USING EXPONENTIAL MIXTURE MODELS FOR SUBOPTIMAL DISTRIBUTED DATA FUSION

Simon J. Julier

Department of Computer Science
University College London
Malet Place
London WC1E 6BT, UK
S.Julier@cs.ucl.ac.uk

Tim Bailey

Australian Centre for Field Robotics
The Rose Street Building J04
The University of Sydney
Sydney, 2006 NSW, Australia
tbailey@acfr.usyd.edu.au

Jeffrey K. Uhlmann

Department of Computer Science
327 Engineering Building West
University of Missouri — Columbia
Columbia, MO 65211-2060, USA
uhlmannj@missouri.edu

ABSTRACT

In this paper we investigate the use of Exponential Mixture Densities (EMDs) as suboptimal update rules for distributed data fusion. We show that EMDs have a pointwise bound "from below" on the minimum value of the probability distribution. However, the distributions are not bounded from above and thus can be interpreted as a fusion operation.

1. INTRODUCTION

One of the key enabling technologies for sensor networks is distributed data fusion. A sensor network consists of a set of fusion nodes. A node received information collected locally (from sensors) and remotely (from other fusion nodes). This information is fused and the estimates are distributed to other nodes. To maintain scalability and robustness, nodes should only maintain local knowledge of the network: the only information they have about the network topology is their list of neighbours. However, because estimates rather than observations are distributed between nodes, the information is *not* conditionally independent. There are several reasons for this including commonality of process noise (in the target motion) and the fact that nodes repeatedly fuse the estimates they receive from their neighbours.

In such circumstances, fusion can take place using a modified form of Bayes Rule [1]. Let A and B be two nodes in the network. At time step k node A constructs its estimate $P\left(\mathbf{x}_k|\mathbf{Z}_a^k\right)$, based on the history of all information it has received, $\mathbf{Z}_a^k$. B has constructed its own state estimate $P\left(\mathbf{x}_k|\mathbf{Z}_a^k\right)$ using its own measurement history $\mathbf{Z}_b^k$. In general the sets are not disjoint; they contain common information and are thus not independent of one another. The update rule [1]

$$P\left(\mathbf{x}_k|\mathbf{Z}_a^k\bigcup\mathbf{Z}_b^k\right)\propto\frac{P\left(\mathbf{x}_k|\mathbf{Z}_a^k\right)P\left(\mathbf{x}_k|\mathbf{Z}_b^k\right)}{P\left(\mathbf{x}_k|\mathbf{Z}_a^k\bigcap\mathbf{Z}_b^k\right)},\qquad(1)$$

divides by the common information and thus "cancels it out".

In principle $P\left(\mathbf{x}_k|\mathbf{Z}_a^k\bigcap\mathbf{Z}_b^k\right)$ can be calculated by maintaining the probability density between the states of all nodes in the entire network. Local algorithms have been derived in the special cases that the network is fully-connected [2] or tree-connected [3, 4]. However, if the network topology contains cycles, no local algorithm exists [5] and *approximations* of (2) must be used.

In this paper we consider the problem of developing a fusion algorithm which yields an estimate which is guaranteed to satisfy some notion of consistency. This problem was first considered by Uhlmann in the context of SLAM [6] who developed an algorithm known as Covariance Intesection (CI). CI was derived by considering the first two moments of the estimates to be fused and it generates an estimate such that the mean squared error in the estimate is never underestimated. However, CI can only utilise the first two moments of the distribution; it cannot exploit any other information such as a probabilistic description of the prior states. Seeking to generalise CI, both Mahler [7] and Hurley [8] noted that, if one assumes the estimates are Gaussian distributed, the CI update is equivalent to constructing the Exponential Mixture of the two densities. Empirical tests suggest that the algorithm yields usable estimates in distributed fusion networks with cycles [9, 10]. However, little theoretical analysis has been carried out to determine if EMDs truly offer a means of "robustly" fusing estimates which contain unknown dependencies and, if so, what notion of "robustness" is preserved.

This paper performs some preliminary analysis of the theoretical properties of EMDs and discusses how they might be related to robust fusion problems. Section 2 introduces the EMD and discusses some of its use in the literature and some of the properties which have been identified. Most of these results are in terms of Kullback-Liebler divergence. The interpretation of these results to inference problems is not always clear. Therefore, in Section 3 we derive pointwise bounds and discuss the relevance of these for fusion problems. The results are discussed in Section 4.

2. EXPONENTIAL MIXTURE DENSITIES

The exponential mixture density takes a logarithmic convex combination of the input estimates. Rewriting (1) using the more compact notation

$$P_c(\mathbf{x}) \propto \frac{P_a(\mathbf{x})\, P_b(\mathbf{x})}{P_{a\bigcap b}(\mathbf{x})}, \tag{2}$$

the Exponential Mixture Density (EMD) is

$$P_c(\mathbf{x}) = \frac{1}{N_c} P_a(\mathbf{x})^{\omega} P_b(\mathbf{x})^{(1-\omega)}. \tag{3}$$

The free parameter $\omega \in [0,1]$ is chosen to minimise some measure of uncertainty in $P_c(\mathbf{x})$. N_c is the normalising constant and its value is given by

$$N_c = \int P_a(\mathbf{x})^{\omega} P_b(\mathbf{x})^{1-\omega} d\mathbf{x}.$$

The EMD arises in two distinct contexts: logarithmic opinion pools and bounds on the Bayes Probability of Error in signal detecton and hypothesis testing.

2.1. Logarithmic Opinion Pools

Logarithmic opinion pool are a method used to perform expert fusion [11]. Expert fusion occurs when the output of a set of inference algorithms, each optimised to detect different classes or events, are to be fused together. Each expert utilises the same observations and the estimates are thus not independent. Let $P_\alpha(\mathbf{x})$ be the probability distribution generated by the αth expert and let $D(\cdot\|\cdot)$ be the Kullback-Leibler divergence. The opinion pool constructs the estimate $P_c(\mathbf{x})$ such that

$$P_c(\mathbf{x}) = \arg\min_{P_c(\mathbf{x})} \sum w_\alpha D\left(P_c(\mathbf{x}) \| P_\alpha(\mathbf{x})\right)$$

where the weights w_α are chosen to reflect the confidence in expert α. Under the constraint that $P_c(\mathbf{x})$ is normalised, it can be shown that the result is the EMD. Heskes describes a method for calculating the weights which attempts to minimise the Kullback-Leibler divergence between the true probability distribution and $P_c(\mathbf{x})$ [11]. He also proves that the Kullback-Leibler divergence of the logarithmic opinion pool cannot exceed the average Kullback-Leibler divergence of the individual experts, showing that fusion improves the performance of the recognition algorithm.

2.2. Upper Bound on Bayes Probability of Error

The second main use for EMDs are in hypothesis testing and signal detection [12–14]. Given a two class problem, with hypotheses H_a and H_b such that

$$H_a : \mathbf{x} \approx P_a(\mathbf{x})$$
$$H_a : \mathbf{x} \approx P_b(\mathbf{x})$$

each with *a prior* probability $\pi_a = \pi_b = \frac{1}{2}$, it can be shown that the lower bound on the Bayes Probability of Error, $P(e)$, is $P(e) \leq N_c$ for all $\omega \in [0,1]$ [12]. Cover and Thomas proved that, if $P_a(\mathbf{x})$ and $P_b(\mathbf{x})$ are discrete and the decision regions are convex, the upper error could be achieved if the samples were distributed according to (3) [13] and the value of $\omega = \omega^\star$ was chosen such that

$$D\left(P_c(\mathbf{x}) \| P_a(\mathbf{x})\right) = D\left(P_c(\mathbf{x}) \| P_b(\mathbf{x})\right)$$

In other words, the EMD is equi-distant from both $P_a(\mathbf{x})$ and $P_b(\mathbf{x})$. This work was extended by Dabak, who considered the problem of building robust detectors. He proved that (3) is an exponential curve in the set of all probability measures absolutely continuous with respect to $P_a(\mathbf{x})$ and $P_b(\mathbf{x})$ [14]. It encodes the transformations over which the performance of the optimal detector remains unchanged. Furthermore, these results assume nothing about the independence of observations in the detection tests. Therefore, it relaxes the assumptions that the distribution is discrete and the decision region is convex. Furthermore, it does not rely on the assumptions that the observations are independently and identically distributed or that the distributions have to be parametric.

The above review suggests that EMDs have a number of properties which make them advantageous as a basis for robust fusion algorithms. First, they have been used in fusion problems (expert fusion) where multiple dependent probabilities must be fused together. Second, the have a strong connection with information theory and have the ability to encode "worst case" distributions for fusion and inference problems. Finally, the computational form is particularly attractive — because it resembles Bayes Rule, many nonlinear filtering methods (such as particle filters) could be adapted to use this form.

However, although the previous analysis has shown that the performance measures can be defined in terms of the Kullback-Leibler divergence we seek stronger conditions to identify what is robust in these algorithms. Our preliminary analysis has identified pointwise bounds induced by the EMD.

3. POINTWISE BOUNDS INDUCED BY EMDS

In this section we identify pointwise lower and upper bounds on EMDs and argue that these can be used to define notions of consistency and fusion.

3.1. Lower Bounds and a Notion of Consistency

Let $P'_c(\mathbf{x})$ be the unnormalised probability density function,

$$P'_c(\mathbf{x}) = P_a(\mathbf{x})^{\omega} P_b(\mathbf{x})^{1-\omega}. \tag{4}$$

Theory 1. *$P'_c(\mathbf{x})$ always lies* between *$P_a(\mathbf{x})$ and $P_b(\mathbf{x})$,*

$$\min\{P_a(\mathbf{x}), P_b(\mathbf{x})\} \leq P'_c(\mathbf{x}) \leq \max\{P_a(\mathbf{x}), P_b(\mathbf{x})\} \forall \mathbf{x}. \tag{5}$$

Proof. Suppose $P_a(\mathbf{x}) \leq P_b(\mathbf{x})$ and consider the lower bound:

$$P_a(\mathbf{x}) \leq P_a(\mathbf{x})^{\omega} P_b(\mathbf{x})^{1-\omega}.$$

Dividing both sides by $P_a(\mathbf{x})$,

$$1 \leq \left(\frac{P_b(\mathbf{x})}{P_a(\mathbf{x})}\right)^{1-\omega}.$$

Since $P_b(\mathbf{x}) > P_a(\mathbf{x})$ and $\omega \leq 1$, the inequality holds true. Similarly, the upper bound is

$$P_a(\mathbf{x})^{\omega} P_b(\mathbf{x})^{1-\omega} \leq P_b(\mathbf{x}).$$

Dividing through by $P_b(\mathbf{x})$ leads to

$$\left(\frac{P_a(\mathbf{x})}{P_b(\mathbf{x})}\right)^{\omega} \leq 1.$$

Since $P_b(\mathbf{x}) > P_a(\mathbf{x})$ and $\omega \geq 0$, this inequality holds true. □

The above inequality on the unnormalised distribution $P'_c(\mathbf{x})$ can be used to derive an inequality on the normalised distribution $P_c(\mathbf{x})$.

Theory 2. $P_c(\mathbf{x})$ *obeys the inequality [14, 15] that*

$$P_c(\mathbf{x}) \geq \min\{P_a(\mathbf{x}), P_b(\mathbf{x})\}\,\forall \mathbf{x}. \tag{6}$$

Proof. From the lower inequality in (5), the only way in which (6) can be violated is if $N_c > 1$. Now, $P'_c(\mathbf{x})$ is convex in ω and so

$$\begin{aligned} N_c &\leq \int \omega P_a(\mathbf{x}) + (1-\omega) P_b(\mathbf{x}) d\mathbf{x} \\ &\leq \omega \int P_a(\mathbf{x}) d\mathbf{x} + (1-\omega) \int P_b(\mathbf{x}) d\mathbf{x} = 1. \end{aligned}$$

□

This property naturally leads to a definition of consistency: an update rule is consistent if the probability of finding that the state is at $\mathbf{x}$ is not *reduced* as a result of the update. The justification for this interpretation arises from the fact that the difficulty with fusing estimates which are dependent upon one another is that it is possible to *overestimate* the amount of information available and thus *underestimate* the uncertainty in the update. This underestimation is characterised by increasing the mass in certain regions more than it should and, corresondingly, reducing the mass in other regions more than it should. Because the mass at each point cannot be reduced below the smallest value in either $P_a(\mathbf{x})$ or $P_b(\mathbf{x})$, the estimate cannot be underestimated.

The lower bound property naturally leads to a definition of consistency. However, it does not necessarily lead to a usable fusion algorithm. For example, taking the weighted average of the prior distributions,

$$P_c(\mathbf{x}) = \omega P_a(\mathbf{x}) + (1-\omega) P_b(\mathbf{x}), \tag{7}$$

also yields a consistent estimate. However, the performance of using estimators using this linear opinion pool are known to be extremely poor [11] and this leads to the question of defining a notion of fusion.

3.2. Upper Bound and a Notion of Consistency

As explained above, a fusion operation reduces the mass in some regions of the state space and increases it in others. Theory 2 places limits on the lower bound of the distribution. However, it places no limits on the behaviour of the upper bound. This, in turn, can be used to define a notion of fusion. Specifically,

Theory 3. *It is possible for an* $\mathbf{x}$ *to exist such that*

$$P_c(\mathbf{x}) \geq \max\{P_a(\mathbf{x}), P_b(\mathbf{x})\} \tag{8}$$

Proof. Suppose the above inequality holds at some point $\mathbf{x}^\star$. Then

$$P'_c(\mathbf{x}^\star) \geq N_c \max\{P_a(\mathbf{x}^\star), P_b(\mathbf{x}^\star)\}.$$

Let $P_a(\mathbf{x}^\star)$ be the larger of the two values on the RHS. Therefore, the inequality holds true if $N_c \leq \frac{P'_c(\mathbf{x}^\star)}{P_a(\mathbf{x}^\star)}$. From the convexity of $P'_c(\mathbf{x}^\star)$,

$$\frac{P'_c(\mathbf{x}^\star)}{P_a(\mathbf{x}^\star)} \leq \frac{\omega P_a(\mathbf{x}^\star) + (1-\omega) P_b(\mathbf{x}^\star)}{P_a(\mathbf{x}^\star)}$$

Since $P_a(\mathbf{x}^\star) \geq P_b(\mathbf{x}^\star)$ is the larger of the two terms, the value of the right hand side cannot exceed 1. Since (6) proves that $N_c \leq 1$, (8) can be true. □

One case where this condition is guaranteed to hold true is if the distributions intersect. Let $(^{x})\star$ be an intersection point. Therefore, $P_a(\mathbf{x}^\star) = P_b(\mathbf{x}^\star) = P'_c(\mathbf{x}^\star)$ and

$$P_c(\mathbf{x}) = \frac{1}{N_c} P_a(\mathbf{x}).$$

In consequence, the mass can increase in some places above what was originally there leading to a potential gain in information.

4. DISCUSSION

This paper has considered the problem of fusing estimates in a distributed data fusion network. These problems are characterised by the fact that the marginal, but not the joint, distributions are known. We have examined the properties of Exponential Mixture Densities in the existing literature. We hope that these properties may relate to the problem of conservative fusion and promote future discussion on a rigorous definition of "conservativeness". Our current interpretation of EMD fusion is as a bound on probabilities; it yields bounds on the update "from below" and shows that the estimate can exceed the prior estimates "from above". These properties suggest that the algorithm maintains a level of consistency and provides information gain at the same time. For Gaussian data fusion, EMDs are equivalent to the covariance intersection algorithm, and their conservative fusion properties extend elegantly to non-Gaussian problems.

In conclusion, we believe that EMDs have the potential to provide consistent and conservative data fusion of information with unknown correlations. We propose that (6) is a rational working definition of "conservativeness" in this case. However, a rigorous definition and proof of conservative fusion remains an open problem. Furthermore, the results derived in Section 3 only exploit the *convex* nature of the EMDs and not the specific form of the EMDs themselves. We believe that the Kullback-Leibler divegence-minimising properties and likelihood invariants play an important role, but the precise nature of this relationship is currently unclear.

Acknowledgements

We would like to thank Dave Nicholson at BAE Systems Filton, Neil Gordon at DSTO, Paul Thompson and Ben Upcroft at the Australian Centre for Field Robotics, and Don Johnson at Rice University, for their help in writing this paper.

5. REFERENCES

[1] Martin E. Liggins II, Chee-Yee Chong, Ivan Kadar, Mark G. Alford, Vincent Vannicola, and Stelios Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 95–107, 1997.

[2] B. S. Y. Rao, H. F. Durrant-Whyte, and A. Sheen, "A fully decentralized multi-sensor system for tracking and surveillance," *International Journal of Robotics Research*, vol. 12, no. 1, pp. 20–45, 1991.

[3] S. Grime and H. F. Durrant-Whyte, "Data fusion in decentralized sensor fusion networks," *Control Engineering Practice*, vol. 2, no. 5, pp. 849–863, 1994.

[4] L. Ong, B. Upcroft, M. Ridley, T. Bailey, S. Sukkarieh and H. Durrant-Whyte, "Decentralised data fusion with particles," in *Proceedings of the Australasian Conference on Robotics and Automation 2004*, Claude Sammut, Ed., 5–7 December 2004.

[5] Simukai W. Utete, *Network Management in Decentralised Sensing Systems*, Ph.D. thesis, Robotics Research Group, Department of Engineering Science, University of Oxford, January 1995.

[6] J. K. Uhlmann, *Dynamic Map Building and Localization for Autonomous Vehicles*, Ph.D. thesis, University of Oxford, 1995.

[7] R. Mahler, "Optimal / robust distributed data fusion: A unified approach," in *The Proceedings of AeroSense: The 14th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, Ivan Kadar, Ed., Orlando, FL, USA, August 2000 2000, SPIE, vol. 4052 of *Signal Processing, Sensor Fusion, and Target Recognition IX*, pp. 128–138, SPIE.

[8] M. Hurley, "An information-theoretic justification for covariance intersection and its generalization," in *Proceedings of the 2002 FUSION Conference*, Annapolis, MD, USA, 7–11 July 2002.

[9] I. Hwang, K. Roy, H. Balakrishnan, and C. Tomlin, "A distributed multiple-target identity management algorithm in sensor networks," in *The Proceedings of the 43rd IEEE Conference on Decision and Control*, December 2004.

[10] Simon J. Julier, "An empirical study into the use of chernoff information for robust, distributed fusion of gaussian mixture models," in *Proceedings of the 2006 FUSION Conference*, Florence, Italy, 10–13 July 2006.

[11] Tom Heskes, "Selecting weighting factors in logarithmic opinion pools," *Advances in Neural Information Processing Systems*, , no. 10, pp. 266–272, 1998.

[12] Martin E. Hellman and J. Raviv, "Probability of error, equivocation and the chernoff bound," *IEEE Transactions on Information Theory*, vol. IT–16, pp. 368–372, July 1970.

[13] Thomas Cover and Jay Thomas, *Elements of Information Theory*, chapter 12, pp. 312–315, Wiley-Interscience, 12 August 1991.

[14] A G. Dabak, *A Geometry for Detection Theory*, Ph.D. thesis, Dept. of Electrical and Computer Engineering, Rice University, Houston, TX, USA, September 1992.

[15] Paul Thompson, "General conservative fusion *is* conservative," Tech. Rep., Australian Field Robotics Centre, University of Sydney, Sydney, NSW, Australia, 27 February 2006.