Optimal Bayesian Fusion of Multiple Unreliable Classifiers

Jean Dezert
ONERA/DTIM, 29 Av. Division Leclerc
92320 Châtillon, France
Jean.Dezert@onera.fr

Abstract – *Most of modern multitarget target tracking and recognition systems integrate different kind of sensors (imaging, optical, radar, IR , etc). The major problem in such new systems is to find how to fuse optimally measurements, decisions/classifications or estimates provided by the different sources of information involved in the global system. One of the main difficulty is to take into account in the fusion process the reliability of each source of information. We present in this paper the optimal bayesian fusion rule (OBFR) for the case of unreliable multi-classifier problem. Validation of OBFR through Monte Carlo simulations is presented.*

Keywords: Bayesian fusion, bayesian theory, target classification, multisensor system, system reliability.

1 Introduction

We consider a system based on N different unreliable sources of information (sensors, human experts, AI analyzers or whatever). Each source of information (i.e. a sensor coupled with its own processing unit) $s_n, n = 1 \ldots, N$ provides a decision A_n on the true nature w of the target T under consideration with given reliability weights $r_n = (1-\alpha_n, 1-\beta_n) \in [0;1]^2$. α_n and β_n correspond to Type I and Type II errors often referred to false alarm and miss probabilities in engineering [9] (see discussion in the sequel). w belongs to a given finite set $W = \{w_1, w_2, \ldots, w_M\}$ called the world (or frame) of discernement of the problem.

In our Bayesian framework, we assume that prior probabilities $p_i \triangleq P\{w = w_i\}, i = 1, \ldots, M$ are known with $\sum_{i=1}^M p_i = 1$ if we consider a close-world W or $\sum_{i=1}^M p_i < 1$ if we consider an open-world. In the open-world case, the list of w_i is not exhaustive. Since we can always introduce the complement hypothesis $w_0 =$ "Not a w_i target" with probability $p_0 = 1 - \sum_{i=1}^M p_i$, the initial open-world W can always be replaced by the new close-world $W_0 = \{w_0, w_1, w_2, \ldots, w_M\}$. Hence if we need to deal with an open-world, we will just have to deal with $M+1$ hypotheses rather than M in OBFR formulaes developped in the sequel. A simple classification system could be

$$\begin{aligned} W = \{ & w_1 = \text{“Fighter”}, \\ & w_2 = \text{“Small civilian jet”}, \\ & w_3 = \text{“Civilian air carrier”}, \\ & w_4 = \text{“Bomber”}, \\ & w_5 = \text{“Air-to-air missile”}, \\ & w_6 = \text{“Helicopter”}\} \end{aligned}$$

Each source s_n, $(n = 1, \ldots, N)$ asserts either $A_n \triangleq$ "$w \in W_n$" $\equiv$ "$w \notin W_n^c$" or its negation $\neg A_n \triangleq$ "$w \notin W_n$" $\equiv$ "$w \in W_n^c$" about the nature of the target. W_n and W_n^c are *disjoint subsets* of W with $W_n \cup W_n^c = W$. The assertion A_n or $\neg A_n$ can change with time k because of the dynamic of target, environmental conditions, etc. For notation convenience, time index k will be omitted in the sequel.

This modelling is more general than the classical one which usually provides only a decision on a focal hypothesis w_i at a time. In the classical multi-sensor detection problem, we look for the best decision between hypothesis $H_0 =$ "no target" and $H_1 =$ "presence of a target" [3, 16, 19]. Usually in practice, each classifier is only able to discriminate between several subsets of W rather than all elements w_i separately. Assertions A_n or $\neg A_n$ are in general not sure and we have to deal with classification errors of type I or II which are characterized by the *false alarm probability* (i.e. the probability that source s_n asserts A_n while A_n is physically not valid (false) - we will write $A_n = \bar{V}$)

$$\alpha_n \triangleq P\{\text{accept} A_n | A_n = \bar{V}\}$$

and the miss probability (the probability that source s_n asserts $\neg A_n$ while A_n is physically valid (true); we will then write $A_n = V$):

$$\beta_n = P\{\text{accept}\neg A_n | A_n = V\}$$

α_n and β_n are assumed to be known and $r_n \triangleq (1-\alpha_n, 1-\beta_n)$ is the reliability of s_n.

It must be noted that OBFR developped in the following requires the full knowledge of prior probabilities p_i and r_n. Different approaches based on non-bayesian frameworks (like evidence theory [1, 2, 6], fuzzy sets [5, 11], possibility theory, etc) could also be used when the full knowledge of prior is missing. For notation convenience, we will note

$$W_n^{\delta_n=1} \equiv A_n \quad \text{and} \quad W_n^{\delta_n=0} \equiv \neg A_n$$

which allows to identify directly W_n^1 for assertion A_n with subset W_n and W_n^0 for assertion $\neg A_n$ with subset W_n^c.

The Optimal Bayesian Fusion (OBF) problem consists to compute, for $i = 1, \dots, M$, the fused conditional probabilities $P\{w = w_i | W_1^{\delta_1}, \dots, W_N^{\delta_N}\}$ from local unreliable sources $s_1, \dots, s_N$ with reliability $r_1, \dots, r_N$. More precisely, the derivation must be done from local conditional probabilities $P\{w = w_i | W_1^{\delta_1}\}$, ..., $P\{w = w_i | W_N^{\delta_N}\}$. The purpose of this work is then to find the general relationship of the kind

$$P\{w = w_i | W_1^{\delta_1}, \dots, W_N^{\delta_N}\} = \mathcal{F}[P\{w = w_i | W_1^{\delta_1}\}, \dots, P\{w = w_i | w_N^{\delta_N}\}, r_1, \dots, r_N, W, p_1, \dots, p_M]$$

where $\mathcal{F}[., \dots, .]$ is the OBFR we are searching for. This can be summarized by the block-scheme on figure 1

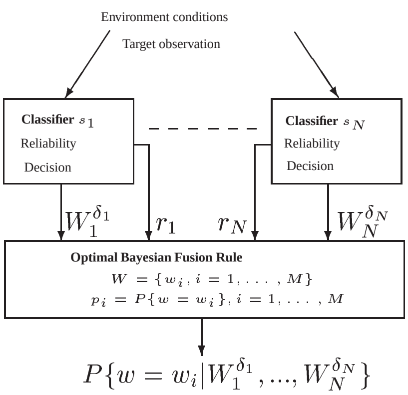


Figure 1: Optimal Bayesian Fusion

Condition of existence of OBFR: The OBFR always exists if the sources are unreliable even if they appear to be incompatible, i.e. $W_1^{\delta_1} \cap, \dots, \cap W_N^{\delta_N} = \emptyset$ when $\delta_1 = \dots = \delta_N$. When the sources are fully reliable ($r_1 = \dots = r_N = (1, 1)$) and if for $\delta_1 = \dots = \delta_N$, one has $W_1^{\delta_1} \cap W_2^{\delta_2}, \dots, \cap W_N^{\delta_N} = \emptyset$ then no theoretical optimal fusion rule exists. Only some heuristic fusion rules can be developped eventually.

2 Simpliest case : OBFR(2)

Consider now the simpliest case for 2 classifiers, denoted by OBFR(2). The condition of existence of OBFR is assumed to be met and we suppose that W has M elements w_i. $I = \{1, 2, \dots, m\}$ is the set of hypothesis indices in W. Source s_1 can discriminate between $W_1^1 \subset W$ and its complement $W_1^0 \triangleq W_1^c$ and s_2 between $W_2^1 \subset W$ and its complement $W_2^0 \triangleq W_2^c$. I_1^1 and I_2^1 are the sets of hypothesis indices involved in W_1^1 and W_2^1. Complement of I_1^1 and I_2^1 in I are denoted I_1^0 and I_2^0. Cardinalities of $W_1^{\delta_1}$ and $W_2^{\delta_2}$ (for $\delta_1, \delta_2 \in \{0, 1\}$) can be any integer between 1 and M with the constraint Card(W_n^1) + Card$(W_n^0) = M$ for $n = 1, 2$. Reliability parameters for s_1 and s_2 are respectively $r_1 = (1 - \alpha_1, 1 - \beta_1)$ and $r_2 = (1 - \alpha_2, 1 - \beta_2)$.

2.1 Derivation of $P\{W_n^{\delta_n}\}$

The prior marginal probabilities of classifier assertion $P\{W_1^{\delta_1}\}$ and $P\{W_2^{\delta_2}\}$ taking into account reliability parameters are obtained by elementary probability calculus [9]. We get for $n = 1, 2$ and for $\delta_n \in \{0, 1\}$,

$$P\{W_n^{\delta_n}\} = \sum_{a_n \in \{0,1\}} \big(f_n[\Delta_n, \delta_n] \sum_{i \in I_n^{a_n}} p_i \big) \tag{1}$$

where Δ_n is the Kronecker indicator function defined here as $\Delta_n = 1$ if $a_n = \delta_n$ or 0 otherwise. The function $f_n[\Delta_n, \delta_n]$ is defined as

$$f_n[\Delta_n, \delta_n] = (1 - \alpha_n)^{\Delta_n(1-\delta_n)} (1 - \beta_n)^{\Delta_n \delta_n} \cdot {\beta_n}^{(1-\Delta_n)(1-\delta_n)} {\alpha_n}^{(1-\Delta_n)\delta_n} \tag{2}$$

which corresponds to

$$f_n[\Delta_n, \delta_n] = \begin{cases} 1 - \alpha_n & \text{if} \quad \Delta_n = 0 \text{ and } \delta_n = 0 \\ 1 - \beta_n & \text{if} \quad \Delta_n = 0 \text{ and } \delta_n = 1 \\ \beta_n & \text{if} \quad \Delta_n = 1 \text{ and } \delta_n = 0 \\ \alpha_n & \text{if} \quad \Delta_n = 1 \text{ and } \delta_n = 1 \end{cases}$$

When $\alpha_n = \beta_n$, $f_n[\Delta_n, \delta_n]$ depends only on Δ_n and if $\alpha_n = \beta_n = 0.5$ then $f_n[\Delta_n, \delta_n] = 0.5$. It can be easily checked that

$$\sum_{\delta_n \in \{0,1\}} P\{W_n^{\delta_n}\} = 1$$

2.2 Derivation of $P\{W_n^{\delta_n} | w = w_i\}$

The derivation of likelihood $P\{W_n^{\delta_n} | w = w_i\}$ is computed using the total probability theorem by introducing the truth or falsety of assertion as follow

$$\begin{aligned} P\{W_n^{\delta_n} | w = w_i\} = & \\ & P\{W_n^{\delta_n} | w = w_i, W_n^{\delta_n} = V\} P\{W_n^{\delta_n} = V | w = w_i\} \\ & + P\{W_n^{\delta_n} | w = w_i, W_n^{\delta_n} = \bar{V}\} P\{W_n^{\delta_n} = \bar{V} | w = w_i\} \end{aligned} \tag{3}$$

$P\{W_n^{\delta_n} = \mathcal{A} | w = w_i\}$ when $\mathcal{A}_n = V$ (truth) or $\mathcal{A}_n = \bar{V}$ (falsety) is given by

$$P\{W_n^{\delta_n} = \mathcal{A}_n | w = w_i\} = (1 - \Delta_n[i])^{1-v_n} \Delta_n[i]^{v_n} \tag{4}$$

which is either equal to 0 or 1 and where $v_n = 1$ if $\mathcal{A}_n = V$ or 0 otherwise. $\Delta_n[i]$ is a new indicator function introduced here for notation convenience defined by $\Delta_n[i] = 1$ if $i \in I_n^{\delta_n}$ or 0 otherwise.

$P\{W_n^{\delta_n} | w = w_i, W_n^{\delta_n} = V\}$ is given by

$$\begin{cases} 0 & \text{if } i \notin I_n^{\delta_n} \\ 1 - \beta_n & \text{if } i \in I_n^{\delta_n} \text{ and } \delta_n = 1 \\ 1 - \alpha_n & \text{if } i \in I_n^{\delta_n} \text{ and } \delta_n = 0 \end{cases}$$

$P\{W_n^{\delta_n}|w=w_i, W_n^{\delta_n}=\bar{V}\}$ is given by

$$\begin{cases} 0 & \text{if } i \in I_n^{\delta_n} \\ \alpha_n & \text{if } i \notin I_n^{\delta_n} \text{ and } \delta_n = 1 \\ \beta_n & \text{if } i \notin I_n^{\delta_n} \text{ and } \delta_n = 0 \end{cases}$$

Note : When $i \notin I_n^{\delta_n}$, the conditioning event "$w = w_i$" $\cap$ "$W_n^{\delta_n} = V$" can never occur (impossible event). The event "$w = w_i$" $\cap$ "$W_n^{\delta_n} = \bar{V}$" when $i \in I_n^{\delta_n}$ is also an impossible event. Therefore, theoretically, $P\{W_n^{\delta_n}|w = w_i, W_n^{\delta_n} = V\}$ when $i \notin I_n^{\delta_n}$ and $P\{W_n^{\delta_n}|w = w_i, W_n^{\delta_n} = \bar{V}\}$ when $i \in I_n^{\delta_n}$ are not defined. These "probabilities" can however be set to any arbitrary finite value without affecting the global result since they are multiplied by a zero factor in (3). This justifies our choice to set $P\{W_n^{\delta_n}|w = w_i, W_n^{\delta_n} = V\}$ and $P\{W_n^{\delta_n}|w = w_i, W_n^{\delta_n} = \bar{V}\}$ to zero when $i \notin I_n^{\delta_n}$ and $i \in I_n^{\delta_n}$ respectively.

Using previous expressions in (3), we get for $P\{W_n^{\delta_n}|w = w_i\}$

$$\begin{cases} (1-\alpha_n)^{1-\delta_n}(1-\beta_n)^{\delta_n} & \text{if } i \in I_n^{\delta_n} \\ \alpha_n^{\delta_n}\beta_n^{1-\delta_n} & \text{if } i \notin I_n^{\delta_n} \end{cases}$$

which can be rewritten as

$$P\{W_n^{\delta_n}|w = w_i\} = f_n[\Delta_n[i], \delta_n] \tag{5}$$

with $f_n[.,.]$ defined in (2).

2.3 Derivation of $P\{w = w_i|W_n^{\delta_n}\}$

The conditional probability of a focal hypothesis "$w = w_i$" given the assertion $W_n^{\delta_n}$ provided by s_n for $n = 1, 2$ follows from Bayes' rule. One gets

$$P\{w = w_i|W_n^{\delta_n}\} = \frac{P\{W_n^{\delta_n}|w = w_i\}p_i}{P\{W_n^{\delta_n}\}} \tag{6}$$

where $p_i \triangleq P\{w = w_i\}$ are assumed to be known; $P\{W_n^{\delta_n}|w = w_i\}$ is given by (5) and $P\{W_n^{\delta_n}\}$ is given by (1). The normalization constant $P\{W_n^{\delta_n}\}$ can be computed by (1) or by

$$P\{W_n^{\delta_n}\} = \sum_i P\{W_n^{\delta_n}|w = w_i\}p_i \tag{7}$$

and the following equality always holds for any $W_n^{\delta_n}$

$$\sum_i f_n[\Delta_n[i], \delta_n]p_i = \sum_{a_n \in \{0,1\}} \big(f_n[\Delta_n, \delta_n] \sum_{i \in I_n^{a_n}} p_i\big)$$

2.4 Derivation of $P\{W_1^{\delta_1}, W_2^{\delta_2}\}$

The prior probability of joint assertions $P\{W_1^{\delta_1}, W_2^{\delta_2}\}$ is given by

$$\sum_{(a_1,a_2)\in\{0,1\}^2} \big(f_1[\Delta_1, \delta_1]f_2[\Delta_2, \delta_2] \sum_{i \in I_1^{a_1} \cap I_2^{a_2}} p_i\big) \tag{8}$$

where $\{0,1\}^2$ represents the set of all couples (i, j) taking their values in $\{0,1\} \times \{0,1\}$. One has also

$$\sum_{(\delta_1,\delta_2)\in\{0,1\}^2} P\{W_1^{\delta_1}, W_2^{\delta_2}\} = 1$$

2.5 Derivation of $P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\}$

By introducing the truth or falsety of joint assertion under consideration and using the total probability theorem, the likelihood $P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\}$ is the given by

$$P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\} = \sum_{(\mathcal{A}_1,\mathcal{A}_2)\in\{V,\bar{V}\}^2} \big(P\{W_1^{\delta_1}, W_2^{\delta_2}|w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\} \cdot P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w_i\}\big) \tag{9}$$

where $(\mathcal{A}_1, \mathcal{A}_2) \in \{V, \bar{V}\}^2$ represents all combinations of truth and falsety for possible joint assertions $(W_1^{\delta_1}, W_2^{\delta_2})$. We must note that there exists *only one non null term* $P\{W_1^{\delta_1}, W_2^{\delta_2}|w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\}P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w_i\}$ in the previous summation. The choice of this non null term depends on the membership of w_i with the intersection of $W_1^{\delta_1} = \mathcal{A}_1$ with $W_2^{\delta_2} = \mathcal{A}_2$ for a particular $(\mathcal{A}_1, \mathcal{A}_2) \in \{V, \bar{V}\}^2$. More precisely, the joint conditional probabilities $P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w = w_i\}$ are theoretically given by

$$\begin{cases} 1 & \text{if } i \in \mathcal{I}_1^{\delta_1,v_1} \cap \mathcal{I}_2^{\delta_2,v_2} \\ 0 & \text{if } i \notin \mathcal{I}_1^{\delta_1,v_1} \cap \mathcal{I}_2^{\delta_2,v_2} \end{cases}$$

where the new set $\mathcal{I}_n^{\delta_n,v_n}$ is defined by

$$\mathcal{I}_n^{\delta_n,v_n} \triangleq v_n I_n^{\delta_n} \cup (1 - v_n)I_n^{1-\delta_n} \tag{10}$$

which is only a concise form to indicate that $\mathcal{I}_n^{\delta_n,v_n} = I_n^{\delta_n}$ when $v_n = 1$ (i.e. $\mathcal{A}_n = V$) or $I_n^{1-\delta_n}$ when $v_n = 0$.

By scanning all combinations of binary values for δ_n and v_n (for $n = 1, 2$), it can be shown that $P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w_i\}$ can be rewritten as

$$\prod_{n=1,2} (1 - \Delta_n[i])^{1-v_n} \Delta_n[i]^{v_n}$$

Taking into account (4), we finally get for $P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w_i\}$

$$\prod_{n=1,2} P\{W_n^{\delta_n} = \mathcal{A}_n|w = w_i\} \tag{11}$$

Two cases must now be considered for the derivation of $P\{W_1^{\delta_1}, W_2^{\delta_2}|w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\}$

- **Case 1**: $i \notin \mathcal{I}_1^{\delta_1,v_1} \cap \mathcal{I}_2^{\delta_2,v_2}$

In this case, $P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w = w_i\} = 0$. This corresponds to the impossible event

$$\text{“}w = w_i\text{”} \cap \text{“}W_1^{\delta_1} = \mathcal{A}_1\text{”} \cap \text{“}W_2^{\delta_1} = \mathcal{A}_2\text{”}$$

$P\{W_1^{\delta_1}, W_2^{\delta_2}|w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\}$ is not defined in theory but can be set to any finite value since it is multiplied by a zero factor in derivations. We set it to zero.

• **Case 2**: $i \in \mathcal{I}_1^{\delta_1,v_1} \cap \mathcal{I}_2^{\delta_2,v_2}$

In this case, $P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w_i\} \equiv 1$. Using Bayes' rule, $P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\}$ is given by

$$\begin{aligned} &P\{W_1^{\delta_1}|w = w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2, W_2^{\delta_2}\} \\ &\cdot P\{W_2^{\delta_2}|w = w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\} \end{aligned} \quad (12)$$

If there is *no feedback (communications)* between sources s_n, the irrelevant conditioning terms can be removed in previous formulaes and one gets

$$\begin{aligned} &P\{W_1^{\delta_1}, W_2^{\delta_2}|w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\} = \\ &\prod_{n=1,2} P\{W_n^{\delta_n}|w_i, W_n^{\delta_n} = \mathcal{A}_n\} \end{aligned} \quad (13)$$

The non null term $P\{W_1^{\delta_1}, W_2^{\delta_2}|w_i, W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2\}P\{W_1^{\delta_1} = \mathcal{A}_1, W_2^{\delta_2} = \mathcal{A}_2|w_i\}$ entering in (9) can be expressed as,

$$\prod_{n=1,2} P\{W_n^{\delta_n}|w_i, W_n^{\delta_n} = \mathcal{A}_n\}P\{W_n^{\delta_n} = \mathcal{A}_n|w_i\}$$

which finally implies, because of (3),

$$P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\} = \prod_{n=1,2} P\{W_n^{\delta_n}|w = w_i\}$$

or equivalently,

$$P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\} = \prod_{n=1,2} f_n[\Delta_n[i], \delta_n] \quad (14)$$

2.6 Derivation of $P\{w = w_i|W_1^{\delta_1}, W_2^{\delta_2}\}$

The conditional probabilities of focal hypothesis "$w = w_i$" given the joint unreliable assertions $W_1^{\delta_1}$ and $W_2^{\delta_2}$ are obtained by the Bayes' rule as follows.

$$P\{w = w_i|W_1^{\delta_1}, W_2^{\delta_2}\} = \frac{P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\}p_i}{P\{W_1^{\delta_1}, W_2^{\delta_2}\}}$$

$P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\}$ is given by (14); p_i is known and the normalization constant is given by (8) or equivalently by

$$P\{W_1^{\delta_1}, W_2^{\delta_2}\} = \sum_i P\{W_1^{\delta_1}, W_2^{\delta_2}|w = w_i\}p_i \quad (15)$$

An other useful expression (for practical implementation) of $P\{w = w_i|W_1^{\delta_1}, W_2^{\delta_2}\}$ is

$$\frac{p_i \prod_{n=1,2} f_n[\Delta_n[i], \delta_n]}{\sum_{i=1,M} p_i \prod_{n=1,2} f_n[\Delta_n[i], \delta_n]} \quad (16)$$

2.7 Final expression of OBFR(2)

The OBFR(2) rule consists to express $P\{w = w_i|W_1^{\delta_1}, W_2^{\delta_2}\}$ as a function $\mathcal{F}[.]$ of $P\{w = w_i|W_n^{\delta_n}\}$, r_n and p_i. Using algebraic manipulations on previous formulaes, one gets the final general OBFR(2) result

$$P\{w = w_i|W_1^{\delta_1}, W_2^{\delta_2}\} = \frac{p_i^{-1} \prod_{n=1,2} P\{w_i|W_n^{\delta_n}\}}{\sum_{i=1,M} p_i^{-1} \prod_{n=1,2} P\{w_i|W_n^{\delta_n}\}} \quad (17)$$

If we assume *uniform* prior $p_i = 1/M$, then terms p_i^{-1} can be removed in the OBFR(2) formula above. The validation of theoretical OBFR(2) via Monte-Carlo simulation is presented in next section.

2.8 Simulation results of OBFR(2)

Comparison of Monte Carlo simulations results with theoretical results will now be presented for cases of a bi-classifier system. Each independent classifier can have different or same performance as the other one. Both cases are presented.

2.8.1 OBFR(2) with 2 different classifiers

We consider here the general case with two different classifiers s_1 and s_2 having different reliability factors r_1 and r_2. The frame of discernement W under consideration is supposed to have only 9 focal hypotheses, i.e. $W = \{w_1, \ldots, w_9\}$. s_1 is able to discriminate between subset B and its complementary $\bar{B}$ with reliability $r_1 = (0.60, 0.75)$ and s_2 between subset C and its complementary $\bar{C}$ with reliability $r_2 = (0.90, 0.80)$ where

$$B = \{w_2, w_4, w_5\} \quad \text{and} \quad C = \{w_1, w_2, w_3, w_4\}$$

The reliability factors imply $\alpha_1 = P\{B|B = \bar{V}\} = 0.40$, $\beta_1 = P\{\bar{B}|B = V\} = 0.25$, $\alpha_2 = P\{C|C = \bar{V}\} = 0.10$ and $\beta_2 = P\{\bar{C}|C = V\} = 0.20$.

• **OBFR(2) with non uniform prior**

Consider the following *non uniform prior probabilities*

$$p_1 = 0.05, p_2 = 0.30, p_3 = 0.05, p_4 = 0.40, p_5 = 0.05,$$
$$p_6 = 0.05, p_7 = 0.02, p_8 = 0.05, p_9 = 0.03$$

Our simulation are based on samples drawing and occurences counting to estimate the experimental probabilities in one side and the implementation of theoretical relationships previously presented in othre side. Monte Carlo simulations are based on 50000 samples.

Figure 2 shows Monte-Carlo and theoretical results for estimation of fused conditional probabilities of focal hypothesis "$w = w_i$".

Figure 3 presents a comparison between marginal conditional probabilities $P\{w = w_i|W_i{}^{\delta_i}\}$ (for $\delta_i \in \{0, 1\}$ and $i = 1, 2$) and the joint conditional probabilities $P\{w = w_i|W_1{}^{\delta_1}, W_2{}^{\delta_2}\}$. As we observe, one gets a very good agreement between Monte Carlo results and theoretical ones. This validates our development of the optimal Bayesian fusion. Other simulations results not reported here confirm this conclusion.

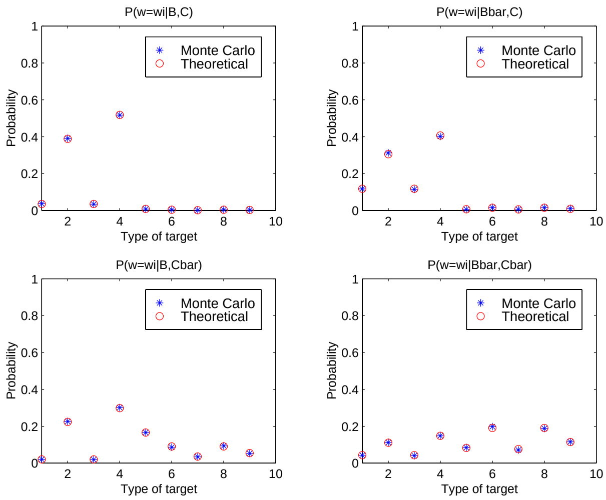


Figure 2: $s_1 \neq s_2, r_1 \neq r_2$ with non-uniform p_i

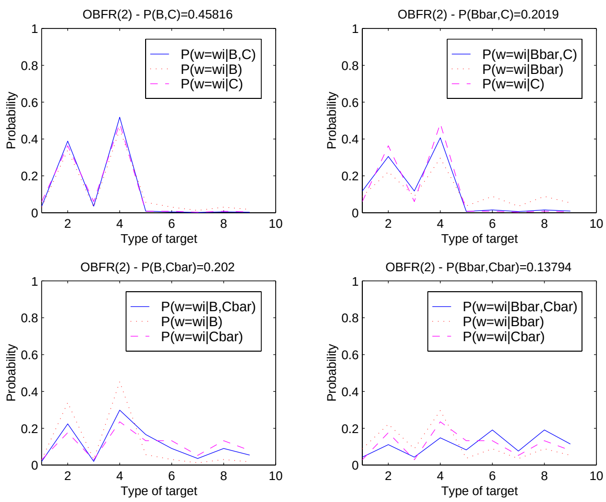


Figure 3: $P\{w = w_i | {W_1}^{\delta_1}, {W_2}^{\delta_2}\}$ vs. $P\{w = w_i | {W_i}^{\delta_i}\}$

Since pretty good reliability parameters have been chosen in this example, Bayesian fusion rule always emphasizes (in this case) the probabilities of all hypotheses w_i belonging to intersection of joint assertion ${W_1}^{\delta_1} \cap {W_2}^{\delta_2}$. However, the improvement on these probabilities provided by optimal fusion rule is not that much.

Note : Continuous plots have been drawn on figure 3 to facilitate the comparison however it is obvious that only values at abscisses w_i must be considered for the comparison of conditional pmf (probability mass functions).

- **OBFR(2) with uniform prior**

Focal hypotheses $w_i, i = 1, \dots, 9$ have now *uniform priors* $p_i = 1/9, i = 1, \dots, 9$. s_1 and s_2 have same pretty good reliability factors as before. Results plotted on figures 4 and 5 confirm our previous remark. In this particular case, we can better distinguish the improvement obtained by the OBFR(2) on hypotheses w_i belonging to intersection of joint assertion ${W_1}^{\delta_1} \cap {W_2}^{\delta_2}$.

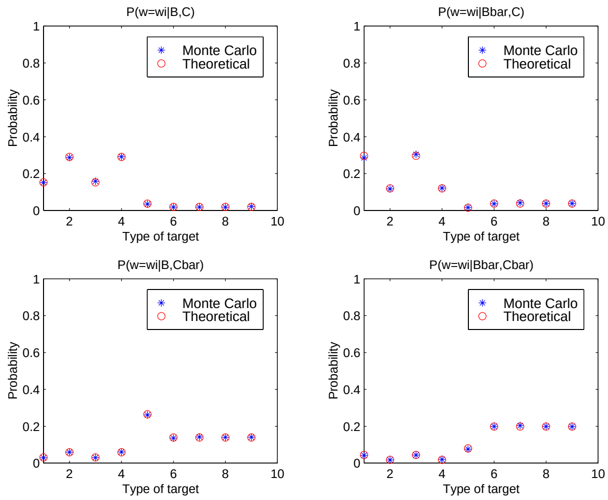


Figure 4: $s_1 \neq s_2, r_1 \neq r_2$ with uniform p_i

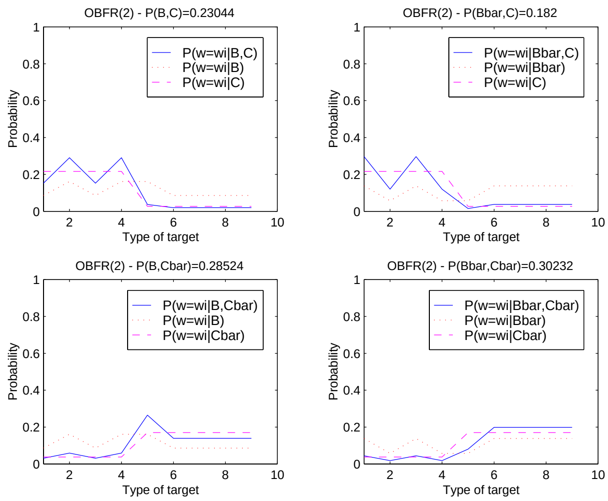


Figure 5: $P\{w = w_i | {W_1}^{\delta_1}, {W_2}^{\delta_2}\}$ vs. $P\{w = w_i | {W_i}^{\delta_i}\}$

2.8.2 OBFR(2) with similar classifiers

Consider now two sensors s_1 and s_2 having the same discrimination capacity (i.e. $(W_1 = B) \equiv (W_2 = C)$) but not necessarily same reliability parameters r_1 and r_2. Such sensors are said similar; In our simulations, we have chosen $(W_1 = B) \equiv (W_2 = C) = \{w_2, w_4, w_5\} \subset W$ and all our theoretical predictions match very well with Monte Carlo results. Due to space limitation, we don't have included Monte Carlo results here. Only theoretical fusion results are shown on figure 6 to see the benefit brought by the OBFR(2) for several typical cases.

- **OBFR(2) with $r_1 \neq r_2$ and non uniform p_i**

Figure 6.1 plots result obtained with non uniform p_i chosen as in previous section and reliabilities $r_1 = (0.60, 0.75)$ and $r_2 = (0.90, 0.80)$. In most of cases, s_1 and s_2 provide the same assertions since $P\{W_1 = B, W_2 = C\} = 0.46255$. But since s_2 is more reliable than s_1, the pmf $P\{w_i | W_1 = B, W_2 = C\}$ is a little bit better than $P\{w_i|W_2 = C\}$ for $w_i \in C$. The improvement of OBFR(2) is very limited in this case. In the dual case, where classifiers claim both $W_1 = W_2 = \bar{B} \equiv \bar{C}$, we can see also a small improvement for $w_i \in \bar{B}$. When similar classifiers claim paradoxal/conflictual assertions (i.e. $W_1 = B$ and $W_2 = \bar{C}$ or $W_1 = \bar{B}$ and $W_2 = C$), then no improvement is obtained at all since there is no $w_i \in B \cap \bar{B}$. The fusionned pmf only follows the conditional pmf corresponding to the best classifier which intuitively makes sense.

- **OBFR(2) with $r_1 = r_2$ and non uniform p_i**

Figure 6.3 plots results obtained with non uniform prior distribution and good and same reliabilities ($r_1 = r_2 = (0.90, 0.90)$). When both classifiers agree, one has a small improvement for all w_i which support the assertions. When classifiers disagree, the fusionned pmf tries to reduce the max performances of both classifiers (i.e. $\forall i, P\{w_i|W_1, W_2\} \leq \max(P\{w_i|W_1\}, P\{w_i|W_2\})$). In the dual case, where we force the same classifiers to have same poor reliability parameters ($r_1 = r_2 = (0.10, 0.10)$), one gets results plotted on figure 6.5. Dual conclusions can be drawn when classifiers agree (i.e. a small improvement for all w_i which do not support the assertions), but same concluding remark when classifiers disagree.

- **OBFR(2) with $r_1 \neq r_2$ and uniform p_i**

Figure 6.2 plots results obtained with uniform prior p_i with pretty good different reliability parameters (i.e. $r_1 = (0.60, 0.75)$ and $r_2 = (0.90, 0.80)$). Same concluding remarks as for non uniform case hold.

- **OBFR(2) with $r_1 = r_2$ and uniform p_i**

Figure 6.4 plots results obtained with uniform p_i and same good reliabilities ($r_1 = r_2 = (0.90, 0.90)$). Figure 6.6 plots results obtained with uniform p_i and same bad reliabilities $r_1 = r_2 = (0.10, 0.10)$). Same concluding remarks as for the non uniform case hold. It is worthwhile to note that OBFR now generates (because of the uniform prior condition) the full ignorance pmf (taken as uniform pmf if we admit the principle of sufficient reason) when classifiers are in full contradiction. This result makes sense with our logical intuition.

3 General case : OBFR(N)

The extension of optimal Bayesian fusion rule to general case of the N unreliable classifiers problem follows directly from previous results. Details of derivations will be omitted in the sequel due to space limitation. We will just indicate important results for the OBFR(N).

3.1 Derivation of $P\{W_1^{\delta_1}, \ldots, W_N^{\delta_N}\}$

Relation (1) can easily be extended for $N > 2$ as follows to provide $P\{W_1^{\delta_1}, \ldots, W_N^{\delta_N}\}$,

$$\sum_{(a_1,\ldots,a_N)\in\{0,1\}^N} \prod_{n=1,N} f_n[\Delta_n, \delta_n] \sum_{i\in I_1^{a_1}\cap\ldots\cap I_N^{a_N}} p_i$$

3.2 Derivation of $P\{W_1^{\delta_1}, \ldots, W_N^{\delta_N}|w_i\}$

$P\{W_1^{\delta_1}, \ldots, W_N^{\delta_N}|w = w_i\}$ is given by

$$\prod_{n=1,N} P\{W_n^{\delta_n}|w = w_i\} = \prod_{n=1,N} f_n[\Delta_n[i], \delta_n]$$

3.3 Derivation of $P\{w_i|W_1^{\delta_1}, \ldots, W_N^{\delta_N}\}$

$P\{w = w_i|W_1^{\delta_1}, \ldots, W_N^{\delta_N}\}$ is given by

$$P\{w = w_i|W_1^{\delta_1}, \ldots, W_N^{\delta_N}\} = \frac{p_i \prod_{n=1,N} f_n[\Delta_n[i], \delta_n]}{k}$$

with $$k = \sum_{i=1,M} p_i \prod_{n=1,N} f_n[\Delta_n[i], \delta_n]$$

3.4 Final expression of OBFR(N)

The general OBFR(N) is then given by

$$P\{w = w_i|W_1^{\delta_1}, \ldots, W_N^{\delta_N}\} = \frac{p_i^{1-N} \prod_{n=1,N} P\{w_i|W_n^{\delta_n}\}}{K_N} \tag{18}$$

with $K_N = \sum_{i=1,M} p_i^{1-N} \prod_{n=1,N} P\{w_i|W_n^{\delta_n}\}$

If we assume *uniform* prior $p_i = 1/M$, then terms p_i^{1-N} can be removed in the OBFR(N) formula above. This formula coincides exactly with Demspter-Shafer rule of combination when basic mass assignments become basic conditional probabilities.

Since the list of w_i is exhaustive and w_i are mutually exclusive, any disjunction of general hypothese can be easily evaluated from (18) because of additivity property of probabilities. Hence for example, the fusioned conditional probability $P\{A|W_1^{\delta_1}, \ldots, W_N^{\delta_N}\}$ will be computed by

$$\sum_{w_i \subset A} \Big[\frac{p_i^{1-N}}{K_N} \prod_{n=1,N} P\{w_i|W_n^{\delta_n}\}\Big]$$

Several $N > 2$ multi-classifier bayesian fusion problems have been simulated. All predicted theoretical results have shown a very good agreement with all Monte Carlo simulation results but are not reported here due to space limitation.

3.5 A note on implementation of OBFR(N)

OBFR(N) can be implemented directly following (18) or through any sequential ways involving combinations of clusters of sub-joint assertions. Final OBFR result will not depend on the order of combinations. Consider as example, the four unreliable classifier problem. The direct OBFR(4) is given by (18), i.e.

$$P\{w = w_i|W_1^{\delta_1}, \ldots, W_4^{\delta_4}\} = \frac{p_i^{-3} \prod_{n=1,4} P\{w_i|W_n^{\delta_n}\}}{K_4}$$

This formula can however, by example, be rewritten as

$$\frac{p_i^{-2}\prod_{n=1,3}P\{w_i|W_n^{\delta_n}\}\cdot p_i^{-1}P\{w_i|W_4^{\delta_4}\}}{K_4}$$

By introducing the K_3 constant, one has now

$$\frac{\frac{p_i^{-2}\quad{}_{n=1,3}P\{w_i|W_n^{\delta_n}\}}{K_3}\cdot p_i^{-1}P\{w_i|W_4^{\delta_4}\}}{K_4/K_3}$$

which corresponds to nothing but the OBF between one classifier providing $W_4^{\delta_4}$ and a cluster of 3 classifiers providing joint assertions $(W_1^{\delta_1}, W_2^{\delta_2}, W_3^{\delta_3})$. The same reasoning is valid for any classifiers clustering choice.

4 Conclusion

The optimal bayesian fusion rule for the general unreliable multi-classifier fusion problem has been fully developped here. Our theoretical results have been intensively validated through comparison with Monte carlo simulation results. We have proved that a very good accuracy of theoretical prediction with experience is achievable. The Optimal Baysian fusion rule (OBFR) is actually very easy to implement and requires only a very low cost of computation even for the general case of N classifiers. This OBFR is very useful to optimize detection/classification performances of future multi-sensor/classifier systems and also to solve many practical problems arising in multi-sensor systems. We have pointed out the typical behavior of OBFR for several different practical cases involving same or different sensors having either same, or different reliability parameters.

References

[1] Bogler P.,"Shafer-Dempster Reasoning with Applications to Multisensor Target Identification Systems", IEEE Trans. on SMC, Vol. 17, Dec. 1987.

[2] Buede D.M., Girardi P.,"A Target Identification Comparison of Bayesian and Dempster-Shafer Multisensor Fusion", IEEE Trans. on SMC, Vol. 27, no 5, pp. 569-577, 1997.

[3] Chair Z., Varshney P.K.,"Optimal Data Fusion in Multiple-Sensor Detection Systems", IEEE Trans. on AES, Vol. 22, no 1, pp. 98-101, 1986.

[4] Dasarathy B.,"Decision Fusion", IEEE Computer Society Press, Los Amitos, CA, 1994.

[5] Dasarathy B.,"Fuzzy Evidential Reasoning Approach to Target Identity and State Fusion in Multisensor Environments", Optical Engineering, Vol. 36, no 3, pp. 669-683, 1997.

[6] Fixsen D., Mahler R.P.S.,"The Modified Dempster-Shafer Approach to Classification", IEEE Trans. on SMC, Vol. 27, no 1, pp. 96-104, 1997.

[7] Hall D.L., Llinas J.,"An Introduction to Multisensor Data Fusion", Proc. of the IEEE, Vol. 85, no 1, pp. 6-23, 1997.

[8] Kittler J.,"Combining Classifiers: A Theoretical Framework", Pattern Analysis and Applications, Vol. 1, no 1, pp. 18-27, 1998.

[9] Li X.R.,"Probability, Random Signals, and Statistics",CRC Press, 1999.

[10] Lynch R.S., Willett P.K."Bayesian Classification Using Mislabeled Training Data and a Noninformative Prior", Journal of Franklin Institute, July, 1999.

[11] Nauck D., Kruse R.,"A Neuro-Fuzzy Method to Learn Fuzzy Classification Rules from Data", Fuzzy Sets and Systems, Vol. 89, no 3, pp.277-288, 1997.

[12] Rao N.S.V.,"Distributed Decision Fusion Using Empirical Estimation", IEEE Trans. on AES, Vol. 33, no 4, pp. 1106-1114, Oct. 1997.

[13] Rao N.S.V.,"A Fusion Method that Performs Better than Best Sensor", Proceedings Fusion'98, pp 19-26, Las Vegas, July 1998.

[14] Shafer G.,"The Combination of Evidence", International Journal of Intelligent Systems, Vol. 1, pp. 155-179, 1986.

[15] Tartakovsky A., Li X.R.,"Sequential Testing of Multiple Hypotheses in Distributed Systems", Proc. of Fusion 2000, Vol. 2, Paris, July 2000.

[16] Tenney R.R., Sandell N.R.Jr.,"Optimal Detection with Distributed Sensors", IEEE Trans. on AES, Vol. 17, no 4, pp. 501-510, 1981.

[17] Thomopoulos S.C.A., Viswanathan R., Bougoulias B.K.,"Optimal and Suboptimal Distributed Decision Fusion", IEEE Trans. on AES, Vol. 25, no 5, pp. 761-765, 1989.

[18] Valet L., Mauris G., Bolon Ph.,"A Statistical Overview of Recent Literature in Information Fusion", Proceedings of Fusion 2000 Conf, Vol. 1, Paris, July 2000.

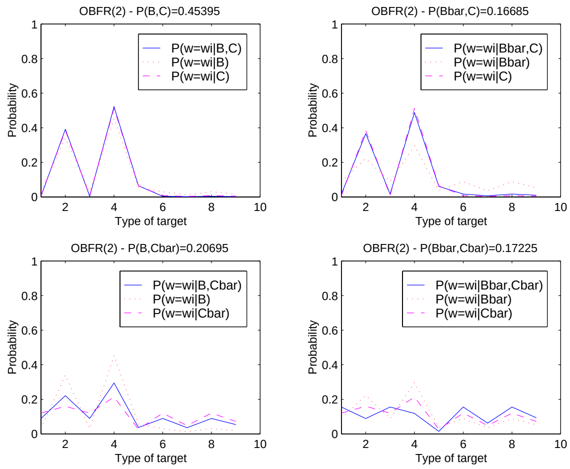
[19] Varshney P.K.,"Distributed Detection and Data Fusion", Springer-Verlag, New York, 1997.

[20] Waller W.G., Jain A.K.,"On the Monotonicity of the Performance of Bayesian Classifiers", IEEE Trans. on IT, Vol. 24, no 3, pp. 392-394, May 1978.

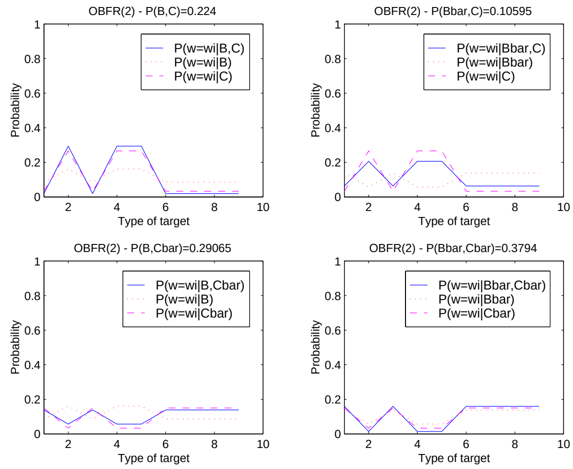
[21] Willett P.K.,"Detection Fusion Under Dependence", Proc. of Fusion 2000, Vol. 2, Paris, July 2000.

[22] Williams M.L., Wilson R.C., Hancock E.R.,"Multisensor Fusion with Bayesian Inference", Lecture Notes in Computer Science, Vol. 1296, pp. 25-32, 1997.

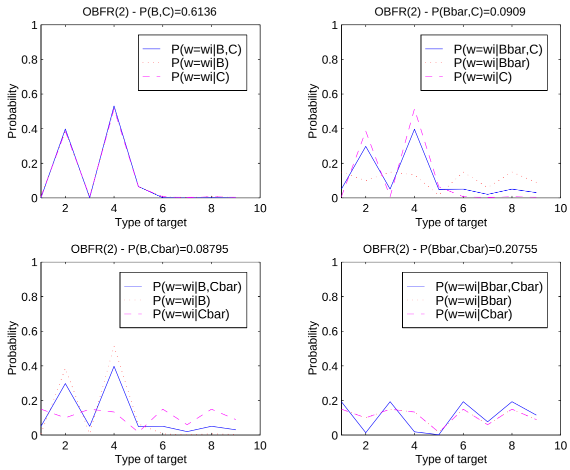
[23] Zhu Y.M., Li X.R., You Z.,"Unified Fusion Rule in Multisensor Network Decision Systems", Proc. of Fusion 2000 Conf, Vol. 2, Paris, July 2000.



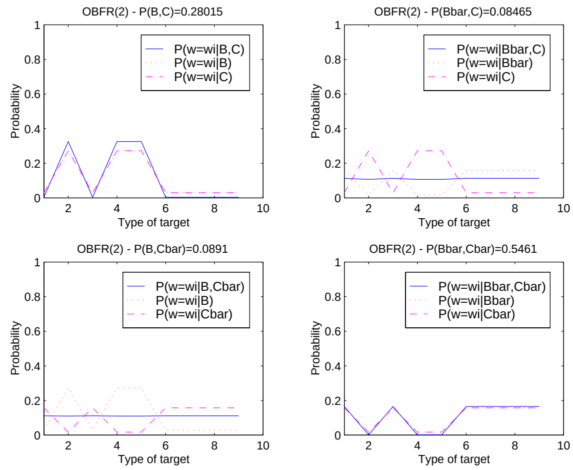
6.1: $r_1 \neq r_2$; non uniform p_i



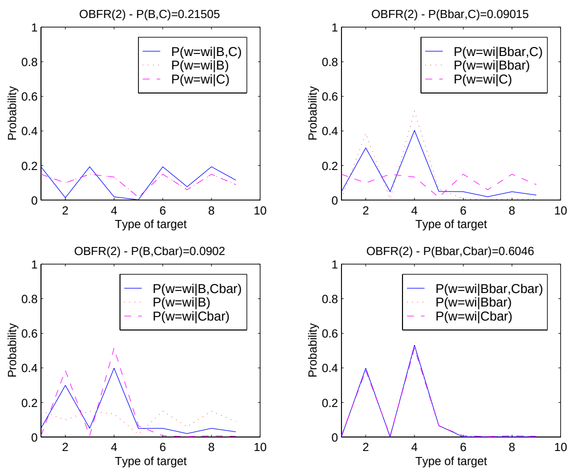
6.2: $r_1 \neq r_2$; uniform p_i



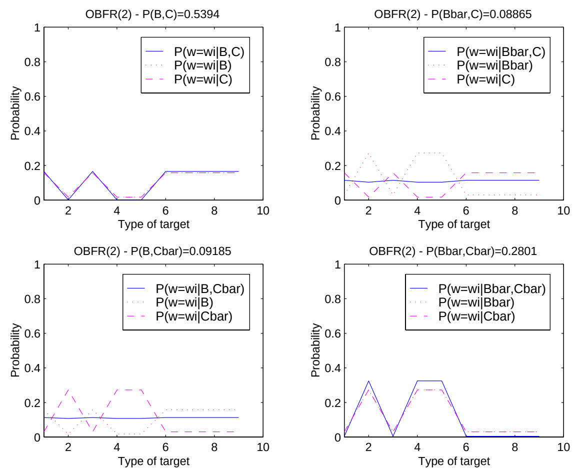
6.3: $r_1 = r_2 =$ "good"; non uniform p_i



6.4: $r_1 = r_2 =$ "good"; uniform p_i



6.5: $r_1 = r_2 =$ "poor"; non uniform p_i



6.6: $r_1 = r_2 =$ "poor"; uniform p_i

Figure 6: Fusion Results with similar classifiers ($s_1 = s_2$)