On Sample-Based Implementation of Decision Fusion Functions

Nagi Rao
(Nageswara S.V. Rao)
Center for Engineering Science Advanced Research
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6355
email: raons@ornl.gov

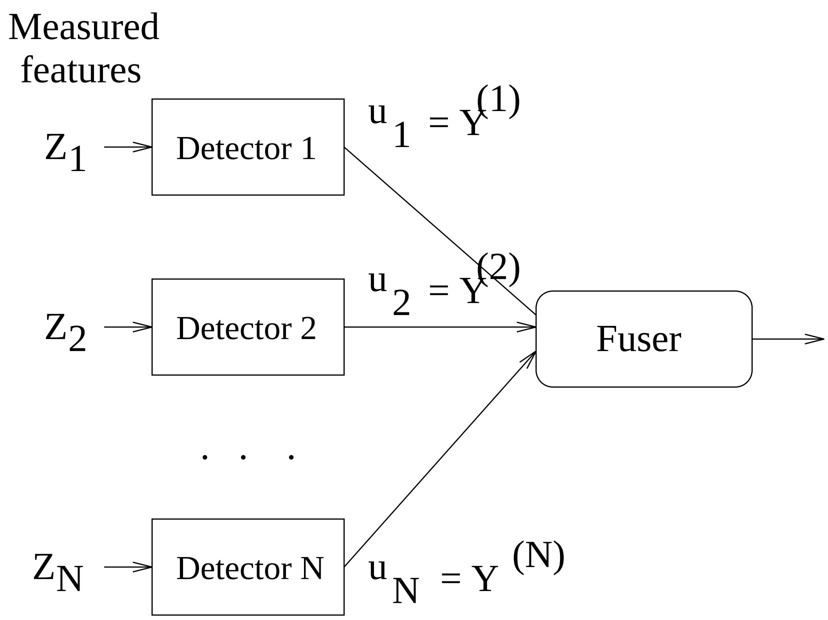
May 15-16, 2001
ONR/GTRI Workshop on
Target Tracking and Senso Fusion
Monterey, CA

This research is sponsored by
Ballistic Missile Defense Organization
Office of Naval Research
U.S. Department of Energy
Office of Science, Engineering Research Program

Presentation Outline

1. **Distributed Detection Problem**
   1.1 Formulation and Examples
   1.2 Motivation
   1.3 Sample-Based Implementation
2. **Smooth Fusion Functions**
   2.1 Sample-Based Solutions
   2.2 Lipschitz Conditions
3. **Non-Smooth Fusion Functions**
   3.1 Examples of Non-Smooth Fusion Rules
   3.2 Sample Size Estimation
4. **Conclusions**

Detection Problem: 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 recognition (Chow 1965)

   Newer applications are being found in diverse areas
2. Of particular importance to distributed sensor systems
   Extensively studied over the past decade
   Dasarathy (1994), Varshney (1996), Rao (1997)

Examples

1. **Intruder Detection System:**
   Detectors monitor workspace from different vantage points
   Each detector is equipped with sensors, algorithms
   H_0: intruder is present; H_1: intruder is not present
   D_i: u_i is generated probabilistically; $P_i(u|H_0)$, $P_i(u|H_1)$
   Question: Can individual results be combined to obtain more reliable decision ?

2. **DNA Analysis System:** (Uberbacher and Mural 1991)
   Each detector is a software program that examines a segment of human DNA sequence
   H_0: Segment is protein-coding region; H_1: otherwise
   Question: Can different programs be combined to obtain improved performance ?

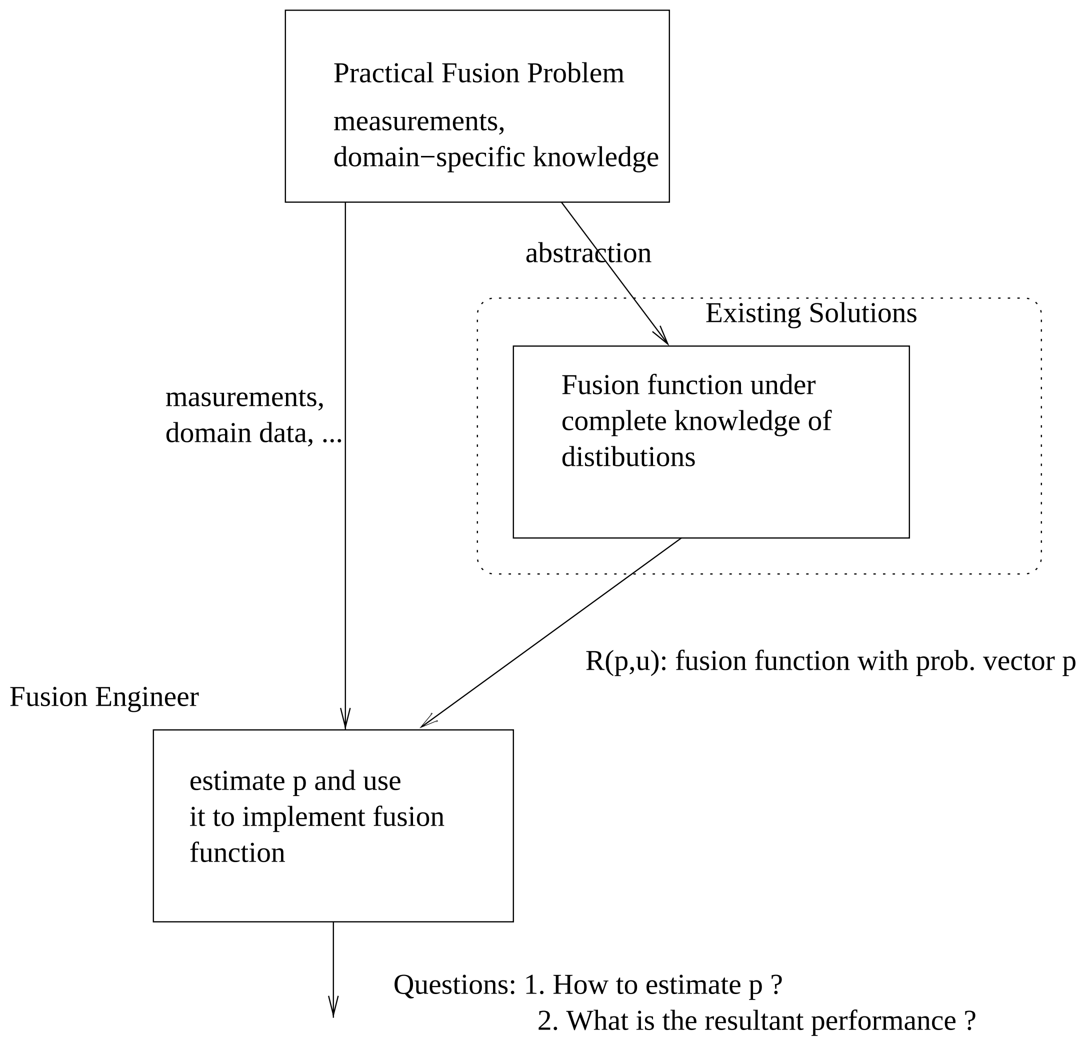
Note:
In the examples, systems are available
— data can be collected.

Example Fusion Rule: (Hashlamoun and Varshney, 1993)
Average-cost criterion is optimized by the likelihood ratio test

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

where
C_{kj}: cost of deciding H_k when H_j is true, $k, j = 0, 1$

A Scenario in Fusion Engineering



Motivation and Summary

Majority of Existing Analytical Results:
(i) Fusion functions are derived under complete knowledge of distributions
(ii) Probabilities are explicitly used – their estimators are used in their place.

In Several Practical Cases:
(i) Detector system is available, hence experimental samples can be collected
(ii) Probability distributions are estimated

Question: How can one use analytical results in practical cases ?

Our Results:
Fusion functions for known distributions can be implemented using sample:
(i) simple averages in place of probabilities work well for smooth laws
but not for non-smooth ones;
(ii) for non-smooth case, estimation is based on cost functional
somewhat weaker performance guarantees.

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.

Substitution formulation:

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 $R(p, u)$?

Performance Criteria I

Cost Functional: 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.

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

Performance Criteria II

Cost Functional: $R(p)$ minimizes certain cost functional C:
$C(R(p, .)) \leq C(f)$ for all

Performance Criterion II:
$R(\hat{p}, u)$ implements $R(p, u)$ with precision ϵ and confidence $1 - \delta$ if

$$P[C[R(\hat{p}, u)] - C[R(\hat{p}, u)] > \epsilon] < \delta$$

for sufficiently (but finite) large sample of size $l < \infty$.

Informally,
based on a sufficiently large sample, cost of $R(\hat{p}, u)$ is within ϵ of optimal with probability $1 - \delta$.

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):

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

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

Result:
For any $r > 2$, consider a sample of size

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

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

Result 1:
Consider a sample of size

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

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

$$1 - 2n\delta$$

or

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

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 1:

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

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

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

$$1 - n\delta$$

or

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

Proof Outline: Theorem 1

1. Conditions $\sup_i |p_i - \bar{p}_i| < \epsilon$ and $|R(p)| \geq L\sqrt{n}\epsilon$, ensure that $R(p,u) > 0$ and $R(\bar{p},u) > 0$ yield the same result.

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

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

   $R(\bar{p},u)$ implements $R(p,u)$ with the required precision.

3. Lower bound for ϵ by noting that $2|R(p,u)| \geq |R(\bar{p},u)|$ which implies $|R(p,u)| \geq \frac{1}{r}|R(\bar{p},u)|$ for any $r \geq 2$.

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

Note:
Tighter sample bounds may be possible in particular cases,
e.g. Theorem 1

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 $\bar{P}(A)$ be fraction of times event A took place in the sample.

(i) Given a training sample of size

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

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

(ii) Given a training sample of size

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

the empirical test $\bar{P}(u \cap H_1)\bar{P}(H_0) - \tau \bar{P}(u \cap H_0)\bar{P}(H_1) > 0$ 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.

Result:
Given a training sample of size

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

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

Limitation of Means-Based Methods

Existing analytical smaple-based results use smoothness properties

Simple Monitoring Example: $R(p, u)$ does not have Lipschitz property
Monitoring area: two non-overlapping regions
– two detectors, one for each region
detection if computed probability is above a threshold
Appropriate fusion rule: $R(p_1, p_2) = (p_1 > t) \vee (p_2 > t)$
– not Lipschitz in p, since it is discontinuous at (t, t).

Explanation:
1. Proximity of mean $\bar{p}$ to p is by CLT and not by $R(\bar{(p)})$ to $R(p)$.
2. If proximity of $\hat{p}$ to p is enforced by R:
performance of $R(\hat{p}, .)$ approaches that of $R(p, .)$.

Example: $R(.)$ is 1 in $[p - \alpha, p + \alpha]$ and zero elsewhere
As $\alpha \to 0$, $R(\bar{p}, u) \neq R(p, u)$ and this method yields very high error.

Motivation: For non-smooth fusion functions
we consider class of functions with bounded variation

Bounded Variation Property

One-dimensional function $h : [-A, A] \mapsto \Re$.
Partition of $[-A, A]$: Set of points $P = \{x_0, x_1, \ldots, x_n\}$ such that

$$-A = x_0 < x_1 < \ldots < x_n = A$$

$\mathcal{P}[-A, A]$: all possible partitions

Bounded-Variation: $g : [-A, A] \mapsto \Re$ is of bounded variation:
if there exists M such that for any $P = \{x_0, x_1, \ldots, x_n\}$, we ha ve

$$\sum(P) = \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})| \leq M.$$

Multiple Dimensions:
A function $g : [-A, A]^d \mapsto \Re$ is of bounded variation if it is so in each of its input variable for every value of the other input variables.

Useful facts:
(i) not all continuous functions are of bounded variation,
e.g. $g(x) = x \cos(\pi/(2x))$ for $x \neq 0$ and $g(0) = 0$;
(ii) differentiable functions on compact domains are of bounded variation;
(iii) absolutely continuous functions, which include Lipschitz functions, are of bounded variation.

Non-Smooth Fusion Functions

Compute $\hat{p}$ such that

$$C(R(\hat{p})) = \min_{p \in [0,1]^n} \frac{1}{l} C(R(p, u^i))$$

based on the sample. Then $R(\hat{p}, u)$ is used in place of $R(p, u)$.

Theorem 2:
$R(p, u)$ is of bounded variation with respect to p
$C \leq M$ is of bounded bounded variation.
Given a sample of size

$$s = \frac{256M^2}{\epsilon^2} \left[4n \ln \left(\frac{8eM}{\epsilon} \right) + \ln(16/\delta) \right],$$

we have

$$\mathbf{P}[C(R(\hat{p})) - C(R(p)) > \epsilon] < \delta.$$

Proof of Theorem2: Preliminaries

For $\mathcal{G} = \{g : X \mapsto \Re\}$ and $S = \{x_1, x_2, \ldots, x_m\} \subseteq X$:

Pseudo-shattering: S is pseudo-shattered by F
if there are real numbers r_1, r_2, $\ldots$, r_m
such that for each $b \in \{0,1\}^m$ there is a function g_0 in $\mathcal{G}$ with

$$\operatorname{sgn}(f_b(x_i) - r_i) = b_i$$

for $1 \leq i \leq m$.

Pseudo-Dimension: $\mathcal{G}$ has pseudo-dimension d: maximum cardinality of a subset S of X that is pseudo-shattered by $\mathcal{G}$

Examples: Function classes with known pseudo-dimension
1. Feedforward sigmoidal neural networks
2. Vector spaces

Proof of Theorem2: Outline

Denote $C(R(p,.))$ by $C(R(p,.))$
$CR(p)$ is of bounded variation with respect to p. Consider the function class

$$\mathcal{CR} = \{CR(q,.) : q \in [0,1]^n\}.$$

Let $\widehat{CR}(q) = \frac{1}{l}C(R(p,u^i)$.

From Vapnik (1982), we have

$$\begin{aligned}&\mathbf{P}\left[CR(\hat{p}) - CR(p) > \epsilon\right] \\ &\quad \leq \mathbf{P}\left[\sup_{q\in[0,1]^n} |CR(q) - \widehat{CR}(q)| > \epsilon/2\right].\end{aligned}$$

From Haussler (1992), we obtain

$$\begin{aligned}&\mathbf{P}\left[CR(\hat{p}) - CR(p) > \epsilon\right] \\ &\quad \leq 2E\left[2\min(\mathcal{N}(\epsilon/2, \mathcal{CR}, d_{L^1}))\right] e^{\frac{-\epsilon^2 l}{256M^2}}.\end{aligned}$$

We next show that

$$\mathcal{N}(\epsilon, \mathcal{CR}, d_{L^1(P)}) \leq 4\left(\frac{4eM}{\epsilon}\ln\frac{4eM}{\epsilon}\right)^{2n},$$

for any P, which yields the required sample size.

Proof of Theorem2: Outline

$CR(.) = C(R(.))$ is of bounded variance:
represented as a sum of two monotone functions $CR = R_1 + R_2$.
For $i = 1, 2$, let $\mathcal{R}_i = \{R_i(q, .) : q \in [0, 1]^n\}$,
functions composed by a monotone function $R_i(.)$ with identity $I(.)$

q forms a linear space: by Anthony and Bartlett (1999)

$$\mathrm{Pdim}(\mathcal{R}_i) = \mathrm{Pdim}(\{q\}) \leq \mathrm{Pdim}([0, 1]^n) = n.$$

From Haussler (1992)

$$\mathcal{N}\left(\epsilon, \mathcal{R}_i, d_{L^1(P)}\right) \leq 2\left(\frac{2eM}{\epsilon}\ln\frac{2eM}{\epsilon}\right)^n$$

for any measure P.

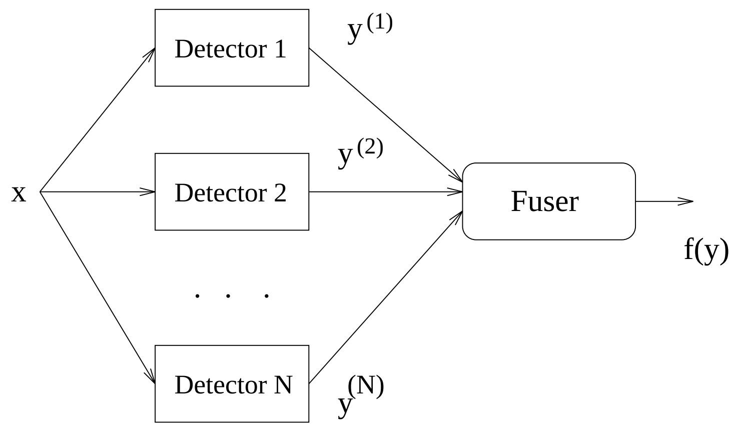
Since $CR = R_1 + R_2$ we obtain

$$\begin{aligned}
&\mathcal{N}\left(\epsilon, \mathcal{CR}, d_{L^1(P)}\right) \\
&\quad \leq \mathcal{N}(\epsilon/2, \mathcal{R}_1, d_{L^1(P)})\mathcal{N}(\epsilon/2, \mathcal{R}_2, d_{L^1(P)}) \\
&\quad \leq 4\left(\frac{4eM}{\epsilon}\ln\frac{4eM}{\epsilon}\right)^{2n}
\end{aligned}$$

for any P.
The sample bound follows from

$$\begin{aligned}
\delta &= 2E\left[2\min(\mathcal{N}(\epsilon/2, \mathcal{CR}, d_{L^1}))\right] e^{\frac{-\epsilon^2 l}{256M^2}} \\
&\leq 16\left(\frac{8eM}{\epsilon}\ln\frac{8eM}{\epsilon}\right)^{2n} e^{\frac{-\epsilon^2 l}{256M^2}}
\end{aligned}$$

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)

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

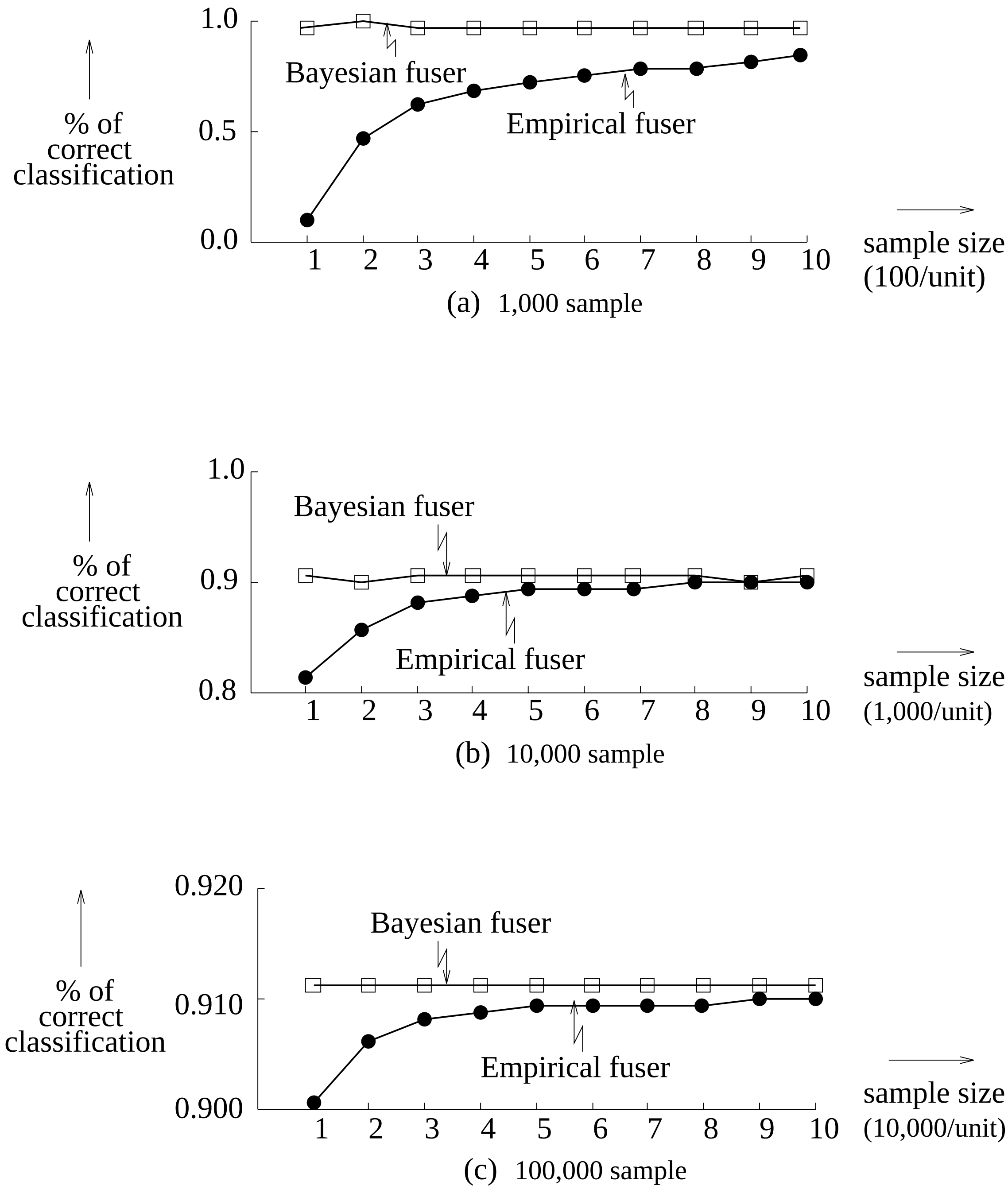


Figure 1: Relative performance of the Bayesian fuser and empirical fuser with training.

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.

<u>Conclusions</u>

Summary:
For a very general class of fusion rules,
existing ones can be converted into sample-based ones.

Next Step: Sample size can be sharpened by using domain specific data.

Future Directions:

1. Empirical estimation methods represent other extreme:
   – operate directly on the data (no estimation of prob.)
   What are boundaries of performance between two methods ?

2. Combination of detection and tracking in a close-loop