CONF-961214--1

Distributed Decision Fusion Using Empirical Estimation †

Nageswara S. V. Rao
Center for Engineering Systems Advanced Research
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6364
nrao@dizzy.epm.ornl.gov

Submitted to *1996 IEEE/SICE/RSJ International Conference on Multisensor Fusion and Integration for Intelligent Systems*, December 8-11, 1996, Washington D.C.

†Research sponsored by the Engineering Research Program of the Office of Basic Energy Sciences, of the U.S. Department of Energy, under Contract No. DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp.

Distributed Decision Fusion Using Empirical Estimation [1]

Nageswara S. V. Rao
Center for Engineering Systems Advanced Research
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6364
raons@ornl.gov

Abstract

The problem of optimal data fusion in multiple detection systems is studied in the case where training examples are available, but no *a priori* information is available about the probability distributions of errors committed by the individual detectors. Earlier solutions to this problem require some knowledge of the error distributions of the detectors, for example, either in a parametric form or in a closed analytical form. Here we show that, given a sufficiently large training sample, an optimal fusion rule can be implemented with an arbitrary level of confidence. We first consider the classical cases of Bayesian rule and Neyman-Pearson test for a system of independent detectors. Then we show a general result that any test function with a suitable Lipschitz property can be implemented with arbitrary precision, based on a training sample whose size is a function of the Lipschitz constant, number of parameters, and empirical measures. The general case subsumes the cases of non-independent and correlated detectors.

Keywords and Phrases: Distributed decision fusion, empirical estimation, Bayesian rule, Neyman-Pearson test.

[1]Research sponsored by the Engincering Research Program of the Office of Basic Energy Sciences, of the U.S. Department of Energy, under Contract No. DE-AC05-96OR22464 with Lockheed Martin Energy Research Corp.

1 Introduction

The problem of fusing the decisions made by a set of distributed sensors or agents has been extensively studied in areas such as political economy models (Grofman and Owen [8]), reliability (von Neumann [21]), forecasting (Granger [7]), pattern recognition (Chow [4]), neural networks (Hashem *et al.* [9]), decision fusion (Dasarathy [5]), etc. One of the well-studied problems in the subarea of decision fusion deals with combining the decisions taken by the individual detectors or agents (Chair and Varshney [3], Thomopoulos *et al.* [18], Drakopoulos and Lee [6]). Typically, in this case the decision rule is in the form of a Bayesian rule [3] or Neyman-Pearson test [18, 6]. Such rule can be derived both in the case of independent and correlated individual decisions. In either case, some knowledge of the underlying probabilities is needed for an accurate implementation of the test; typically, analytical expressions for the error distributions in a computationally convenient form are needed. It is generally understood that the knowledge about the system can be utilized by the designer to obtain the required information about the probability distributions. In turn, this knowledge, could be based on the experience with the system, possibly in the form of empirical data generated by the system during experimentation or operation.

In this paper, we study systematic methods to utilize the empirical data to implement a fusion rule in the case no information about the probability distributions is available (along the lines of Naim and Kam [11]). Empirical and structural risk minimization have been used to solve a number of problems using empirical data (Vapnik [20]), and, more recently, similar formulations have been studied under the topic of machine learning (Blumer *et al.* [2], Natarajan [12]). If no constraints are placed on the underlying probabilities, the empirical data, which is finite, can only result in an approximate implementation of the required fusion rule (which is typically chosen from a class of functions). The degree of approximation between a fusion rule that can be obtained if the underlying probabilities are known and its empirical implementation based on a finite sample, depends on the sample size. We obtain the required sample sizes in order to arbitrarily bound the probability of disagreement between a fusion rule its empirical implementation. We wish to emphasize here that the proposed technique is to be used mainly when accurate estimates of the probabilities are either not available by other means or computationally difficult. For example, for a system of independent detectors if the exact analytical form of the probabilities are available, the methods of Chair and Varshney [3] and Thomopoulos *et al.* [18] could be used to implement the required fusion rule [2]; these methods, however, cannot be applied if the probabilities are unknown. In some cases, typically in a system of non-independent detectors, even if the suitable analytical forms of the distributions are available, the problem of implementing a Bayesian test could be computationally intractable (NP-Complete) [16, 19]. In such cases, Monte Carlo simulation can be used to generate the empirical data and the methods proposed here can still be used to implement the test in some cases (as illustrated in the example of Section 4). However, in some other situations, such as those employing the correlation coefficients methods (Drakopoulos and Lee [6]), the proposed empirical estimation method is not practical since the sizes of the samples to be generated by Monte Carlo methods grow

[2] Solutions to similar problems in the area of pattern recognition [4], and group decision models [8] yield almost identical analytical solutions.

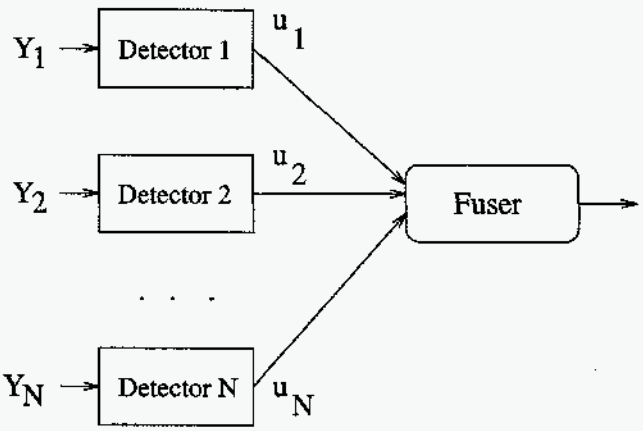

Figure 1: Parallel sensor suite.

exponentially with the number of parameters.

Consider a distributed detection system consisting of N sensors or detectors and a fusion center in the form of a parallel suite [5] as shown in Fig. 1. Only the fusion problem is addressed here in that the local detectors are given (i. e., we do not address the problem of designing the local detectors). We consider a simple hypothesis testing problem whereby the hypothesis H_0 is to be tested against the alternative hypothesis H_1. Each detector D_i, for $i = 1, 2, \ldots, N$ makes a decision $u_i \in \{H_0, H_1\}$, and the fusion center receives $u = (u_1, u_2, \ldots, u_N)$ and outputs either H_0 or H_1 by suitably using the information u. Let π_i denote the *a priori* probability of H_i, for $i = 0, 1$. For example, in one of the general formulations, the average-cost criterion is optimized by the likelihood ratio test [10] given by

$$T(u) = \frac{P(u|H_1)}{P(u|H_0)} > \frac{\pi_0(C_{10} - C_{00})}{\pi_1(C_{01} - C_{11})} \tag{T.1}$$

where C_{kj} denotes the cost of deciding H_k when H_j is true, $k, j = 0, 1$, and $P(\{u|H_i\}) = P(u|H_i)$ denotes [3] the conditional probability of u given H_i, $i = 1, 2$. The decision of fusion center is H_1 if the above test evaluates to true and is H_0 otherwise. If the underlying probabilities are available in a convenient form, then $T(u)$ can be computed at given u. One of the most studied formulations of this problem deals with the case where u_i's are independent, in which case $T(u)$ takes a simple form in terms of products [5].

In this paper, we consider that the probabilities needed to evaluate the tests of the form (T.1) are unknown, but a *sample* is available in the form of $(u^1, H^1), (u^2, H^2), \ldots, (u^l, H^l)$, where $u^i \in \{H_0, H_1\}^N$ is the ith example and $H^i \in \{H_0, H_1\}$ is the corresponding correct hypothesis. In object recognition systems that are required to detect when an object belonging to a certain object class enters a workspace, each individual detector could base its decision on possibly different object features. Examples in such system can be obtained by sensing the objects that belong to the object class and also the objects that do not. This formulation has been motivated by the sensor fusion problems that arise in robotic applications, where the individual sensors have been built and mounted on the robot. As an example, consider

[3] Here $P(A)$ denotes the probability measure of A which is a set belonging to a σ-algebra on the underlying event space ([1]). For a singleton $A = \{a\}$, we denote $P(A)$ by $P(a)$ with the usual abuse of notation.

training a mobile robot system equipped with an array of ultrasonic and infrared sensors to distinguish between the doors that are wide enough to go through from the narrower ones. An accurate probabilistic model of sensor errors is difficult to obtain, but a training sample can be easily obtained in a laboratory, for example, by placing boxes to create doors of required widths. Typically in these applications, obtaining accurate probabilistic models of the sensors is a more challenging task than performing experiments by sensing objects with known features and collecting the empirical data (situation in other applications, however, could be significantly different).

In our formulation only a *finite sample* is given as opposed to the formulae for the underlying probabilities required to implement Bayesian rule or Neyman-Pearson test. As a result only an approximate implementation of the required test is possible. We characterize the degree of approximation in terms of a confidence parameter which is a function of the sample size.

The fusion rule for decision problems is often expressed in terms of the probabilities, $p = (p_1, p_2, \ldots, p_n)$, and the data, $u = (u_1, u_2, \ldots, u_N)$, in the form

$$R(p, u) > 0, \tag{T.2}$$

where the decision is H_1 if the inequality is true and is H_0 otherwise. This form of the test captures many decision rules, for example, the test (T.1) can be easily converted into this form. In some typical cases, $n = 2N$ or $n = 2^N$. In the case the underlying probabilities are known, these expression for $R(p, u)$ for given u can be explicitly evaluated. In our case, we use estimators $\hat{p}_i$ based on the sample $(u^1, H^1), (u^2, H^2), \ldots, (u^l, H^l)$, and employ the *empirical implementation* $R(\hat{p}, u)$ of $R(p, u)$, where $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_N)$. If some of the probabilities are known in suitable forms, they can be directly used in the test, and the empirical estimates are used only for the unknown ones. Now we consider the performance measure

$$E_u[|\Theta[R(p, u)] - \Theta[R(\hat{p}, u)]|] = \sum_u |\Theta[R(p, u)] - \Theta[R(\hat{p}, u)]| P(u),$$

which is the expected error associated with $R(\hat{p}, u)$, where $\Theta[x]$ is 1 if x is non-negative and 0 otherwise. We define that $R(\hat{p}, u)$ *implements* $R(p, u)$ *with confidence* $1 - \lambda$ if

$$P[\Theta[R(p, u)] \neq \Theta[R(\hat{p}, u)]] < \lambda$$

or equivalently

$$E_u[|\Theta[R(p, u)] - \Theta[R(\hat{p}, u)]|] < \lambda$$

for sufficiently (but finite) large sample of size $l < \infty$. Informally, this condition means that based on a sufficiently large sample, it should be possible to ensure that both $R(p, u)$ and $R(\hat{p}, u)$ yield the same result with a probability of at least $1 - \lambda$.

In this paper, we show that the above criterion can indeed be guaranteed under the assumption that the function $R(p, u)$ is Lipschitz with respect to p, i. e. these exists a positive constant L such that

$$|R(p + \Delta p, u) - R(p, u)| < L||\Delta p||$$

for all Δp, u, where $||\Delta p||$ denotes the Euclidean norm of Δp in $\Re^n$. The sample size required to ensure a confidence of $1-\lambda$ is given by (see Section 3.2 for details)

$$l = \left\lceil \frac{r^2 n L^2}{2(R(\hat{p},u))^2} \ln(2n/\lambda) \right\rceil$$

for any $r > 1$, where $\lceil x \rceil$, for real x, denotes the smallest integer larger than or equal to x. We show that several existing fusion rules satisfy the required Lipschitz condition. In particular, in the case of independent detectors this condition is satisfied for the Bayesian case [3] and also for Neyman-Pearson test [18]. Also this condition is satisfied in the case of non-independent detectors formulated in terms of correlation coefficients [6].

In several simple cases, the required test may not satisfy Lipschitz condition. For example, consider a decision rule $R_1(u,p)$ that returns 1 if $p \geq 1/2$ and 0 otherwise. This function has a discontinuity at $p = 1/2$ which violates the Lipshitz condition since L is unbounded. In several cases, a practical solution could be obtained by using a smoothing function. For instance, here $R_1(u,p)$ can be replaced by $R_2(u,p) = \frac{1}{1+e^{-C(p-1/2)}}$, for a suitably large $C \geq 1$, which yields the Lipschitz constant $L = C/4$; this solution is useful if the likely values of p are away from $1/2$.

Our sample size estimate may not be tight in specific cases. Tighter bounds on the sample sizes than that indicated above can be obtained by exploiting the specific structure of $R(p,u)$ rather than using the general formula (as illustrated in Section 3.1).

The present formulation is very similar in spirit to that of Naim and Kam [11]. Our results are applicable to any test that satisfies the Lipschitz condition, including cases where the statistical independence is not satisfied, whereas the formulation of [11] is based on independence. In terms of the nature of the results, we provide finite sample estimates for guaranteeing specified confidence levels, and no small sample analysis is presented for the method of [11] (asymptotic convergence of their method can be asserted using the law of large numbers).

The organization of this paper is as follows. Some preliminaries are presented in Section 2. The solution to the decision fusion problem is presented in Section 3. In section 3.1, we consider the well-studied cases of independent detectors; although these cases can be derived as corollaries to the general result, we illustrate how the special nature of this formulation can be used to get sharper bounds on the sample size. In Section 3.2, the sample size estimates are derived for the empirical implementation of any test with the Lipschitz property. In Section 4, an example is illustrated to compare the proposed fusion rule with the existing rules based on Bayesian rule.

2 Preliminaries

We now present two lower bounds for the probability of simultaneous occurrence of a set of events in terms of the bounds for the occurrence of the individual events; one bound is tighter for small number of events and the other for large.

Lemma 2.1 *Consider events A_i, $i = 1, 2, \ldots, N$ such that $P(A_i) \geq 1 - \delta$, where A_i is a set belonging to a suitable σ-algebra. Then we have* [4]

(i) $P(A_1 \cap A_2 \cap \ldots \cap A_N) \geq 1 - N\delta$

(ii) $P(A_1 \cap A_2 \cap \ldots \cap A_N) \geq 1/2^N - \delta(1 - 1/2^N)$

(iii) The estimate in (ii) is increasingly better than (i) for large values of N, in particular for any N such that

$$\delta \geq \frac{(1 - 1/2^N)}{N - (1 - 1/2^N)}.$$

Proof: We have $P(\tilde{A}_i) < \delta$, for $i = 1, 2, \ldots, N$, where $\tilde{A}_i$ is the complement of the event A_i. Then $P(\tilde{A}_1 \cup \ldots \cup \tilde{A}_N) < N\delta$ which shows Part (i).

We show Part (ii) by induction on N. As a base case, $N = 1$, we have

$$P(A_1) \geq 1 - \delta \geq 1/2 - \delta/2$$

which is true since $\delta \leq 1$. For $N = 2$ we have $2P(A_1 \cap A_2) + P(A_1 \Delta A_2) = P(A_1) + P(A_2)$ which yields

$$P(A_1 \cap A_2) \geq \frac{1}{2}[P(A_1) + P(A_2) - 1] = 1/2 - \delta.$$

Now we have $1/2 - \delta \geq 1/4 - \delta 3/4$ which shows that the claim is true for $N = 2$. For the inductive step assume that the claim is true for $N = k$. We now show that the calim is true for $N = k + 1$ as follows.

$$\begin{aligned} P(A_1 \cap A_2 \cap \ldots \cap A_k \cap A_{k+1}) &\geq \frac{1}{2}[P(A_1 \cap A_2 \cap \ldots \cap A_k) + P(A_{k+1}) - 1] \\ &\geq \frac{1}{2}\left[1/2^k - \delta(1 - 1/2^k) + 1 - \delta + 1\right] \\ &= 1/2^{k+1} - \delta(1 - 1/2^{k+1}). \end{aligned}$$

Hence part (ii) is true. Part (iii) follows by direct algebraic manipulation. □

Notice that the condition in (iii) above becomes $\delta > 0$ as $N \to \infty$; some values for the lower bound on δ are shown in Table 1.

3 Empirical Implementation of Fusion Rules

We obtain the sample sizes needed to ensure that a fusion rule can be implemented based on a sample with a given value of confidence. We first consider the cases of independent detectors and then discuss the general case applicable to any fusion rule that satisfies the Lipschitz condition.

[4] One can also derive the bound $P(A_1 \cap A_2 \cap \ldots \cap A_N) \geq 1/2^{N-1} - \delta$ in an identical fashion as suggested by an anonymous reviewer; the present formulae are chosen for their convenient form in the present context.

N	lower bound	N	lower bound
1	1.000000	80	0.012658
2	0.600000	90	0.011236
3	0.411765	100	0.010101
4	0.306122	200	0.005025
5	0.240310	300	0.003344
10	0.110991	400	0.002506
20	0.052632	500	0.002004
30	0.034483	600	0.001669
40	0.025641	700	0.001431
50	0.020408	800	0.001252
60	0.016949	900	0.001112
70	0.014493		

Table 1: Some values for the lower bound of δ.

3.1 Independent Hypotheses

We consider the formulation of Chair and Varshney [3] for a system of independent learners. In the case the *a priori* distributions of the hypotheses are known, the fusion rule can be expressed in the form, for $n \geq 1$

$$\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i > 0 \qquad \text{(T.3.1)}$$

where q_i and s_i are the probabilities of suitable events (defined precisely in [3]), say denoted by Q_i and S_i respectively. An empirical implementation of (T.3.1) is given by

$$\prod_{i=1}^{n} \hat{q}_i - \prod_{i=1}^{n} \hat{s}_i > 0$$

where $\hat{q}_i$ and $\hat{s}_i$ are empirical estimates of q_i and s_i respectively based on the l-sample (i. e. , $\hat{q}_i$ is the fraction of the times the event Q_i took place in the sample). For this rule the sample size to ensure the required confidence is given in the following theorem.

Theorem 3.1 *For any $r > 2$, consider a sample of size*

$$l = \left\lceil \frac{1}{2\epsilon_L^2} \ln(2/\delta) \right\rceil$$

where

$$\epsilon_L = \left(1 + \frac{|\prod_{i=1}^{n} \hat{q}_i - \prod_{i=1}^{n} \hat{s}_i|}{r+2} \right)^{1/n} - 1.$$

Then the test $\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i > 0$ can be implemented by the empirical measures of q_i and s_i computed based on the sample with the confidence $1 - 2n\delta$ or $1/2^{2n} - \delta(1 - 1/2^{2n})$.

Proof: The proof is separated into two parts. First, we compute sufficiency conditions on $|q_i - \hat{q}_i|$ and $|s_i - \hat{s}_i|$ such that $\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i > 0$ and $\prod_{i=1}^{n} \hat{q}_i - \prod_{i=1}^{n} \hat{s}_i > 0$ yield the same result. Second we compute the sample size require to ensure the required sufficiency conditions.

The test (T.3.1) and its empirical implementation yield the same result under the conditions

$$|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} \hat{q}_i| < \frac{1}{r}|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i| \tag{3.1.1}$$

and

$$|\prod_{i=1}^{n} s_i - \prod_{i=1}^{n} \hat{s}_i| < \frac{1}{r}|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i| \tag{3.1.2}$$

for any $r > 2$. Consider the constraint $|q_i - \hat{q}_i| < \epsilon$ for all $i = 1, 2, \ldots, n$ which implies (Theorem 3.1 of [17])

$$|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} \hat{q}_i| \leq (1+\epsilon)^n - 1.$$

Thus the condition $|q_i - \hat{q}_i| < \epsilon$ and $|s_i - \hat{s}_i| < \epsilon$, for all $i = 1, 2, \ldots, n$, and

$$\epsilon = \left(1 + \frac{|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i|}{r}\right)^{1/n} - 1 \tag{3.1.3}$$

ensures that conditions (3.1.1) and (3.1.2) are satisfied.

For the second part, by Hoeffding's inequality (for example, we use the form given in Appendix B of Pollard [13]) we have

$$P[|q_i - \hat{q}_i| > \epsilon] < 2e^{-2\epsilon^2 l}$$

and

$$P[|s_i - \hat{s}_i| > \epsilon] < 2e^{-2\epsilon^2 l}$$

for any $i \in \{1, 2, \ldots, n\}$. Then $2e^{-2\epsilon^2 l} = \delta$ yields the sample size of

$$l = \left\lceil \frac{1}{2\epsilon^2} \ln(2/\delta) \right\rceil. \tag{3.1.4}$$

Hence by Part (i) and (ii) of Lemma 2.1, with confidence $1 - 2n\delta$ or $1/2^{2n} - \delta(1 - 1/2^{2n})$ we have $|q_i - \hat{q}_i| < \epsilon$ and $|s_i - \hat{s}_i| < \epsilon$ simultaneously for all $i = 1, 2, \ldots, n$.

Now to obtain the sample size stated in the theorem, we employ $\epsilon_L \leq \epsilon$ in (3.1.4) which guarantees the required condition. Notice that from (3.1.1) and (3.1.2) we have

$$\begin{aligned} |\prod_{i=1}^{n} \hat{q}_i - \prod_{i=1}^{n} \hat{s}_i| &\leq |\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i| + \frac{2}{r}|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i| \\ &= \frac{r+2}{r}|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i|. \end{aligned}$$

Now ϵ_L is obtained by replacing $|\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i|$ in the expression for ϵ by $\frac{r}{r+2}|\prod_{i=1}^{n} \hat{q}_i - \prod_{i=1}^{n} \hat{s}_i|$. □

Remark 3.1 The sample size given by the expression (3.1.4) based on ϵ yields smaller sample size compared to the one based on ϵ_L given in Theorem 3.1. But the latter has more practical utility in that it can be used to estimate the precision with which the test is implemented with a given sample of size l. Notice that the expression for l based on ϵ_L involves only the estimated quantities; thus, using $\hat{q}_i$ and $\hat{s}_i$ one can compute ϵ_L which can be used in the expression for l to compute the confidence achieved at a given sample size. Such computation is not possible based on (3.1.4) since it involves the unknown ϵ.

Remark 3.2 By denoting the confidence $1 - 2n\delta$ or $1/2^{2n} - \delta(1 - 1/2^{2n})$ by $1 - \lambda$ one can obtain following expressions for the sample sizes

$$l = \left\lceil \frac{1}{2\epsilon_L^2} \ln(4n/\lambda) \right\rceil$$

or

$$l = \left\lceil \frac{1}{2\epsilon_L^2} \ln\left(\frac{2(1 - 1/2^{2n})}{\lambda - (1 - 1/2^{2n})} \right) \right\rceil$$

respectively where the sample sizes are expressed directly in terms of λ.

We now consider the formulation of Thomopoulos *et al.* [18] which does not require the *a prori* probabilities. In this case the fusion rule can be expressed as the Neyman-Pearson test in the form

$$\prod_{i=1}^{n} q_i - \tau \prod_{i=1}^{n} s_i > 0 \qquad \text{(T.3.2)}$$

where the positive real τ is fixed by the type I and II errors. The empirical implementation of this test can be obtained as in the above case.

Corollary 3.1 *Consider a sample of size*

$$l = \left\lceil \frac{1}{2\epsilon_L^2} \ln(2/\delta) \right\rceil$$

where

$$\epsilon_L = \left(1 + \frac{|\prod_{i=1}^{n} \hat{q}_i - \tau \prod_{i=1}^{n} \hat{s}_i|}{1 + \tau} \right)^{1/n} - 1.$$

Then the test $\prod_{i=1}^{n} q_i - \tau \prod_{i=1}^{n} s_i > 0$ *can be implemented by using the empirical measures* $\hat{q}_i$ *and* $\hat{s}_i$ *based on the sample with confidence* $1 - 2n\delta$ *or* $1/2^{2n} - \delta(1 - 1/2^{2n})$.

Proof: The outline of the proof is identical to that in Theorem 3.1. The condition $|q_i - \hat{q}_i| \leq \epsilon$ and $|s_i - \hat{s}_i| \leq \epsilon$ for all $i = 1, 2, \ldots, n$ yields

$$\left| \prod_{i=1}^{n} \hat{q}_i - \prod_{i=1}^{n} q_i - \tau \left(\prod_{i=1}^{n} \hat{s}_i - \prod_{i=1}^{n} s_i \right) \right| \leq (1 + \tau)[(1 + \epsilon)^n - 1].$$

Under this condition the test (T.3.2) and its empirical implementation yield the same result if $(1+\tau)[(1+\epsilon)^n - 1] \leq |\prod_{i=1}^{n} q_i - \tau \prod_{i=1}^{n} s_i|$. This yields an expression for ϵ as in Theorem 3.1. Then, the value of ϵ_L is obtained by noting that $2\left|\prod_{i=1}^{n} q_i - \tau \prod_{i=1}^{n} s_i\right| \geq \left|\prod_{i=1}^{n} \hat{q}_i - \tau \prod_{i=1}^{n} \hat{s}_i\right|$. □

Remark 3.3 A slightly different execution of Corollary 3.1 can be obtained by using $r > 2$ which yields

$$\epsilon_L = \left(1 + \frac{|\prod_{i=1}^{n} \hat{q}_i - \tau \prod_{i=1}^{n} \hat{s}_i|}{(r+2)(1+\tau)}\right)^{1/n} - 1.$$

Compared to the quantity in the Corollary 3.1, this yields a smaller value and hence requires a larger sample to ensure the same confidence.

3.2 General Case

In this section, we consider an empirical implementation of a general test

$$R(p,u) > 0, \tag{T.3.3}$$

where $p = (p_1, p_2, \ldots, p_n)$ is the vector of probabilities with respect to which $R(p,u)$ is Lipschitz with constant L. Let p_i correspond to an event P_i and let the empirical estimate $\hat{p}_i$ of p_i be the fraction of the time the event P_i takes place in the sample.

Theorem 3.2 *Consider a decision rule $R(p,u)$ with Lipschitz constant L. For any $r \geq 2$, given a training sample of size*

$$l = \left\lceil \frac{r^2 n L^2}{2(R(\hat{p},u))^2} \ln(2/\delta) \right\rceil$$

the empirical implementation $R(\hat{p},u) > 0$ implements the test $R(p,u) > 0$ with confidence $1 - n\delta$ or $1/2^n - \delta(1 - 1/2^n)$.

Proof: The outline of the proof follows that of Theorem 3.1. First, we show that the conditions $\sup_i |p_i - \hat{p}_i| < \epsilon$ and $|R(p)| \geq L\sqrt{n}\epsilon$, ensure that $R(p,u) > 0$ and $R(\hat{p},u) > 0$ yield the same result. Notice that $\sup_i |p_i - \hat{p}_i| < \epsilon$ implies $||p - \hat{p}|| < L\sqrt{n}\epsilon$. Then by the Lipschitz property

$$|R(\hat{p},u) - R(p,u)| \leq ||p - \hat{p}|| \leq L\sqrt{n}\epsilon.$$

Thus if $|R(p,u)| \geq L\sqrt{n}\epsilon$, both $R(p,u)$ and $R(\hat{p},u)$ yield the same result.

Then for $\epsilon = \frac{|R(p,u)|}{L\sqrt{n}}$, given a sample of size

$$l = \left\lceil \frac{1}{2\epsilon^2} \ln(2n/\delta) \right\rceil \tag{3.2.1}$$

$R(\hat{p},u)$ implements $R(p,u)$ with the required precision. Now we obtain a lower bound for ϵ by noting that $2|R(p,u)| \geq |R(\hat{p},u)|$ which implies $|R(p,u)| \geq \frac{1}{r}|R(\hat{p},u)|$ for any $r \geq 2$. The

sample size in the hypothesis is obtained by using the lower bound $\frac{|R(\hat{p},u)|}{rL\sqrt{n}}$ for ϵ. □

Remark 3.4 The test (T.3.1) can be shown to be Lipschitz and the Theorem 3.2 can be applied. For a scalar ϵ, let $p+\epsilon$ denote the vector obtained by adding ϵ to each component of p. For example, we can show that $|R(p+\epsilon,u) - R(p,u)| \leq 2[(1+\epsilon)^{n/2} - 1]$, where $n/2$ refers to the quantity in test (T.3.4). By noting that $(1+\epsilon)^{n/2} - 1 = \sum_{i=1}^{n/2} \binom{n/2}{i} \epsilon^i$ which is upper-bounded by $\epsilon \sum_{i=1}^{n/2} \binom{n/2}{i} \leq 2^{n/2}\epsilon$, we can obtain $L = 2^{n/2}$ which can be used to obtain the sample size based on Theorem 3.2. The result of Theorem 3.1 based on the special structure of (T.3.1) yielded us a sharper result on the sample size.

Now consider the following particular form of the test, for a positive real τ,

$$P(u|H_1) - \tau P(u|H_0) > 0 \qquad (T.3.4)$$

which is the Neyman-Pearson test from which several specific tests (including the ones in the last two sections) can be derived.

Corollary 3.2 *Consider a decision rule $P(u|H_1) - \tau P(u|H_0) > 0$, where τ is a positive real number. Let the empirical estimate $\hat{P}(A)$ of $P(A)$ be the fraction of times the event A took place based on the sample.*

(i) Given a training sample of size

$$l = \left\lceil \frac{4(1+\tau)^2}{[\hat{P}(u|H_1) - \tau\hat{P}(u|H_0)]^2} \ln(4/\lambda) \right\rceil$$

the empirical rule $\hat{P}(u|H_1) - \tau\hat{P}(u|H_0) > 0$ implements the test $P(u|H_1) - \tau P(u|H_0) > 0$ with confidence $1-\lambda$.

(ii) Given a training sample of size

$$l = \left\lceil \frac{72(1+\tau)^2}{[\hat{P}(u \cap H_1)\hat{P}(H_0) - \tau\hat{P}(u \cap H_0)\hat{P}(H_1)]^2} \ln(8/\lambda) \right\rceil$$

the empirical test $\hat{P}(u \cap H_1)\hat{P}(H_0) - \tau\hat{P}(u \cap H_0)\hat{P}(H_1) > 0$ implements the test $P(u|H_1) - \tau P(u|H_0) > 0$ with confidence $1-\lambda$.

Proof: For Part (i), the sample size is estimated based on the two probabilities $p_1 = P(u|H_1)$ and $p_2 = P(u|H_0)$ (hence $n = 2$), and noting that for $|\hat{p}_1 - p_1| < \epsilon$ and $|\hat{p}_2 - p_2| < \epsilon$, we have

$$|\hat{p}_1 - \tau\hat{p}_2 - (p_1 - \tau p_2)| = |\hat{p}_1 - p_1 - \tau(\hat{p}_2 - p_2)| = (1+\tau)\epsilon.$$

Now $1+\tau$ is used in the place of the Lipschitz constant in the Theorem 3.2 (with $r = 2$) to show Part (i) (in a strict sense, we have not established that $1+\tau$ is a suitable Lipschitz constant, but the derivation is valid under slightly weaker condition).

For Part (ii), we use the following form of the given test

$$P(u \cap H_1)P(H_0) - \tau P(u \cap H_0)P(H_1) > 0$$

where we estimate the probabilities $p_1 = P(u \cap H_1)$, $p_2 = P(H_0)$, $p_3 = P(u \cap H_0)$, and $p_4 = P(H_1)$ (hence $n = 4$). Now we have

$$|\hat{p}_1\hat{p}_2 - \tau\hat{p}_3\hat{p}_4 - (p_1p_2 - \tau p_3p_4)| \leq (1+\tau)(\epsilon^2 + 2\epsilon) \leq 3(1+\tau)\epsilon$$

and sample size is obtained by using $3(1+\tau)$ in place of the Lipschitz constant. □.

Remark 3.5 In the two cases of the Corollary 3.2 it can be shown that the quantities $(1+\tau)$ and $3(1+\tau)$ respectively yield the required Lipschitz constants.

Remark 3.6 In Corollary 3.2, $n = 2$ for (i) and $n = 4$ for (ii), for which the bound in Part (i) of Lemma 2.1 is better than the bound in Part (ii).

We now consider the fusion rule of Drakopoulos and Lee [6] which makes use of the correlation coefficients to obtain the fusion rule. The set of correlation coefficients are given by

$$\mathcal{C} = \{P[u_{i_1}...u_{i_k}|H_j] | \{i_1, \ldots, i_k\} \subseteq \{1, \ldots, N\}, j = 0, 1\}.$$

The fusion test is given by

$$P(u|H_1) - \tau P(u|H_0) > 0 \tag{T.3.5}$$

for suitable τ such that for $j = 0, 1$

$$P(u|H_j) = \sum_{I \subseteq A_0} (-1)^{|I|} P\left[\prod_{i \in A_1 \cup I} u_i | H_j\right]$$

where $A_k = \{i : u_i = k\}$ (see [6] for details), and I, of cardinality $|I|$, varies over all subsets of A_0. In our method, we estimate the correlation coefficients based on empirical data and use them to implement the test.

Corollary 3.3 *Consider a decision rule $P(u|H_1) - \tau P(u|H_0) > 0$ implemented in terms of the correlation coefficients. Let empirical estimate $\hat{P}(A)$ of $P(A)$ be the fraction of times the event A took place based on the empirical sample, and for $j = 0, 1$*

$$\hat{P}(u|H_j) = \sum_{I \subseteq A_0} (-1)^{|I|} \hat{P}\left[\prod_{i \in A_1 \cup I} u_i | H_j\right].$$

Given a training sample of size

$$l = \left\lceil \frac{r^2 2^{3N-1}(1+\tau)^2}{[\hat{P}(u|H_1) - \tau\hat{P}(u|H_0)]^2} \ln(2/\delta)) \right\rceil$$

the empirical rule implements $P(u|H_1) - \tau P(u|H_0) > \tau$ with confidence $1/2^N - \delta(1 - 1/2^N)$.

Proof: Here the critical point is that the Lipschitz constant is given by $(1+\tau)2^N$ as shown below. First we have for $|P(.) - \hat{P}(.)| < \epsilon$

$$|\hat{P}(u|H_j) - P(u|H_j)| = \sum_{I \subseteq A_0} (-1)^{|I|} \left(\hat{P}\left(\prod_{i \in A_1 \cup I} u_i | H_j \right) - P\left(\prod_{i \in A_1 \cup I} u_i | H_j \right) \right] \leq \sum_{I \subseteq A_0} \epsilon \leq 2^N \epsilon.$$

This condition yields

$$\hat{P}(u|H_1) - \tau \hat{P}(u|H_0) - [P(u|H_1) - \tau P(u|H_0)] \leq 2^N (1+\tau)\epsilon.$$

The rest of the proof follows by the application of Theorem 3.2 with $n = 2^N$. □

Remark 3.7 In Corollary 3.3, $n = 2^N$ for which the bound in Part (ii) of Lemma 2.1 is better than that in Part (i).

4 Example

We consider a system with 5 detectors, $D_1, \ldots, D_5$. All simulations are carried out using pseudo random number generators. The hypotheses H_1 and H_0 are generated with equal probabilities. This behavior is simulated by generating a uniformly distributed random variable over the interval $[0, 1]$ and checking to see if it lies in the interval $[0, 0.5]$. Each detector is given the hypothesis as input and it produces an output that disagrees with the input according to a probabilistic strategy. The detectors D_i, $i = 1, 2, \ldots, 5$ introduces an error as follows: with probability of $1 - i/10$ it passes on the input to output, and with probability $i/10$ it passes on the opposite of the input. The individual detector behavior is implemented by generating a uniform random variable in the range $[0, D]$ and checking if it falls in the interval $[0, iD/10]$. It is assumed that the pseudo random number generator yields independent outputs and the error processes of the individual detectors are probabilistically independent.

A sequence of examples has been generated and given as input to both the Bayesian fuser and the proposed empirical fuser. The Bayesian fuser is implemented by using the analytical formulae for the distribution of errors under the assumption of independence between the various detectors. The Bayesian test is given by $P(H_0|u) \geq P(H_1|u)$ where $u = (u_1, u_2, \ldots, u_5)$ and u_i is the output of D_i. Due to the independence and the property $P(H_0) = P(H_1) = 1/2$, this test is equivalent to the test:

$$\prod_{i=1}^{5} P_i(u_i|H_0) \geq \prod_{i=1}^{5} P_i(u_i|H_1)$$

where the $P_i(.)$ corresponds to D_i, $i = 1, 2, \ldots 5$, such that, for $j, k \in \{H_0, H_1\}$,

$$P_i(j|k) = \begin{cases} i/10 & \text{if } j \neq k \\ 1 - i/10 & \text{if } j = k \end{cases}.$$

For the empirical fuser, the probabilities are estimated based on the sample seen so far. Each example is given as input to both the fusers and their outputs are computed. An average

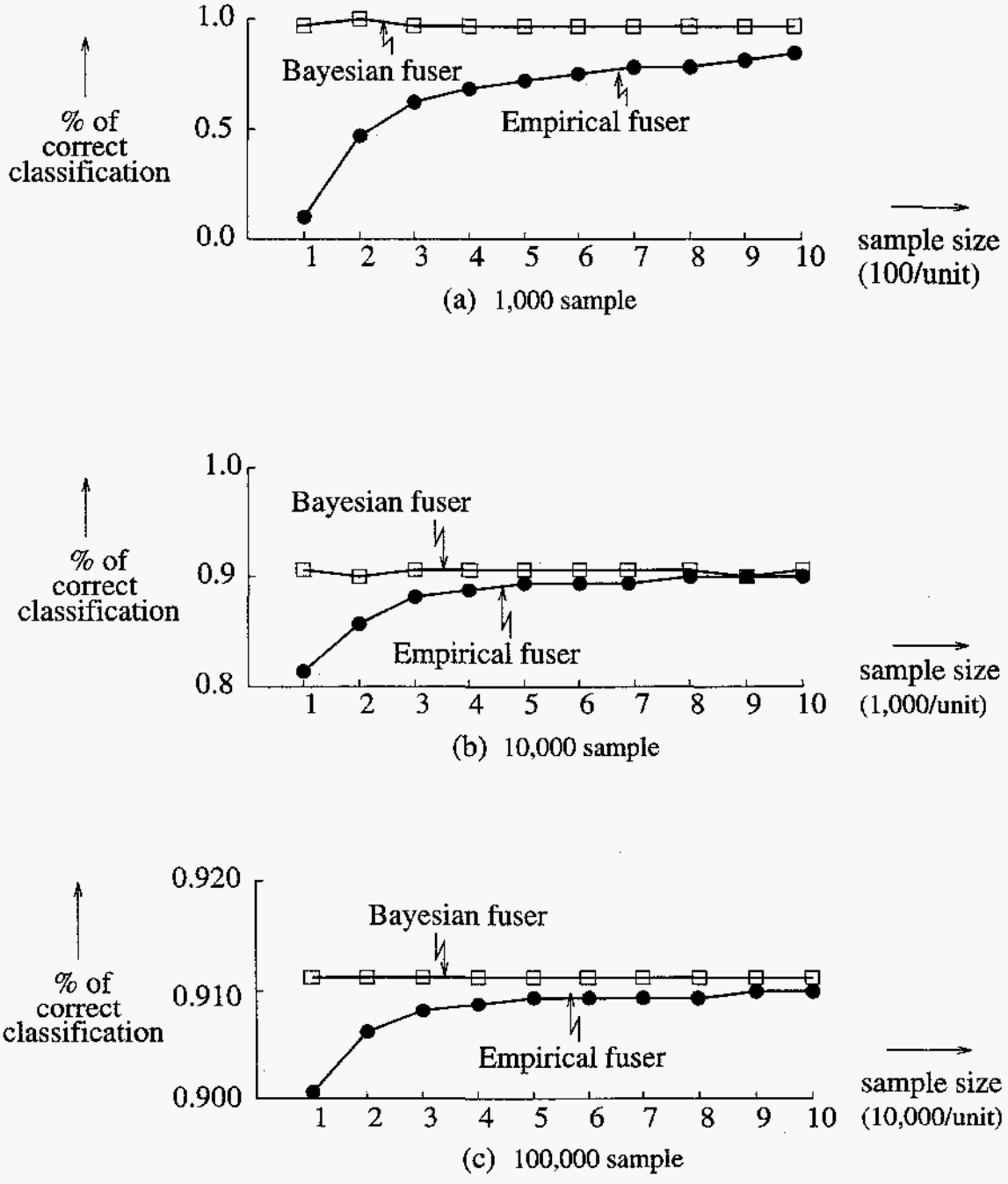

Figure 2: Relative performance of the Bayesian fuser and empirical fuser with training.

percentage of correctly classified examples (among the sample seen so far) is computed and shown in Fig. 2. The plots in Fig. 2 (a), (b) and (c) as shown for 1,000, 10,000, and 100,000 examples respectively. The performance of the empirical fuser has shown an improving trend as the training progressed. After 1,000 examples, the empirical fuser achieved the performance within 10% of that of the Bayesian fuser as shown in Fig. 2(a). After 10,000 examples the performance of the empirical user is approximately within 1% of that of Bayesian fuser (Fig. 2(b)). After 100,000 examples, the performance of the empirical fuser is within 0.1% of that of the Bayesian fuser. This simulation has been repeated with different starting seeds for the pseudo random number generator with almost identical qualitative behavior.

The estimates for the sample sizes needed to ensure a confidence of 90% is of the order of few millions for this system. Thus the performance of the method is better than that indicated by the samples estimates; it is, however, unclear if it an artifact of this specific example.

The training program has been implemented on SPARC workstation IPX. The execution time for the training with 100,000 examples, including the time required to generate the samples, is of the order of 20 seconds.

We wish to note here that in some cases, the problem of implementing a Bayesian fuser could be computationally expensive (since the problem could be NP-complete). In such cases, the information about the *a priori* probabilities and error distributions can be used to generate examples by using Monte Carlo methods, and the empirical fuser can be trained with the generated examples (as shown in the above example). The performance of such methods depends on the ease with which the examples can be generated and the characteristics of the programs that generate the pseudo random variables.

5 Conclusions

We have studied the problem of optimal data fusion in multiple detection systems in the case when training examples are available, but no information is available about the probability of errors committed by the individual detectors. Most existing solutions to this problem require some knowledge (for example, either a parametric or an analytical form) of the error distributions of the detectors. We showed that given sufficiently large training sample, an optimal fusion rule can be implemented with an arbitrarily high level of confidence. Since no information about the underlying probabilities is available, an *exact* implementation of the optimal rule which is chosen from a set of functions is not possible based on a finite set of examples. We showed a general result that any test function with a suitable Lipschitz property can be implemented based on a training sample with an arbitrarily high precision, where the sample is a function of the Lipschitz constant and the number of parameters. The general case subsumes the cases of non-independent and correlated detectors, but specific properties of the tests can be used to obtain sharper results than those yielded by the general result. Two other approaches based on empirical risk minimization and nearest neighbor rule for solving the present problem have been recently studied by Rao and Iyengar [15].

The proposed method is useful in systems where either the underlying probabilities are not known or the Bayesian test is too difficult to implement. As illustrated in Section 4, the proposed method can only approach the Bayesian method in performance after a significant

amount of training.

There are two generalizations of the formulation studied here. Rao [14] discusses the estimation of fusion rules in multiple sensor systems, and Rao and Oblow [17] discuss empirical implementation of optimal fusion rules for a system of probably approximately correct learners. These results are more general in that the finiteness properties (of the σ-algebra on the event space in formulation studied here) are not satisfied and consequently the sample size estimates are presented in terms of the Vapnik and Chervonenkis dimension [20].

There are several directions for future investigations. Extension of the proposed method for non-Lipschitz cases will be interesting. Also, although the Lipschitz property is sufficient for implementing an empirical fuser, it is clearly not necessary. We suspect that some "local" properties can replace the Lipschitz property (which is a global property). When the decision amounts to choosing one hypothesis from a *continuum* (instead of choosing one of the two alternatives), we are faced with more difficult problems. It would be of interest to investigate the multiple detectors systems to handle such formulations.

6 Acknowledgements

We also deeply appreciate the comments of Max Morris, Jacob Barhen, V. Ram Uppuluri and Vladimir Protopopescu on several parts of this paper.

References

[1] P. Billingsley. *Probability and Measure.* John Wiley and Sons, New York, second edition, 1986.

[2] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the Association of Computing Machinery*, 36(4):929–965, 1989.

[3] Z. Chair and P.K. Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Trans. Aerospace Electronic Syst.*, 22(1):98–101, 1986.

[4] C. K. Chow. Statistical independence and threshold functions. *IEEE Trans. Electronic Computers*, EC-16:66–68, 1965.

[5] B. V. Dasarathy. *Decision Fusion.* IEEE Computer Society Press, Los Alamitos, California, 1994.

[6] E. Drakopoulos and C. C. Lee. Optimal multisensor fusion of correlated local decision. *IEEE Trans. Aerospace Electronics Syst.*, 27(4):593–605, 1991.

[7] C. W. J. Granger. Combining forecasts — twenty years later. *J. of Forecasting*, 8:167–173, 1989.

[8] B. Grofman and G. Owen, editors. *Information Pooling and Group Decision Making.* Jai Press Inc., Greenwich, Connecticut, 1986.

[9] S. Hashem, B. Schmeiser, and Y. Yih. Optimal linear combinations of neural networks: An overview. In *Proceedings of 1994 IEEE Conf. on Neural Networks*, pages 1507–1512. 1994.

[10] W. A. Hashlamoun and P.K. Varshney. Further results on distributed bayesian signal detection. *IEEE Trans. on Information Theory*, 39(5):1660–1661, 1993.

[11] A. Naim and M. Kam. On-line estimation of probabilities for Bayesian distributed detection. *Automatica*, 30(4):633–642, 1994.

[12] B. K. Natarajan. *Machine Learning: A Theoretical Approach.* Morgan Kaufmann Pub. Inc., San Mateo, California, 1991.

[13] D. Pollard. *Convergence of Stochastic Processes.* Springer-Verlag, New York, 1984.

[14] N. S. V. Rao. Fusion rule estimation in multiple sensor systems with unknown noise densities. *Journal of Franklin Institute*, 331B(5):509–530, 1995.

[15] N. S. V. Rao and S. S. Iyengar. Distributed decision fusion under unknown distributions. *Optical Engineering*, 1996. in press.

[16] N. S. V. Rao, S. S. Iyengar, and R. L. Kashyap. Computational complexity of distributed detection problems with information constraints. *Computers and Electrical Engineering*, 19(6):445–451, 1993.

[17] N. S. V. Rao and E. M. Oblow. N-learners problem: System of PAC learners. In *Computational Learning Theory and Natural Learning Systems, Vol IV: Making Learning Practical.* MIT Press, 1996. in press.

[18] S. C. A. Thomopoulos, R. Viswanathan, and B. K. Bougoulias. Optimal and suboptimal distributed decision fusion. *IEEE Trans. Aerospace Electronics Syst.*, 25(5):761–765, 1989.

[19] J. N. Tsitsiklis and M. Athans. On the complexity of decentralized decision making and detection problems. *IEEE Transactions on Automatic Control*, AC-30(5):440–446, 1985.

[20] V. Vapnik. *Estimation of Dependences Based on Empirical Data.* Springer-Verlag, New York, 1982.

[21] J. von Neumann. Probabilistic logics and the synthesis of reliable organisms from unreliable components. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 43–98, 1956. Princeton University Press.