Data-Based Approaches to Decision Fusion

Nageswara S.V. Rao
Center for Engineering Systems Advanced Research
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6355

MSF Panel Meeting
Subpanel on Discrimination
August 24-25, 2000, Washington D.C.

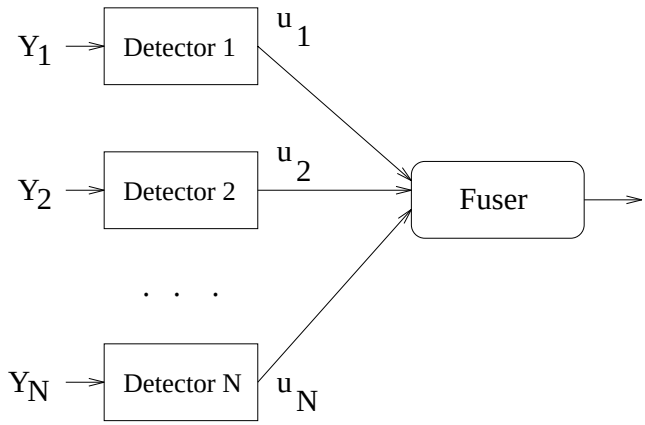
Outline of the Talk

1. **Introduction**
2. **Independent Hypotheses**
3. **Correlation Coefficients Method**
4. **Lipschitz Decision Rules**
5. **Correlation vs. Independence**
6. **Simulation Results**

References

1. E. Drakopoulos and C. C. Lee, Optimal multisensor fusion of correlated local decision. IEEE Trans AES, 27, 4(1991), 593-605.
2. N. S. V. Rao, Distributed decision fusion using empirical estimation, IEEE AES, 33, 4(1997), 1106-1114.

Parallel Sensor Suite



Detector System: $D_1, D_2, \ldots, D_N$

Detector: D_i
Makes a decision $u_i \in \{H_0, H_1\}$

Fusion Center:
Receives $u = (u_1, u_2, \ldots, u_N)$ outputs either H_0 or H_1

Notes:

1. Well-studied problem in different domains:
   democracy models (Condorcet 1786)
   composite methods (Laplace 1818)
   reliability (von Neumann 1956)
   pattern recongition (Chow 1965)
   Newer applications are being found in diverse areas
2. Of particular importance to distributed detection systems
   Extensively studied over the past decade
   Dasarathy (1994), Varshney (1996)

Data-Based Formulation

In our formulation:

(i) Probability distributions are NOT all known and/or validated,
(ii) Detector system is available, hence experimental samples can be collected, and
(iii) Known and validated distributions can be incorprated to better the guarantees.

Existing results:

For many formulations of problem, fusion rules have been derived under the complete knowledge of the distributions

Our Results:

The fusion rules designed under known distributions can be implemented when only a sample is available.

Empirical Implementation

Fusion rules : expressed in terms of
(i) probabilities $p = (p_1, p_2, \ldots, p_n)$
(ii) data $u = (u_1, u_2, \ldots, u_N)$
of the form

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

where

$$\text{decision is} \begin{cases} H_1 & \text{if the inequality is true} \\ H_0 & \text{otherwise.} \end{cases}$$

If probabilities are known:
$R(p, u)$ for given u can be explicitly evaluated.

In our formualtion:

1. Estimators $\hat{p}_i$ are computer based on iid sample
   $(u^1, H^1), (u^2, H^2), \ldots, (u^l, H^l)$
2. Empirical version given by

$$R(\hat{p}, u) > 0$$

is employed

Question: How good is $R(\hat{p}, u)$ compared to $R(p, u)$?

<u>Performance Criteria</u>

<u>Performance Measure:</u> Expected error of $R(\hat{p}, u)$,

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

where $\Theta[x]$ is 1 if x is non-negative and 0 otherwise.

<u>Performance Criterion:</u>
$R(\hat{p}, u)$ implements $R(p, u)$ with <u>confidence</u> $1 - \lambda$ 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,
based on a sufficiently large sample, $R(p, u)$ and $R(\hat{p}, u)$ yield the same result with a probability of at least $1 - \lambda$.

Independent Hypotheses

Formulation of Chair and Varshney (1986):
(i) independent detectors, (ii) a priori distributions are known
Fusion rule is of the form, for $n \geq 1$

$$\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i > 0 \tag{T.3.1}$$

where q_i and s_i are the probabilities of suitable events Q_i and S_i.

Empirical implementation of (T.3.1):

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

where $\hat{q}_i$ and $\hat{s}_i$ are empirical estimates of q_i and s_i respectively based on the l-sample.

Theorem 3.1: Rao (1997)
For any $r > 2$, consider a sample of size

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

where $\epsilon_L = \left(1 + \frac{|\prod_{i=1}^{n} \hat{q}_i - \prod_{i=1}^{n} \hat{s}_i|}{r+2}\right)^{1/n} - 1$. Empirical implementation of $\prod_{i=1}^{n} q_i - \prod_{i=1}^{n} s_i > 0$ has confidence

$$1/2^{2n} - \delta(1 - 1/2^{2n})$$

or

$$1 - 2n\delta$$

Neyman-Pearson Test

Thomopoulos et al. (1989): a prori probabilities not known
Under statistical independence, fusion rule is Neyman-Pearson test in the form

$$\prod_{i=1}^{n} q_i - \tau \prod_{i=1}^{n} s_i > 0 \tag{T.3.2}$$

where the positive real τ is fixed by the type I and II errors.

Corollary 3.1: Rao (1997)
Consider a sample of size

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

where $\epsilon_L = \left(1 + \frac{|\prod_{i=1}^{n} \hat{q}_i - \tau \prod_{i=1}^{n} \hat{s}_i|}{1+\tau}\right)^{1/n} - 1$. Empirical implementation of $\prod_{i=1}^{n} q_i - \tau \prod_{i=1}^{n} s_i > 0$ has confidence confidence

$$1 - 2n\delta$$

or

$$1/2^{2n} - \delta(1 - 1/2^{2n})$$

Neyman-Pearson Test

Particular form of the test, for a positive real τ,

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

Let $\hat{P}(A)$ be fraction of times event A took place in 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 $1 - \lambda$.

Correlation Coefficients Method:

Drakopoulos and Lee (1991):
Method of correlation coefficients

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

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 $A_k = \{i : u_i = k\}$ and I, of cardinality $|I|$, varies over all subsets of A_0.

Corollary 3.3: Rao (1997)
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 $1/2^N - \delta(1 - 1/2^N)$.

Lipschitz Test

$R(p, u)$ is Lipschitz with respect to p:
there 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 $\Re^n$.

Note:

1. $R(p, u)$ must be continuous in p
2. L is "upperbounded" by maximum gradient magnitude wrt p

Current Fusion Rules:

1. Large majority of published fusion rules are Lipschitz
2. There are several examples of non-Lipschitz tests
   some of these tests can be approximated by Lipschitz tests

Sample Size for Lipschitz Test

Theorem 3.2 Rao (1997)
Consider a decision rule $R(p, u)$ with Lipschitz constant L.
For any $r \geq 2$, given a training sample of size

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

$R(\hat{p}, u) > 0$ implements the test $R(p, u) > 0$ with confidence

$$1 - n\delta$$

or

$$1/2^n - \delta(1 - 1/2^n)$$

Proof Outline: Theorem 3.2

1. 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.

2. 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.

3. 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 $r \geq 2$.

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

Note:
Tighter sample bounds may be possible in particular cases,
e.g. Theorem 3.1

Statistical Independence

Distributions:
X - distribution P_X; Y - distribution P_Y
X, Y - joint distributions $P_{X,Y}$

$$P_{X,Y} = P_X P_Y$$

or equivalently

$$P_{X|Y} = P_X \qquad \text{and} \qquad P_{Y|X} = P_Y$$

Densities:
X - density $f_X(.)$; Y - density $f_Y(.)$

$$f_{X,Y}(.) = f_X(.) f_Y(.)$$

Comments:
1. Independence implies joint probabilities are simple products
2. It was invented to derive closed-form solutions for joint distributions
3. Some physical variables can be shown to be statistically independent from first principles

Challenge:
Given small sample of two random variables, can we decide if their distributions are independent ?
– if P_X and P_Y are known, it is easier to check independence with some confidence using hypothesis testing

Independence Implies Zero Covariance

X, Y are statistical independent

$$
\begin{aligned}
Cov(X,Y) &= E[(X - E(X))(Y - E(Y)] \\
&= \int_x \int_y (x - E(x))(y - E(y)) f_{X,Y}(x,y) dx dy \\
&= \int_x \int_y (x - E(x))(y - E(y)) f_X(x) f_Y(y) dx dy \\
&= \int_x (x - E(x)) \int_y (y - E(y)) f_Y(y) dy f_X(x) dx \\
&= 0
\end{aligned}
$$

Zero Covariance Does Not Imply Independence

P_X uniform on $[-1/2, 1/2]$
P_Y uniform on $\{-1/2, 1/2\}$
if $X \in [-1/2, 0)$ then $Y = -1/2$
if $X \in [0, 1/2]$ then $Y = 1/2$

X and Y are not statistically independent since
For $X \in [-1/2, 0)$ we have
$P_{Y=-1/2|X} = 1$ and $P_{Y=1/2|X} = 0$
For $X \in [0, 1/2]$ we have
$P_{Y=-1/2|X} = 0$ and $P_{Y=1/2|X} = 1$

For $X \in [-1/2, 0)$ we have

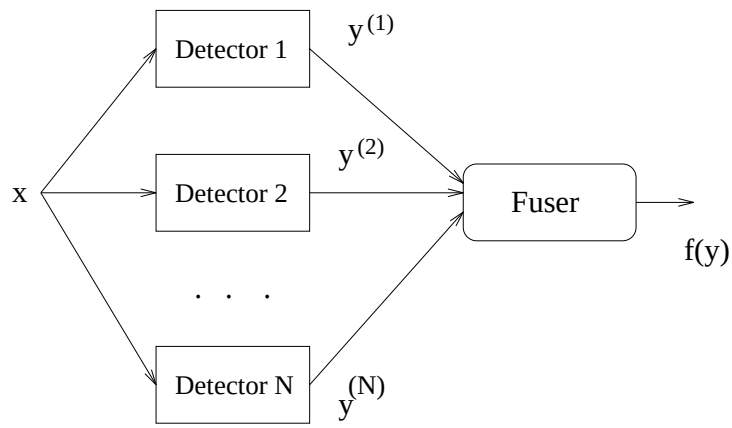
$$P_{Y=-1/2} P_X = 1/2 \neq P_{Y=-1/2|X}$$

Covariance:

$$\begin{aligned} Cov(X,Y) &= E[(X - E(X))(Y - E(Y)] \\ &= \int_x \int_y (x - E(x))(y - E(y)) f_{X,Y}(x,y) dx dy \\ &= 1/2 \int_{x \in [-1/2,0)} x dx - 1/2 \int_{x \in [0,1/2]} x dx \\ &= 0 \end{aligned}$$

Important Point:
One cannot conclude statistical independence based on covariance

Simulation Results: Decision Fusion



Five detectors D_i, $i = 1, 2, \ldots 5$
Input: Boolean with equal probability

Detector D_i:
output is input with prob. $1 - i/10$;
opposite with prob. $i/10$;
Detectors are statistically independent.
$\pi_0 = \pi_1 = 1/2$

Simulation Results: Decision Fusion

Sensor	Probability of Correct Classification
S_1	90.0%
S_2	80.0%
S_3	70.0%
S_4	60.0%
S_5	50.0%

Note:

1. Best classifier:
   correct classification probability - 90%
2. Bayesian fuser:
   empirical performance - 91%

3. Both nearest neighbor and empirical fuser also achieve 91% correct classification.

Decision Fusion (Cntd.)

Percentage of correct classification:

Sample Size	Test set size	Bayesian Fuser	Empirical Decision	Nearest Neighbor	Nadaraya-Watson
100	100	91.91	23.00	82.83	88.00
1000	1000	91.99	82.58	90.39	89.40
10000	10000	91.11	90.15	90.81	91.42
50000	50000	91.19	90.99	91.13	91.14

Bayesian Fuser: Uses probability distribution
(Chair and Varshney 1986)

$$\left.\begin{array}{l}\text{Empirical Decision}\\ \text{Nearest Neighbor}\\ \text{Nadaraya} - \text{Watson}\end{array}\right\}\ \text{Use only the sample}$$

Note: Best performance is achieved by methods that do not explicitly estimate the probabilities.

Decision Fusion (Cntd.)

Percentage of misclassification:

Sample Size	Test set size	S_1	S_2	S_3	S_4	S_5	Nadaraya-Watson
100	100	7.0	20.0	33.0	35.0	55.0	12.0
1000	1000	11.3	18.5	29.8	38.7	51.6	10.6
10000	10000	9.56	20.19	30.38	39.82	49.68	8.58
50000	50000	10.038	20.136	29.854	39.904	50.050	8.860

Note:
The fuser performs better than the best estimator S_1 after 1000 examples.

Conclusions

Summary:

1. Distributed decision fusion problem applicable to cases with unknown error distributions
2. Several fusion rules designed for known distributions case are taken advantage of
3. Finite sample results are obtained (as opposed to asymptotic)

Open Issues:

1. Can we test for statistical independence based on data ?
2. How do we validate probability models ?
3. How do we combine detectors and trackers ?

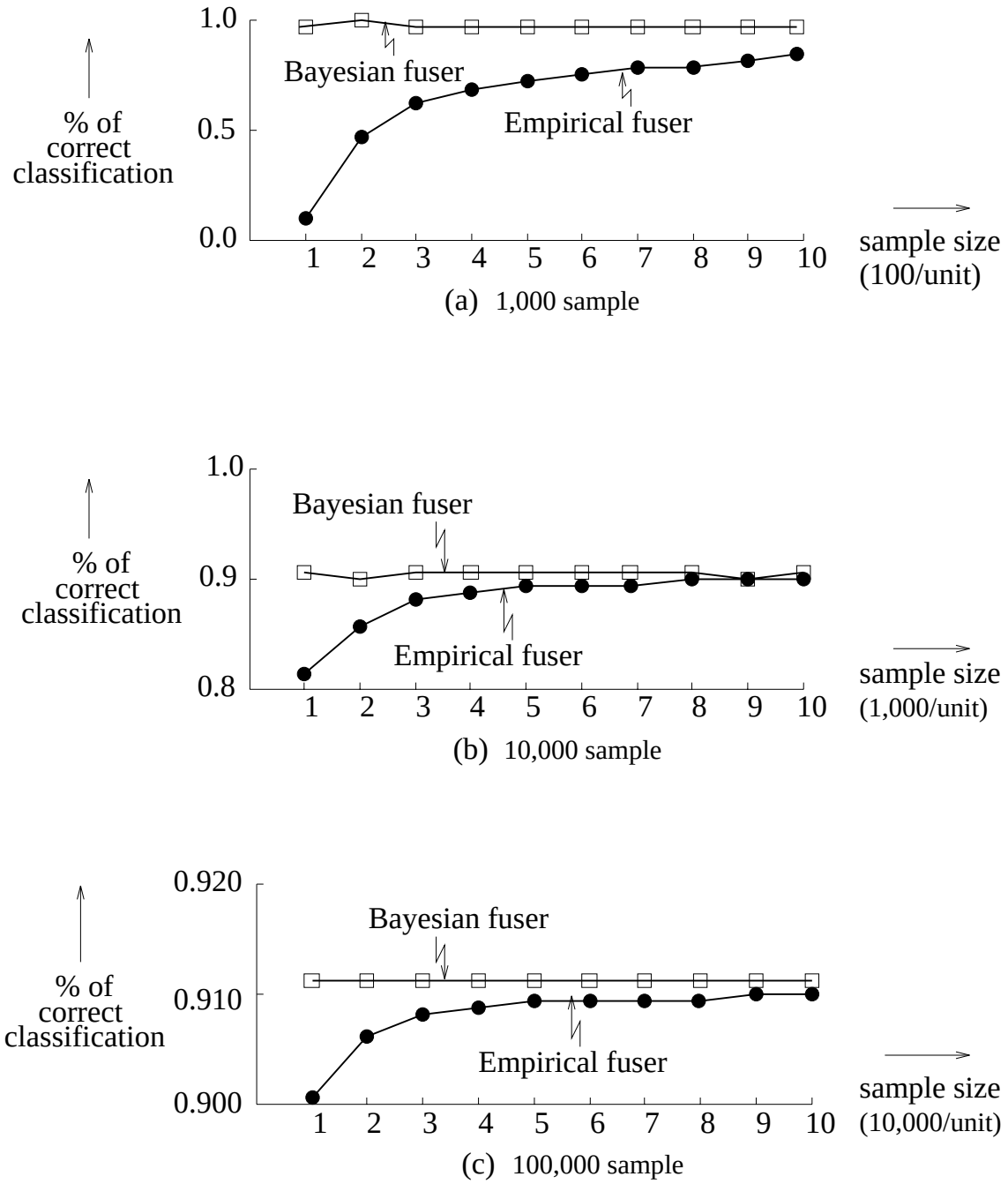


Figure 1: Relative performance of the Bayesian fuser and empirical fuser with training.