E_2|\mu_{s_1}^{(l)}}{|E_1| + |E_2|}$$

$$C_{s_1 s_2}^{(l)}(\lambda_1,\lambda_2) = \rho C_{s_1}(\lambda_1,\lambda_2) + \theta C_{s_2}(\lambda_1,\lambda_2) +$$

$$\rho\left[\mu_{s_1}^{(l)}(\lambda_1) + \alpha(s_1,s_2)\mu_{s_2}^{(l)}(\lambda_1) - \mu_{s_1 s_2}^{(l)}(\lambda_1)\right]\left[\mu_{s_1}^{(l)}(\lambda_2) + \alpha(s_1,s_2)\mu_{s_2}^{(l)}(\lambda_2) - \mu_{s_1 s_2}^{(l)}(\lambda_2)\right]$$

$$+\,\theta\left[\mu_{s_2}^{(l)}(\lambda_1) + \alpha(s_2,s_1)\mu_{s_1}^{(l)}(\lambda_1) - \mu_{s_1 s_2}^{(l)}(\lambda_1)\right]\left[\mu_{s_2}^{(l)}(\lambda_2) + \alpha(s_2,s_1)\mu_{s_1}^{(l)}(\lambda_2) - \mu_{s_1 s_2}^{(l)}(\lambda_2)\right]$$

where we have suppressed the λ argument in the first equation, and $\rho = |E_1|/(|E_1|+|E_2|)$, and $\theta = 1-\rho$.

We see that a state now consists of the mean log-opinion $\mu_s^{(l)}(\lambda)$ for all labels $\lambda \in \Lambda$, the covariance of the opinons $C_s^{(l)}(\lambda_1,\lambda_2)$, and a weight of evidence, $|E|$, 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 E_1 and the experts in E_2.

Many other formulations are possible. For example, we can compute means and covariances of the probabilities instead of the log-probabilites. We omit the formulas here, for lack of space. We could also use only incremental evidence, so that experts in E_1 update using the means of the opinions of E_2, while the experts in 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. Conclusions

We see that the formulas for tracking statistics of sets of experts with probabilistic opinions are conceptually simple. The resulting formulations permit the representation of uncertain information, by recording the degree of concurrence of opinions, as well as recording a current average opinion. Our suggested formulation has been based on combining sets of experts, with each expert using the statistics of the information from the other set to update his opinion. The major contribution of these ideas is that the updating of individual experts can be performed by well-founded probabilistic reasoning, namely Bayes' rule. In this way, although the resulting formulas make good sense from a subjectivist viewpoint, the subjectivity has been given a rigorous foundation based objective probabilistic updating. The key is that the updating takes place over sets of opinions, and not on a single opinion.

The idea of α-independence has been introduced. The definition generalizes the notion of statistical independence, and may apply in situations where independence is too-strong an assumption. This does not mean, necessarily, that α-independence is practical or realistic. The best we can claim is that α-independence may be more realistic than complete (conditional) independence for some applications. The resulting modifications to the formulas are not too startling: the degree of independence is used as a weighting factor to determine the amount of updating that occurs as a result of the new information. However, the α factor can be based on a parameter that is well-defined by the model, and defended by the rigorous derivation of the updating formulas.

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. Professor 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,* In preparation. November, 1985.

[5] Hummel, Robert and M. Landy, ''Evidence as opinions of experts,'' *Proceedings of the ''Uncertainty in AI'' Workshop,* pp. 136-143 (August 8-10, 1986).

[6] Shafer, G., ''Constructive probability,'' *Synthese* **48**, pp. 1-60 (1981).

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