Statistical Approaches to Fusion with Uncertainty

Robert Hummel
Courant Institute
New York University
New York, New York 10012
hummel@cs.nyu.edu

Larry Manevitz
Department of Computer Science
Haifa University
Haifa 31999, Israel
manevitz@robust.nyu.edu

Abstract— **Reasoning with uncertainty is a field with many different approaches and viewpoints, with important applications to sensor design and autonomous system development. We attempt to unify some of the different approaches by introducing a common philosophical framework under which different calculi may be developed. Each calculus reflects different design choices compatible with the philosophical tenets. The tenets postulate that uncertainty in representations can be viewed as a degree of dispersion of opinions, and that the space of opinions operates as a separate sample space distinct from the underlying sample space on which probabilities are normally defined. Different calculi result when choices are made for the representation of the opinions, the method for combining opinions, the method for juxtaposing multiple sets of opinions, and the way of measuring the spread in the opinions.**

I. INTRODUCTION

Many sensor fusion systems make use of an evidential reasoning system, where evidence is combined with current measurements in order to maintain states of belief and confidence in a set of hypotheses. These systems are all motivated by the fact that the sensor is supposed to provide more than simple measurements, but also confidence levels in the measurements, and ultimately confidence levels in propositions and hypotheses developed from those measurements. The fundamental concepts always involve quantities related to the degree of validity of a proposition, such as a probability, and other quantities related to the degree of certainty in the assertion of the degree of belief. Various calculi are used for representing these concepts and performing the calculations, including Bayesian networks, fuzzy logic, Kalman filtering, and the Dempster/Shafer theory of evidence. Each calculus has certain theoretical underpinnings, although a universally accepted methodology is still lacking.

There are really two problems that must be addressed when designing an evidential reasoning theory. First, the philosophical issues of belief, certainty, and confidence must be modeled in a rational manner. Second, the methodology for maintaining and combining states must be determined in a manner that conforms as nearly as possible to the philosophy. One reason for the profusion of different calculi is that both issues present serious difficulties. The philosophical issues present difficulties because different meanings can be ascribed to beliefs and certainties. Although probabilities are likely to be used to develop the calculus, the probabilities must apply to events that are well-defined, and the events will typically involve subjective evaluations that make the theory subject to varying interpretations. The methodological issues are difficult because no matter what scheme is chosen to implement the philosophical foundations, certain approximations will be necessary. Always, the methodology will fall short of the desired goals.

To make the philosophical questions more concrete, consider the difficulty of defining the statement that "There is a 20% probability that this there is a tank in this image." The frequency interpretation of such a sentence means that among 100 images that are more or less precisely the same, roughly 20 will actually contain the image of a tank. The difficulty with this interpretation is that it presupposes the existence of a sufficient number of cases with identical conditions — whereas the statement may be uttered by a knowledgeable photo-interpreter who has never seen precisely such an image before; indeed, there may have never been such an image before! Another interpretation might be termed the "subjective" probability theory, and is founded on work by DeFinetti, Good, Savage, Kyburg, Fisher, and others [1,2,3]. In the "subjective interpretation," the statement can be interpreted to mean something along the lines of "I would accept 1 to 4 odds that there is a tank in the image." (We hope that the person is betting with money, and not with their life!) However, such an interpretation can lead to different measurement methods. For example, if one insists that the bettor should come out even in the average over many bets on the same situation, then the subjective interpretation should yield the same value as computed in the frequency interpretation. On the other hand, one could take a particular situation and devise experiments to find a psychometric function under varying conditions, to find the odds under which a majority of experts would be indifferent when placing "bets." This value need not equal a precise frequency even if the frequency can reasonably be measured statistically.

Manuscript submitted March 1, 1991. Robert Hummel is currently at INRIA, Project Epidaure, BP 105, 78153 Le Chesnay Cedex, France. Robert Hummel was in part supported by Air Force contract F33615-89-1087.

Difficulties also arise in using probabilities, no matter how they are developed, for inferencing. Bayesian analysis and non-Bayesian approaches both have advocates and their specific rationale, but invariably invoke philosophical debates concerning their applicability. We begin with a specific set of philosophical tenets, founded on the notion that the sensor systems yield measurements with variances that can be determined on physical or empirical grounds.

II. TENETS

The tenets are:

(1) That uncertainty can be represented by a distribution of opinions, whereas certainty is represented by unanimity of multiple opinions.

(2) By an "opinion," we mean an estimate of a quantity that is functionally related to a (frequency-based) probability, or is a subjective estimate of a conditional probability, or is an estimate of a well-defined quantity representing the likelihood of a given proposition based upon given evidence.

(3) When opinions are combined in order to make estimates that are conditioned on combinations of evidence, a precise and well-defined combination formula should be used. In the case that the opinions are estimates of probabilities, Bayes' rule should be used to combine pairs of opinions to yield a new opinion.

The propositions can relate to measurement values obtained from different sensors, or can be propositions that are developed and based upon those measurements (such as the presence or absence of a target class).

We present a variety of different calculi that are obtainable from these philosophical foundations, depending upon the values that the opinions are supposed to represent, and depending upon the assumptions used in the updating process. Our intent is to show the utility of the multiple-opinions approach to uncertainty. We do not purport to obviate other philosophical approaches to uncertainty, and we do not support a particular uncertainty calculi over all others. In particular, we do not give a thorough survey of existing calculi nor their relation to the calculi that are derived as a result of the multiple-opinions framework. However, we do show how these philosophical tenets are reflected in the Dempster/Shafer calculus, and to the calculi of the systems approach to combining uncertain estimates, generally known as Kalman filtering. More importantly, the approach leads to a way of categorizing uncertainty calculi, and a way for choosing a calculus that is most appropriate for a specific application.

Certain aspects of each calculus is invariant. They all will represent a state of belief by statistical measures on the distribution of opinions, and they will also use a method to combine opinions based on a rational Bayesian formulation. Combinations of opinions yield a new set of opinions; a meaningful measure of certainty in any single measurement or proposition necessarily entails multiple opinions.

III. DESIGN CHOICES

Each calculus contains four main design choices:

Representation of opinions:

The opinions may represent individual measurements of some sensor (single value or multidimensional value), or might be a probability of some proposition. The representational issue includes the measurement units: for example, for the case of probabilistic measurements of propositions, logarithmic probabilities often are more convenient units.

Statistics maintained:

The most common statistics are simply the mean and covariance of the opinions. However, in the Dempster/Shafer calculus, all statistics of the opinions are maintained. Intermediate systems might provide more specificity concerning the degree of certainty in the opinions.

Combining of sets:

We assume that the combination of belief states, as obtained from individual sensors or different groups of sensors, takes place by combining the the statistics of the multiple opinion sets. One way to combine the sets of opinions is to take the union of all constituent opinions. However, unions do not permit a multiplier effect that can arise due to independent confirmation; in order to permit a more precise analysis, it is necessary for pairs of opinions to be combined through a Bayesian or analytical process. The issue of the combination of the sets concerns the pattern in which pairs of opinions are brought together. The main choices, other than union, involve pairing off opinions, under the assumption that each set of opinions always has n elements, or taking a set product, to form the set of all pairs of opinions with one component from each set.

Combining of opinions:

Once pairs of opinions are brought together, a new opinion must be generated. This can be done by averaging the opinions, finding an intersection, Bayes' formula assuming independence, a Bayesian combination assuming some other parameterized form of independence, or a weighted average.

Each design choice leads to a different calculus, giving a Chinese menu of uncertainty calculi. Different applications will require different design choices. We illustrate a number of different calculi that arise for several specific applications. In this short paper, we mention three such calculi, but omit descriptions of the corresponding applications.

IV. DEMPSTER/SHAFER

The Dempster/Shafer theory of calculus [4], with it's "Dempster rule of combination," represents the following design choices. The opinions necessarily relate to a proposition for which there are multiple possible labels (the frame of discernment). Each opinion is simply a list (i.e., a subset) of the labels that are considered possible. This is often misunderstood, since the values in the Dempster/Shafer calculus resemble probabilities. In reality, the opinions represent lists of possibilities, and preclude the expression of the degree of probability in the individual labels. The statistics are the full set of joint statistics, represented as the percentage of opinions in the set of opinions designating a given subset as the exact list of possible labels. The collection of all of these percentages form the "mass function," which gives the state of belief over the frame of discernment. Sets of opinions are combined by taking the product space of opinions. Each pair of opinions forms a new opinion by intersecting their lists, except that if the intersection is empty, then the new opinion is discarded, and the pair is removed from the product set of experts.

Much has been written in attempting to provide alternate interpretations of the combination formulas, however, whatever interpretation one adopts, the design choices indicated above and the interpretation that it provides must give an isomorphic understanding of the workings of the system. The particular interpretation given here happens to be the one defined by Dempster in his original introduction of the calculus [5,6].

We now specify the interpretation in mathematical notation. More details may be found in an earlier paper [7]. The same notation will be used for the alternative calculi.

Let Λ be the frame of discernment, which is a finite collection of mutually exclusive and exhaustive propositions. That is, for a given situation, exactly one label from the set Λ is true. The set of opinions will be indexed over E, the set of experts. An individual expert in E is denoted by ω. Thus an expert $\omega \in E$ has an opinion on the current situation. The expert might give a probability distribution over Λ, but instead gives a set of possible labels. That is, for each label in $\lambda \in \Lambda$, the expert ω gives a **boolean opinion**, (either 0 or 1), saying whether that label is possible or not in the given circumstances. We denote this (boolean) opinion by $x_\omega(\lambda)$. That is, $x_\omega(\lambda)$ is equal to 1 if and only if expert ω in E believes that label λ in Λ is possible. For labels that the expert rules out, the corresponding value of $x_\omega(\lambda)$ will be zero. We will assume that every expert names at least one label as being possible.

Given a collection of values $\{x_\omega(\lambda)\}$ indexed over ω in E and λ in Λ, the state of the system (i.e., the belief state) is defined by the statistics over E of the opinions. The complete set of joint statistics may be represented by:

$$C(A) = \underset{\omega \in E}{\mathrm{Prob}}(x_\omega(\lambda) = 1 \text{ for all } \lambda \in A),$$

where A ranges over all possible subsets of Λ. These are the full set of joint probabilities. They also happen to equal the **commonality numbers** as defined by Shafer [4], and they are equivalent to the full set of belief values. That is, from the commonality values one can derive the full set of belief values, and vice-versa. Accordingly, the belief state is represented by the full collection of statistics over the set of experts of the boolean opinions given by those experts. The set of masses are also equivalent to these collections; the mass on a subset A can be defined as the probability that an expert names *precisely* A as the subset of possibilities (i.e., that the labels in A are possible and the labels outside of A are ruled out).

Finally, suppose we wish to combine two belief states. One belief state is represented by the statistics of a collection of opinions $\{x_{\omega_1}(\lambda)\}$ for ω_1 in E_1 and λ in Λ, and the other belief state is represented by the statistics of a collection of opinions $\{x_{\omega_2}(\lambda)\}$ indexed over ω_2 in E_2, and λ in Λ. Given two opinions of x_{ω_1} and x_{ω_2}, we define the combined opinion as the pointwise product:

$$x_{(\omega_1,\omega_2)}(\lambda) = x_{\omega_1}(\lambda) \cdot x_{\omega_2}(\lambda).$$

That is, according to the committee consisting of experts ω_1 and ω_2, a label is still possible only if both committee members agree that the label is possible. We then take the set of all combined opinions, where one opinion comes from the collection E_1 and the other opinion comes from the collection E_2. We rule out any combined opinions where the committee rejects all labels. The resulting set of opinions is indexed over a set E, which is a subset of the product set of $E_1 \times E_2$, and can be used to collect statistics over E. The full collection of statistics is used to determine the commonality numbers, which in turn are equivalent to beliefs, which in turn are equivalent to masses. If we suppose that the resulting masses are given by $m(A)$ for A ranging over the subsets of Λ, and that the original masses of the two collections are given by the functions $m_1(A)$ and $m_2(A)$, then we discover that:

$$m(A) = \frac{1}{Z} \cdot \sum_{B \cap C = A} m_1(B) \cdot m_2(C)$$

where

$$Z = 1 - \sum_{B \cap C = \varnothing} m_1(B) \cdot m_2(C)$$

This is precisely the Dempster rule of combination.

We thus see that the Dempster rule of combination is completely compatible with the philosophical tenets defined in Section II, under the design choices of boolean opinions, complete tracking of all statistics, and product rule combination (throwing out opinions when there are no possibilities).

V. LOG PROBABILITIES

Next, suppose that the opinions are represented by the logarithms of probabilities for a particular proposition. We again envision a collection of "experts" E, with each expert ω in E expressing an opinion $x_\omega(\lambda)$ for every label $\lambda \in \Lambda$. The values, however, are not simply the logarithms of estimated conditional probabilities based on the known observations, but rather the log of the ratio of the conditional probability and the prior probability. Specifically, we set

$$x_\omega(\lambda) \approx \frac{\text{Prob}(\lambda \mid \text{Information})}{\text{Prob}(\lambda)},$$

where the "Information" is the information shared by the experts in E, and the denominator is the prior probability of label λ in the absence of any information. Note that these probabilities are defined over the usual sample space of problem instances, and not the set of experts (as was the case with the probabilities used in the previous section). These values are the representation suggested by Charniak [8] for probabilistic reasoning. Each expert will have a different estimate of this log-ratio, and the statistics that we propose to maintain are the mean and variance of these values. Thus if an expert regards a proposition as being four times as likely due to the given measurements as opposed to its probability in the absence of information, then the expert contributes log(4) as the opinion in the set of opinions, from which we measure the mean and variance. Note that if the information has no influence on the prior probability, according to an expert, then the expert's opinion will be log(1), i.e., zero.

The state of the system is then represented by two vectors:

$$\mu(\lambda) = \underset{\omega \in E}{\text{Avg}}(x_\omega(\lambda)),$$

and

$$\sigma(\lambda) = \left[\underset{\omega \in E}{\text{Avg}}(x_\omega(\lambda) - \mu(\lambda))^2\right]^{1/2}.$$

Unlike the Dempster/Shafer representation, which requires 2^N values for the specification of the state, this formulation requires only $2N$ values, where N is the number of possible labels (i.e., $N = \#\Lambda$).

Now, if two such sets of opinions are to be combined, we chose to take the set product of the opinions. If a composite opinion is to be formed from two individual opinions, and if we may suppose a conditional independence between the information sources on which the two experts are basing their opinions, then it can be shown that modulo a uniform additive constant, the two opinions may be summed. This comes from Bayes formula, using conditional independence of the information sources, yielding:

$$\frac{\text{Prob}(\lambda \mid \text{Info}_1, \text{Info}_2)}{\text{Prob}(\lambda)} \;\alpha\; \frac{\text{Prob}(\lambda \mid \text{Info}_1)}{\text{Prob}(\lambda)} \cdot \frac{\text{Prob}(\lambda \mid \text{Info}_2)}{\text{Prob}(\lambda)}$$

The two sides are equal, except for a proportionality constant, which is independent of λ. Taking the log of both sides, we see that it is logical to set

$$x_{(\omega_1,\omega_2)}(\lambda) = x_{\omega_1}(\lambda) + x_{\omega_2}(\lambda).$$

The proportionality constant has been dropped, which means that the opinions can be off by a constant additive amount (but the same constant for all the labels λ). This skewing, it turns out, is unimportant, since we are only concerned in the relative size of the components over different labels λ. This then is the updating method, and we can see that independence (conditioned on every label) is required for the information. A mathematical statement of the independence assumptions says that

$$\text{Prob}(\text{Info}_1 \mid \text{Info}_2, \lambda) = \text{Prob}(\text{Info}_1 \mid \lambda)$$

for all λ in Λ. These are strong requirements, but potentially valid in some circumstances. See Charniak [8] for a further discussion of the applicability, and see a previous paper [9] of ours for an alternative way of weakening the conditional independence requirements.

Using the product formulation for obtaining the the combination set of opinions, we find that the following formulas hold. The mean and standard deviation of the opinions of the experts in set E_1 are denoted by (μ_1, σ_1), and the opinions of the experts in set E_2 give rise to the state (μ_2, σ_2). Then

$$\mu(\lambda) = \mu_1(\lambda) + \mu_2(\lambda),$$
$$\sigma(\lambda) = \left[\sigma_1^2(\lambda) + \sigma_2^2(\lambda)\right]^{1/2}.$$

That is, the resulting calculus can be specified by the statement that mean log-probability opinions should be added, and variances also add. We look for a situation where the resulting mean opinion is either very large positive, or very large negative, with a relatively tight variance (much less than the magnitude of the opinion) in order to conclude that the corresponding proposition is true or false.

KALMAN FILTERING

Finally, suppose that the opinions represent vectors of measurements, obtained from sensors. Based on a set of measurements, for instance, we may have a set of opinions that cluster at a particular vector $\mathbf{v}_1$, with covariance C_1. Another set of opinions, perhaps based on a different technology, yield a mean and covariance of $\mathbf{v}_2$ and C_2. We assume there are an equal number of samples in each set leading to the measurement of the mean and covariance. We regard the mean as a single random variable whose realization is, respectively, $\mathbf{v}_1$ and $\mathbf{v}_2$. The expected error in each realization is related to the observed covariance matrices, C_1 and C_2. The combination method pairs the collection of opinions as a single entity with the second collection of opinions as a single entity, and the combination of those opinions is based on an analysis of the density function for the true value being measured, conditioned on the fact that two realizations of the two separate Gaussian random vectors gave particular values. The resulting density function is

Gaussian, under the assumption that the prior distribution for the measured value is Gaussian. The mean is given by a weighted sum of $\mathbf{v}_1$ and $\mathbf{v}_2$, where the weights are inversely proportional to the covariance matrices. Specifically, the new mean is given by

$$CC_1^{-1}\mathbf{v}_1 + CC_2^{-1}\mathbf{v}_2$$

where $C^{-1} = C_1^{-1} + C_2^{-1}$. The covariance of the distribution is given by C. These are precisely the formulas that arise when using the Kalman filtering approach to combining independent measurements, expressed as a combination formula rather than a sequential filter.

Let us suppose that the covariance matrices are always diagonal. This occurs if and only if the values of the $x_\omega(\lambda)$'s are independent with respect to λ. In this case, we may view the Kalman filtering formulas (which are really for the simplest case of filtering, where the state transition formula is static) in even closer correspondence to our philosophical tenets.

We now assume that the experts are numbered 1 through n. Each expert has an opinion $x_\omega(\lambda)$ for each label λ. We use the pairwise combination method for mixing experts, so that two experts meet only if they have the same number. Thus we have two sets of experts, E_1 and E_2, where each expert in both sets has a number between 1 and n. Two experts $\omega_1 \in$ *E sub 1$ and $ww sub 2* $\in E_2$ meet only if they have the same number, $Num(\omega_1) = Num(\omega_2)$. We thus have n pairs in the pairwise combining.

The updating formula works as follows. Suppose the pair $x_{\omega_1}(\lambda)$ and $x_{\omega_2}(\lambda)$ are to be combined. Our experts decide that an average should be taken in order to get the final value. However, they decide to weight the average unequally, using as weighting coefficients the *inverses* of the variances of the corresponding labels within the respective entire set of experts. That is, the average is:

$$x_{(\omega_1,\omega_2)}(\lambda) = a \cdot x_{\omega_1}(\lambda) + b \cdot x_{\omega_2}(\lambda),$$

where $a+b=1$, and a is proportional to $(\sigma_1(\lambda))^{-2}$ and b is proportional to $(\sigma_2(\lambda))^{-2}$. Here,

$$\sigma_i(\lambda) = \left[\operatorname*{Avg}_{\omega \in E_i}(x_\omega(\lambda) - \mu(\lambda))^2\right]^{1/2}$$

is the standard deviation of the λ'th component within the set of opinions indexed over E_i. This is reasonable, since the degree of certainty that an expert $\omega_i \in E_i$ has in his own opinion is inversely related to the variance of that opinion among his colleagues. Accordingly, a combined opinion is obtained for every pair of experts that meet.

We maintain the mean opinion for each label, and the standard deviation of each opinion, as in the previous section. That is, the states are represented by $(\mu_i(\lambda), \sigma_i(\lambda))$ for $i = 1,2$. It is easy to see that the resulting mean opinion of the collection of experts (n of them) obtained by pairing off the experts in E_1 and E_2 is just

$$\mu(\lambda) = a_\lambda \cdot \mu_1(\lambda) + b_\lambda \cdot \mu_2(\lambda),$$

where a_λ and b_λ are respectively proportional to the inverse of the variances of the labels λ. That is, aa_λ is proportional to $(\sigma_1(\lambda))^{-2}$ and bb_λ is proportional to $(\sigma_2(\lambda))^{-2}$. We thus obtain the formulas:

$$a_\lambda = \frac{\sigma_1^{-2}(\lambda)}{\sigma_1^{-2}(\lambda) + \sigma_2^{-2}(\lambda)},$$

$$b_\lambda = \frac{\sigma_2^{-2}(\lambda)}{\sigma_1^{-2}(\lambda) + \sigma_2^{-2}(\lambda)},$$

which are the same as the Kalman updating formulas. Computing the standard deviation of the n resulting opinions yields

$$\sigma^{-2}(\lambda) = \sigma_1^{-2}(\lambda) + \sigma_2^{-2}(\lambda)$$

as desired.

Thus for the case of independent labels, the mean and standard deviations of the opinions, with pairwise combining, correspond to the static Kalman filtering equations, in the case where opinions are combined by weighted averaging, with the weights proportional to the inverse of the variance of the corresponding label among the opinions within the set of experts.

REFERENCES

[1] I.J. Good, "The measure of a non-measurable set," in *Logic, Methodology, and Philosophy of Science,* E. Nagel, P. Suppes, and A. Tarski (Eds.). Stanford University Press, 1962, pp. 319-329.

[2] H. E. Kyburg, Jr., "Bayesian and non-Bayesian evidential updating," University of Rochester Dept. of Computer Science Technical Report 139, July, 1985.

[3] L. J. Savage, *The Foundations of Statistics,* New York: John Wiley & Sons, 1954.

[4] G. Shafer, *A mathematical theory of evidence,* Princeton, N.J.: Princeton University Press, 1976.

[5] A. P. Dempster, "Upper and lower probability inferences based on a sample from a finite univariate population," *Biometrika,* Vol. **54**, 1967, pp. 515-528.

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

[7] R. A. Hummel, M. S. Landy, "A statistical viewpoint on the theory of evidence," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 1988, pp. 235-247.

[8] E. Charniak, "The Bayesian basis of common sense medical diagnosis," *Proceedings of the AAAI,* Washington, D.C., 1983, pp. 70-73.

[9] R. Hummel, L. Manevitz, "Combining bodies of dependent information," *Proceedings of the Tenth International Joint Conference on Artificial Intelligence,* Milan, August, 1987.