# Comparing Fusors within a Category of Fusors

**Steven N. Thorsen**
Air Force Institute of Technology
Department of Mathematics and Statistics
2950 Hobson Way
WPAFB, OH 45433-7765
USA
steven.thorsen@afit.edu

**Mark E. Oxley**
Air Force Institute of Technology
Department of Mathematics and Statistics
2950 Hobson Way
WPAFB, OH 45433-7765
USA
mark.oxley@afit.edu

**Abstract –** *Category Theory is used to describe a category of fusors. The category is formed from a model of a process beginning with an event and leading to the final labeling of the event. Although many techniques of fusing information have been developed the inherent relationships among different types of fusion techniques (fusors) have not yet been fully explored. In this paper, a foundation of fusion is presented, definitions developed, and a method of measuring the performance of fusors is given. Functionals on receiver operating characteristic (ROC) curves are developed to form a partial ordering of a set of classifier families. The functional also induces a category of fusion rules. The treatment includes a proof of how to find the Bayes optimal classifier (or Bayes Optimal fusor, if available) from a ROC curve.*

**Keywords:** information fusion, fusors, ROC, ROC curves, Bayes Optimal.

## 1 Introduction

Information fusion is a rapidly advancing science. Researchers are daily adding to the known repertoire of fusion techniques (fusion rules). An agency that is building a fusion system to detect or identify objects is bound to want to get the best possible result for the money expended. It is with this goal in mind that we need a way to compete various fusion rules for acquisition purposes. It appears that the receiver operating characteristic curves (ROC curves) that can be developed for such systems under test conditions may serve well in this regard. We will demonstrate the development of a functional on ROC curves which will allow us, under certain assumptions and constraints, to compete classifiers, fusors (fusion rules with a constraint), and fusion systems in order to choose the best from among finitely many competitors.

## 2 Category Theory Preliminaries

Category theory is a branch of mathematics useful for determining universal properties of classes. The science of information fusion does not yet know of all the relationships involved between the classes of data and the mappings from one type of data to another. It has been our goal to try to engage the community to think in terms of generalities when studying fusion processes in order to abstract the processes and perhaps gain some clarity of thought, if not genuine insight. I have drawn upon the work of various authors [1, 2, 3, 4]to present the definitions.

**Definition 1 (Category)** *A category $\mathcal{C}$ consists of the following:*

A1. A collection of objects denoted $\mathbf{Ob}(\mathcal{C})$.

A2. A collection of arrows (maps) denoted $\mathbf{Ar}(\mathcal{C})$.

A3. Two mappings, called Domain ($dom$) and Codomain ($cod$), which assign to an arrow $f \in \mathbf{Ar}(\mathcal{C})$ a domain and codomain from the objects of $\mathbf{Ob}(\mathcal{C})$. Thus, for arrow $f$, given by $O_1 \xrightarrow{f} O_2$ , $dom(f) = O_1$ and $cod(f) = O_2$.

A4. A mapping assigning each object $\mathrm{O} \in \mathbf{Ob}(\mathcal{C})$ an unique arrow $1_{\mathrm{O}}$ called the identity arrow, such that

$$\mathrm{O} \xrightarrow{1_{\mathrm{O}}} \mathrm{O}$$

and such that for any existing element, $x$, of O, we have that

$$x \xmapsto{1_{\mathrm{O}}} x.$$

A5. A map, $\circ$ , called composition, $\mathbf{A} \times \mathbf{A} \xrightarrow{\circ} \mathbf{A}$ . Thus, given $f, g \in \mathbf{A}$ with $cod(f) = dom(g)$ there exists an unique $h \in \mathbf{A}$ such that $h = g \circ f$.

Axioms A3-A5 lead to the associative and identity rules:

- **Associative Rule**. Given appropriately defined arrows $f, g$, and $h$ we have that

$$(f \circ g) \circ h = f \circ (g \circ h).$$

- **Identity Rule**. Given arrows $A \xrightarrow{f} B$ and $B \xrightarrow{g} A$ , then there exists identity arrow $1_A$ such that $1_A \circ g = g$ and $f \circ 1_A = f$.

**Definition 2 (Subcategory)** *A subcategory $\mathcal{B}$ of $\mathcal{A}$ is a category whose objects are some of the objects of $\mathcal{A}$ and whose arrows are some of the arrows of $\mathcal{A}$, such that for each arrow $f$ in $\mathcal{B}$, $dom(f)$ and $cod(f)$ are in $\mathbf{Ob}(\mathcal{B})$, along with each composition of arrows, and an identity arrow for each element of $\mathbf{Ob}(\mathcal{B})$.*

A category of interest is the category $\mathbf{Set}$, which has as objects sets and arrows all total functions, with composition of functions as the composition. Clearly this construct has identity arrows and the associative rule applies, so it is indeed a category. The subcategories of interest to us are subcategories of particular types of data sets, denoted $\mathcal{D}$, with objects similar types of data sets and arrows only the identity arrows, and subcategories of particular types of feature sets, denoted $\mathcal{F}$, with objects similar types of feature sets, and arrows only the identity arrows. The objects and arrows of these categories shall correspond to a particular sensor system, so will represent all of the possible data (or feature) sets that can be generated by the sensor-processor system. For example, the data generated by a particular sensor system may be $2x2$ real-valued matrices. In this case, $\mathcal{D} = (\mathbb{R}^{2x2}, \mathbf{id}_{\mathcal{D}}, \mathbf{id}_{\mathcal{D}}, \circ)$ represents the category with only the identities as arrows, and $\circ$ being the usual composition of functions.

A further categorical term that will be useful is that of a **functor**.

**Definition 3 (Functor)** *A **functor** $\mathfrak{F}$ between two categories $\mathcal{A}$ and $\mathcal{B}$ is a pair of maps $(\mathfrak{F}_{\mathbf{Ob}}, \mathfrak{F}_{\mathbf{Ar}})$*

$$\mathbf{Ob}(\mathcal{A}) \xrightarrow{\mathfrak{F}_{\mathbf{Ob}}} \mathbf{Ob}(\mathcal{B})$$

$$\mathbf{Ar}(\mathcal{A}) \xrightarrow{\mathfrak{F}_{\mathbf{Ar}}} \mathbf{Ar}(\mathcal{B})$$

*such that $\mathfrak{F}$ maps $\mathbf{Ob}(\mathcal{A})$ to $\mathbf{Ob}(\mathcal{B})$ and $\mathbf{Ar}(\mathcal{A})$ to $\mathbf{Ar}(\mathcal{B})$ while preserving the associative property of the composition map and preserving identity maps.*

Thus, given categories $\mathcal{A}, \mathcal{B}$ and functor $\mathfrak{F} : \mathcal{A} \longrightarrow \mathcal{B}$, if $\mathrm{A} \in \mathbf{Ob}(\mathcal{A})$ and $f, g, h, 1_{\mathrm{A}} \in \mathbf{Ar}(\mathcal{A})$ such that $f \circ g = h$ is defined, then there exists $\mathrm{B} \in \mathbf{Ob}(\mathcal{B})$ and $f', g', h', 1_{\mathrm{B}} \in \mathbf{Ar}(\mathcal{B})$ such that

*i)* $\mathfrak{F}(\mathrm{A}) = \mathrm{B}$.

*ii)* $\mathfrak{F}(f) = f'$, $\mathfrak{F}(g) = g'$.

*iii)* $h' = \mathfrak{F}(h) = \mathfrak{F}(f \circ g) = \mathfrak{F}(f) \circ \mathfrak{F}(g) = f' \circ g'$.

*iv)* $\mathfrak{F}(1_{\mathrm{A}}) = 1_{\mathfrak{F}(\mathrm{A})} = 1_B$.

**Definition 4 (Natural Transformation)** *Given categories $\mathcal{A}$ and $\mathcal{B}$ and functors $\mathfrak{F}$ and $\mathfrak{G}$ with $\mathcal{A} \xrightarrow{\mathfrak{F}} \mathcal{B}$ and $\mathcal{A} \xrightarrow{\mathfrak{G}} \mathcal{B}$, then a **Natural Transformation** is a family of arrows $\nu = \{\nu_{\mathrm{A}} | \mathrm{A} \in \mathcal{A}\}$ such that for each $f \in \mathbf{Ar}(\mathcal{A})$, $\mathrm{A} \xrightarrow{f} \mathrm{A}'$, $\mathrm{A}' \in \mathcal{A}$, the square*

$$\begin{array}{ccc} \mathfrak{F}(\mathrm{A}) & \xrightarrow{\nu_A} & \mathfrak{G}(\mathrm{A}) \\ {\scriptstyle \mathfrak{F}(f)}\downarrow & & \downarrow{\scriptstyle \mathfrak{G}(f)} \\ \mathfrak{F}(\mathrm{A}') & \xrightarrow{\nu_{\mathrm{A}'}} & \mathfrak{G}(\mathrm{A}') \end{array}$$

*commutes. We then say the arrows $\nu_{\mathrm{A}}$ are the components of $\nu : \mathfrak{F} \longrightarrow \mathfrak{G}$, and call $\nu$ the natural transformation of $\mathfrak{F}$ to $\mathfrak{G}$.*

**Definition 5 (Functor Category $\mathcal{A}^{\mathcal{B}}$)** *Given categories $\mathcal{A}$ and $\mathcal{B}$, the notation $\mathcal{A}^{\mathcal{B}}$ refers to the category of all functors $\mathfrak{F}$, $\mathcal{B} \xrightarrow{\mathfrak{F}} \mathcal{A}$. This category has all such functors as objects and the natural transformations between them as arrows.*

**Definition 6 (Product Category)** *Let $\{\mathcal{D}_i\}_{i=1}^n$ represent a finite collection of data set categories. Then $\prod_{i=1}^n \mathcal{D}_i$ is the corresponding product category.*

## 3 Modelling Fusion within the Event-Decision Model

Let $\mathrm{X}$ be a set of states for some event, and $\mathrm{T} \subset \mathbb{R}$ be a bounded interval of time. Interval $\mathrm{T}$ sorts $\mathrm{X}$ such that we call $\mathrm{E} \subseteq \mathrm{X} \times \mathrm{T}$ an **event-state**. An event-state is then comprised of event-state elements, $e = (x, t)$, where $x \in X$ and $t \in \mathrm{T}$. Thus $e$ denotes a state $x$ at an instant of time $t$. Let $\mathbf{E} = \mathrm{X} \times \mathrm{T}$, be the set of all event-states for an event over time interval $\mathrm{T}$.

The following discussion can be expanded to a finite number of sensors, but for now consider the simple model of a multi-sensor process using two sensors in Figure 1. The sets $\mathrm{E}_i$, for $i \in \{1, 2\}$, are sets of event-states. It

$$\mathrm{E}_1 \xrightarrow{s_1} \mathrm{D}_1 \xrightarrow{p_1} \mathrm{F}_1 \xrightarrow{c_1} \mathrm{L}_1$$

$$\mathrm{E}_2 \xrightarrow{s_2} \mathrm{D}_2 \xrightarrow{p_1} \mathrm{F}_2 \xrightarrow{c_2} \mathrm{L}_2$$

Fig. 1: Simple Model of a Dual-Sensor Process.

is useful to think of $\mathrm{E}_i$ as the set of all possible states of an event (such as an aircraft flying) occurring within sensor $s_i$'s field of view. Given $\mathrm{E}_i$ thus defined, now define a sensor as a mapping from an event-state set to a data set, $\mathrm{D}_i$. The mapping $s_i$ is then a sensor. A data set could be a radar signature return of an object, multiple radar signature returns, a two-dimensional image, or even a video stream over the time period of the event-state set, for example. In any case we would like to extract recognizable features from the data set. Hence, mapping $p_i$ represents a processor which does just that. Processors are mappings from data sets into feature sets, $\mathrm{F}_i$. Finally, from the feature sets we want to determine a label or decision based upon the sensed event-state. This is achieved through use of the classifiers $c_i$ which map the feature set into a label set. The label set $\mathrm{L}_i$ can be as simple as the two-class set {target,non-target} or could have a more complex nature to it, such as the *types* of targets and non-targets in order to define the battlefield more clearly for the warfighter. Now the diagram in Figure 1 represents a simple sensor process pair involving two sensors, two processors, and two classifiers, but can easily be extended to any finite number. Now consider two sensors not necessarily co-located. Hence they may sense different event-state sets. Figure 1 models two sensors with differing fields of view. Performing fusion along any node or edge in this graph will result in an elevated level of fusion [5]–that of situation refinement or threat refinement, since we are not fusing common information about a particular object or objects. There are two other possible scenarios than Figure 1 depicts. The sensors can overlap in their field of view, either partially or fully, in which case fusing the information regarding object event-states within the intersection may be useful. Thus, a fusion process may be used to increase the reliability and accuracy of the system, above that which is possessed by either of the sensors on its own. Let $\mathrm{E}$ represent that event-state set that is common to both sensors, that is, $\mathrm{E} = \mathrm{E}_1 \cap \mathrm{E}_2$. Hence, there are two basic challenges regarding fusion. The first is how to fuse information from multiple sources regarding common event-states (or targets, if preferred) for the purpose of knowing the event-state (presumably for the purposes of tracking, identifying, and estimating future event-states). The second and much more challenging problem is to fuse information from multiple sources regarding event-states not common to all sensors, for the purpose of knowing the state of a situation (the situation-state), such as an enemy situation or threat assessment. We label the two types of fusion scenarios discussed **event-state fusion** and **situation-state fusion**. Therefore, Figure 2 represents the Event-State-Decision model of a dual sensor process. The

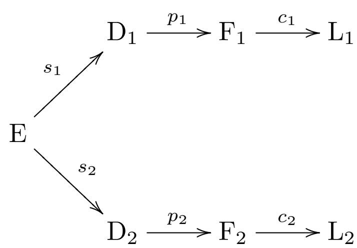


Fig. 2: Dual Sensor Process for Overlapping Field of View.

only restriction necessary for the usefulness of this model is that a common field of view (the event-state) be used. For example, $\mathrm{D}_1$ and $\mathrm{D}_2$ can actually be the same data set under the model, while $s_1$ and $s_2$ could be different sensors.

**Definition 7 (Fusion Rule)** *Let $\prod_{i=1}^{n} \mathcal{D}_i$ be a product category of data (or feature) set categories. Then a fusion rule is a functor $\mathfrak{R} \in \mathcal{D}_0^{\prod_{i=1}^{n} \mathcal{D}_i}$ and $\mathcal{D}_0$ is the resulting data set category.*

The key to this definition is to realize that a fusion rule $\mathfrak{R}$ (see Figure 3) simply combines the inputs from a product category into a resultant data set (or feature set), which is an element of a single data (or feature) set category. There is no restriction on the output with regards to being a "better" output than a system designed without a fusion rule.

We now desire to show how defining a fusor (see Definition 9) as a fusion rule with a constraint changes the sensor process model into an Event-State Fusion model. Continuing to consider the dual sensor process in Figure 2, a fusion rule can be applied to either the data sets or the feature sets. Given a fusion rule $\mathfrak{R}$ for the two data sets as in Figure 3, our model becomes that of Figure 5. A new data set, processor, feature set, and classifier may become necessary as a result of the fusion rule having a

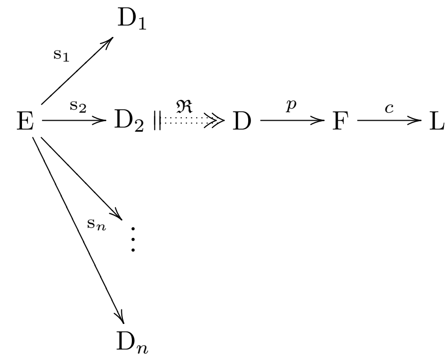


Fig. 3: Fusion Rule on Category of Data Sets.

different codomain than the previous systems. The label set may change also, but for the remainder of this paper we are interested only in a two class label set, that of $\mathrm{L} = \mathrm{L}_1 = \mathrm{L}_2 = \{\mathrm{Target}, \mathrm{Nontarget}\}$. In a homogeneous

$$(\mathrm{D}_1, \mathrm{D}_2) \| \overset{\mathfrak{R}}{\ggg} \mathrm{D}_3$$

Fig. 4: Fusion Rule Applied to Data Sets.

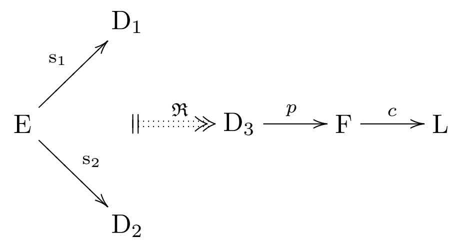


Fig. 5: Fusion Rule Applied within a dual sensor process.

(or within) fusion scenario, the data sets (or feature sets) are the same, $\mathrm{D}_1 = \mathrm{D}_2 = \mathrm{D}_3$. This is true in the case that the sensors used are the same type (that is, they collect the same measurements, but from possibly different locations relative to the overlapping field of view. In the case where the data sets (or feature sets) are truly different, a composite data set (and/or feature set) which is different from the first two (possibly even the product of the first two) is created as the codomain of the fusion rule functor.

Now at this point we may consider, in what way is the process modeled in Figure 5 **superior** to the original processes shown in Figure 2? One way of comparing performance in such systems is to compare the processes' receiver operating characteristics (ROC curves).

### 3.1 Developing a ROC Curve

Setting aside the fusion process for a moment, we focus on the classification process

$$\mathrm{F} \xrightarrow{c} \mathrm{L}.$$

Assume that $\mathrm{F}$ is a probability space, $\mathrm{F}$ can be denoted equivalently as $(\mathrm{F}, \mathscr{B}, \mathrm{Pr})$ where $\mathrm{Pr}$ is a probability measure and $\mathscr{B}$ is the associated $\sigma$-field. Recall that $\mathrm{L}$ is a two-class label set, T=target, N=non-target, and L= $\{\mathrm{T}, \mathrm{N}\}$.

Finally, consider the hypothetical "perfect" classifier $c^*$, the classifier which always matches a feature element with the correct label. Subjecting our processes to tests we can run a collection of features through the classifier and produce a corresponding label. Given $x \in \mathrm{F}$ and using the inverse image of the classifier we can calculate the hit rate,

$$\mathrm{P}_{tp} = \frac{\Pr\{x \mid x \in c^{-1}(\mathrm{T}) \wedge x \in c^{*-1}(\mathrm{T})\}}{\Pr\{x \mid x \in c^{*-1}(\mathrm{T})\}} \tag{1}$$

and the false alarm rate,

$$\mathrm{P}_{fp} = \frac{\Pr\{x \mid x \in c^{-1}(\mathrm{T}) \wedge x \in c^{*-1}(\mathrm{N})\}}{\Pr\{x \mid x \in c^{*-1}(\mathrm{N})\}}. \tag{2}$$

The ordered pair $(\mathrm{P}_{fp}, \mathrm{P}_{tp}) \in [0,1] \times [0,1]$ is the ROC for the system. Now it is desirable for a classifying system to have a parameter associated with the classifier, such that changing the parameter (which is possibly multidimensional) changes the ROC. In such a case, a parameter set $\Theta$ would be chosen such that the associated classifier family $\{c_\theta\}_{\theta\in\Theta}$ continuously maps the feature set into the label set in a bijection, and such that the curve $f = (\mathrm{P}_{fp}(c_\theta), \mathrm{P}_{tp}(c_\theta))$ is the projection of the trajectory $\hat{f} = (\theta, \mathrm{P}_{fp}(c_\theta), \mathrm{P}_{tp}(c_\theta))$ into the $\mathrm{P}_{fp} - \mathrm{P}_{tp}$ plane. In this case we have that

$$\mathrm{P}_{tp}(\theta) = \frac{\Pr\{x \mid x \in c_\theta^{-1}(\mathrm{T}) \wedge x \in c^{*-1}(\mathrm{T})\}}{\Pr\{x \mid x \in c^{*-1}(\mathrm{T})\}} \tag{3}$$

and

$$\mathrm{P}_{fp}(\theta) = \frac{\Pr\{x \mid x \in c_\theta^{-1}(\mathrm{T}) \wedge x \in c^{*-1}(\mathrm{N})\}}{\Pr\{x \mid x \in c^{*-1}(\mathrm{N})\}}. \tag{4}$$

Call such a parameter set an admissible parameter set. Note the parameter need not necessarily be associated with the classifier of the system, but could be associated instead with the sensor(s), processor(s), or any combination of the three. What is key is that the final parameter set must produce a corresponding ROC curve as a continuous curve from $(0,0)$ through $(1,1)$ in the $\mathrm{P}_{fp} - \mathrm{P}_{tp}$ plane as the example in Figure 6 shows. The parameter $\theta$ is the threshold of the ROC. Is there a threshold among a particular family of classifiers that performs best? It is well-known and accepted that the threshold for which the probability of a misclassification (or Bayes error) is minimized is considered best and denoted the Bayes optimal threshold (BOT). That is, if $\theta^*$ is the solution to the problem

$$\begin{aligned}
&\min_{\theta\in\Theta} \quad [\Pr\{x \in \mathrm{F} : (x \in c_\theta^{-1}(\mathrm{T}) \wedge x \in c^{*-1}(\mathrm{N})) \\
&\qquad\qquad \vee (x \in c_\theta^{-1}(\mathrm{N}) \wedge x \in c^{*-1}(\mathrm{T}))\}] \\
&= \min_{\theta\in\Theta} \quad [\Pr\{x \in \mathrm{F} : (x \in c_\theta^{-1}(\mathrm{T}) \wedge x \in c^{*-1}(\mathrm{N}))\} \\
&\qquad\qquad + \Pr\{(x \in c_\theta^{-1}(\mathrm{N}) \wedge x \in c^{*-1}(\mathrm{T}))\}] \\
&= \min_{\theta\in\Theta} \quad [\mathrm{P}_{fp}(\theta)\Pr(\mathrm{N}) + (1 - \mathrm{P}_{tp}(\theta))\Pr(\mathrm{T})]
\end{aligned} \tag{5}$$

where $\Pr(\mathrm{T})$ and $\Pr(\mathrm{N})$ are the prior probabilities of a target class and non-target class, respectively, then $\theta^*$ is the BOT for the family of classifiers $\{c_\theta\}_{\theta\in\Theta}$.

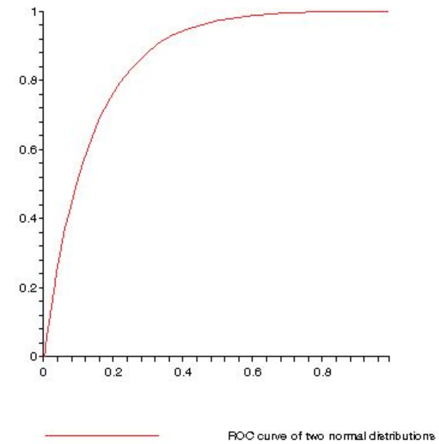



Fig. 6: A Typical ROC Curve

An obvious question at this point is given two families of classifiers, $\{a_\theta\}_{\theta\in\Theta}$ and $\{b_\pi\}_{\pi\in\Pi}$, which classifier is best? This is not an easy problem as seen in [6]. It is tempting to use some measure of the BOT, but notice that the BOT is dependent upon the selection of prior probabilities. The priors are generally not known, so selection of a better classifier based on ROC curves may not be possible, since ROC curves for different families can overlap. Rather, we should ask the question, given an operating threshold of prior probabilities, such as $\Pr(\mathrm{T}) = \frac{1}{4}$, can we choose among competing classifier families one that is superior to the others? One way to answer the question is derived in a very unexpected way.

### 3.2 A Variational Calculus Solution to Determining the Bayes Optimal Threshold of a Classifier Family

We will only consider ROC curves that are smooth (differentiable) over the entire range, i.e., given a ROC curve $f$, $f \in \mathrm{C}^1([0,1])$. Given a diagram describing the family of classifiers $\{c_\theta\}_{\theta\in\Theta}$, $\Theta$ an admissible parameter set, $(\mathrm{F}, \mathscr{B}, \Pr)$ being a probability space of feature vectors, and $\Theta$ an admissible parameter set, there is then a graph $G = \{(\theta, \mathrm{P}_{fp}(\theta), \mathrm{P}_{tp}(\theta) : \theta \in \Theta\}$ which we call the ROC trajectory. The projection of the ROC trajectory onto the $\mathrm{P}_{fp} - \mathrm{P}_{tp}$ plane, $f = \{(\mathrm{P}_{fp}(\theta), \mathrm{P}_{tp}(\theta)) : \theta \in \Theta\}$, is the ROC curve of the classifier family. Hence for $h \in [0,1]$ such that $h = \mathrm{P}_{fp}(\theta)$ for some $\theta \in \Theta$, we have that $[\mathrm{P}_{fp}]^{-1}(h) = \theta$. It is now clear that the BOT of the classifier family $\{c_\theta\}_{\theta\in\Theta}$, $\theta^*$, corresponds to some point $h^* = \mathrm{P}_{fp}(\theta^*) \in [0,1]$. So what can we learn about $h^*$? Consider the problem stated as follows:

> Among all smooth curves whose endpoints lie on the point $(0,1)$ and the ROC curve $y = f(h)$, find the curve for which the functional
>
> $$\mathrm{J}[y] = \int_0^h [\alpha + \beta|y'(t)|]dt \tag{6}$$

has a minimum subject to the constraints:

$$\begin{aligned} y(0) &= 0 \\ y(h) &= \mathrm{P}_{tp}(\theta) \end{aligned} \quad (7)$$

where $h = \mathrm{P}_{fp}(\theta)$ for some $\theta \in \Theta$ and $\beta = 1 - \alpha$ with $\alpha = \Pr(\mathrm{N})$, the prior probability of no target.

This functional is finding the curve with the smallest weighted Manhattan distance from the point $(0, 1)$ to the ROC curve. The constraints show that the curve must begin at $(0, 1)$ and terminate on the ROC curve. Any solution to Equation 6 must solve Euler's equation [7]

$$\mathrm{T}_y - \frac{d}{dt}\mathrm{T}_{y'} = 0. \quad (8)$$

where $\mathrm{T} = \alpha + \beta|y'(t)|$, so that $\mathrm{T}_y = 0$ and $\mathrm{T}_{y'} = \beta sign(y'(t))$. Hence we have that

$$-\frac{d}{dt}sign(y'(t)) = 0 \quad (9)$$

so that $sign(y'(t))$ is constant for all $t \in [0, 1]$. Thus $sign(y'(t))$ can be $0$ or $-1$ since the curve has the constraints of the endpoints $(0, 1)$ and a point on the ROC curve $f$. Now if $sign(y'(t)) = 0$ for all $t$, then $y(0) = y(h) = y(1)$ due to the smoothness of the ROC curve. Thus Equation 6 becomes

$$\mathrm{J}[y] = \alpha h = \Pr(N)\mathrm{P}_{fp}(\theta), \quad (10)$$

with $\mathrm{P}_{fp}(\theta) = 1$. Thus $\Pr(N) = 1$ and the weighted manhattan length of curve $y$ is therefore 1. On the other hand, if $sign(y'(t)) = -1$, then solving Equation 6 directly yields

$$\alpha t|_{t=0}^{t=h} + [\beta(sign(y'(t)))y(t)]_{t=0}^{t=h} \quad (11)$$

which reduces to

$$\mathrm{P}_{fp}(\theta)\Pr(\mathrm{N}) + (1 - \mathrm{P}_{tp}(\theta))\Pr(\mathrm{T}). \quad (12)$$

Notice that Equation 12 is identical to the unminimized Equation 5. Therefore, $h = h^*$ which minimizes Equation 12 corresponds to the BOT, $\theta^*$, of the family of classifiers! The transversality condition of the variation is

$$\begin{aligned} &\alpha + \beta|y'(t)|]_{t=h^*} \\ &\quad + [\beta(f'(t) - y'(t))(\mathrm{sign} y'(t))]_{t=h^*} = 0 \end{aligned} \quad (13)$$

so that

$$f'(t)_{t=h^*} = \frac{\alpha}{\beta}$$

which is

$$f'(t)_{t=h^*} = \frac{\Pr(\mathrm{N})}{\Pr(\mathrm{T})}. \quad (14)$$

So the transversality condition tells us that the BOT of a family of classifiers corresponds to a point on the ROC curve which has as a derivative the prior ratio $\frac{\Pr(\mathrm{N})}{\Pr(\mathrm{T})}$! Therefore, if one presumes a prior ratio of 1, then the point on the curve corresponding to the BOT will have a tangent to the ROC curve with slope 1. For many problems this will make the BOT very easy to find given the graphing capabilities of today's computers, especially when the parameter set, $\Theta$, is multidimensional. This gives us an idea of what would make a good functional for determining which classifier families are more desirable than others. An immediate approach would be to choose a preferred prior ratio and locate the BOTs for each competing classifier family. Since all the BOTs will have the same slope for lines tangent to their ROC curves at that point, the BOT with the tangent line closest to the point (0,1) would be considered the best choice. However, it is still possible that many ROC curves could be constructed so that the BOT for each one has the same tangent line. This would set up a rather large equivalence class of classifier families. This is the same problem faced when using area under the curve (AUC) of a ROC curve as a functional. In both cases the underlying posterior conditional probabilities are unknown and there are just too many possible combinations of posterior distributions that can produce ROC curves with the same AUC (or BOT tangent lines).

### 3.3 A Functional for Comparing Classifier Families

So, what criteria is best in selecting from among competing classifiers? We submit that first of all, among all ROC curves representing the competing classifier families, identifying the the BOT for each ROC is most important, since it is this threshold which minimizes the corresponding Bayes error. We can easily identify this point on a ROC curve presupposing only the prior probabilities $\Pr(\mathrm{N})$ and $\Pr(\mathrm{T})$, as demonstrated earlier. Furthermore, our decision objective is, in addition to minimizing Bayes error, to minimize $\mathrm{P}_{fp}$ while simultaneously maximizing $\mathrm{P}_{tp}$. The supremum BOT among all ROC curves would be the point $(0, 1)$, so we can codify the decision objective mathematically.

**Definition 8 (ROC Functional)** *Let $\{c_\theta\}_{\theta\in\Theta}$ be a classifier family with an admissible parameter set $\Theta$. Let $f$ be the corresponding ROC curve. Given data $\Gamma = (\alpha_0, \beta_0, \gamma)$, where $\alpha_0, \beta_0$ are acceptable levels for $\mathrm{P}_{fp}(\theta), \mathrm{P}_{tp}(\theta)$ respectively and $\Pr(N) = \gamma$, determine the point on the ROC curve, $(\mathrm{P}_{fp}(\theta^*), \mathrm{P}_{tp}(\theta^*))$, as the right endpoint of the smooth curve $y = \hat{y}$ which minimizes the functional:*

$$\mathrm{J}[y] = \int_0^h (\alpha + \beta|y'|)dt$$

*subject to the constraints $y(0) = 1$ and $y(h) = f(h)$ where $h = \mathrm{P}_{fp}(\theta)$ for some $\theta \in \Theta$. Call the minimized right endpoint $(h^*, f(h^*)) = (\mathrm{P}_{fp}(\theta^*), \mathrm{P}_{tp}(\theta^*))$. Let*

$$F(f) = \begin{cases} 0 & \text{if} \quad \mathrm{P}_{fp} > \alpha_0 \text{ or } \mathrm{P}_{tp} < \beta_0 \\ \textit{1-k} & \text{otherwise} \end{cases}$$

*where*

$$k = \sqrt{\mathrm{P}_{fp}(\theta^*)^2 + (1 - \mathrm{P}_{tp}(\theta^*))^2}dt.$$

*Call the functional $F(\,\cdot\,; \alpha_0, \beta_0, \gamma)$ the* ***ROC functional****.*

The ROC functional satisfies the requirements we set forth in our decision objectives. Taking the Euclidean distance between the point $(0, 1)$ to the point on the ROC corresponding to its system's BOT also allows us to make a better preference from among ROC curves, when more than one curve contains a BOT with the smallest weighted Manhattan distance from the point $(0, 1)$.

Now given a finite collection of competing classifier families

$$\mathrm{B} = \{b_1 = \{b_\theta\}_{\theta\in\Theta_1}, b_2 = \{b_\theta\}_{\theta\in\Theta_2}, \ldots, b_n = \{b_\theta\}_{\theta\in\Theta_n}\}$$

where $\{\Theta_1, \Theta_2, \ldots, \Theta_n\}$ is a collection of admissible parameter spaces, we say that for fixed data $(0, \beta_0, \gamma_0)$,

$$b_i \succeq b_j \iff F(f_{b_i}; 0, \beta_0, \gamma_0) \geq F(f_{b_j}; 0, \beta_0, \gamma_0). \quad (15)$$

In this way we have established a partial order on the set $\mathrm{B}$ of competing classifiers. Similarly, since there is a ROC curve associated with each classifier family, we say that

$$f_{b_i} \succeq_{ROC} f_{b_j} \iff F(f_{b_i}; 0, \beta_0, \gamma_0) \geq F(f_{b_j}; 0, \beta_0, \gamma_0). \quad (16)$$

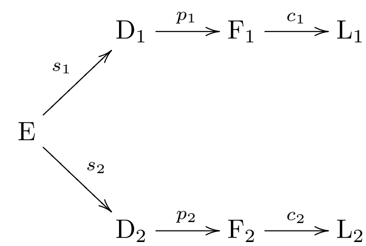
## 4 Fusors

We are now in a position to define a system in which we can compete fusion rules. Suppose we have a system such as that in Figure 2. Each branch has a ROC curve that can be associated with the classifier family, and we now have a viable means of competing each branch. If we can only choose among the two event-decision systems, take the one whose associated ROC functional is greater. Therefore, we can also compete these two event-decision systems with a system that fuses the two data sets (or the feature sets for that matter) by fixing a third classifier family and finding the ROC functional of the event-decision system corresponding to the fused data (features). If the fused system's ROC functional is greater than either of the original two, then the fusion rule is in fact a fusor. Repeating this process on a finite number of fusion rules, we discover a finite collection of fusors with associated ROC functional values. The fusor that is the best choice is then selected by finding the fusor corresponding to the largest ROC functional value.

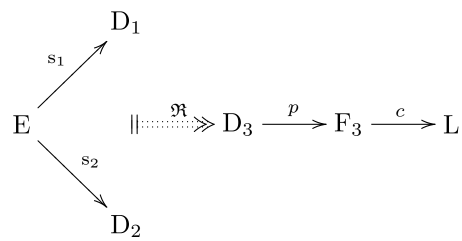
Do you want to change your a priori probabilities? Simply adjust $\gamma$ in the ROC functional's data and recalculate the BOTs for each system. Then calculate the ROC functional for each corresponding ROC and choose the largest value. The corresponding fusor is then the best fusor to select under your criteria. We have for each set of ROC functional data and each finite collection of fusion rules, a partial ordering of fusors.

**Definition 9 (fusor)** *A fusor is a fusion rule of an event-decision process which performs by means of a functional on its corresponding ROC curve better than any branch of the graph of the original processes before applying a fusion rule.*

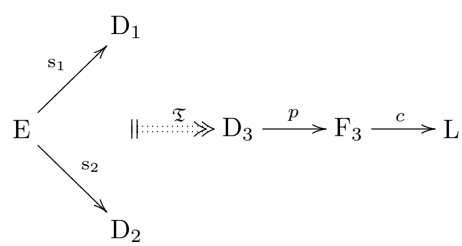
By way of example, suppose we start with the system




and consider a functional $F$ on the ROC curves $f_{c_1}$ and $f_{c_2}$ ($F$ being created under the assumptions and data of the researcher's choice). Then given fusion rules $\mathfrak{R}$ and $\mathfrak{T}$ such that




and




let $f_{\mathfrak{R}}$ and $f_{\mathfrak{T}}$ refer to the corresponding ROC curves to each of the fusion rule's systems (as a possible example of ROC curves of competing fusion rules see Figure 7 ). Then we have that if $F(f_{\mathfrak{R}}) \geq F(f_{c_i})$ for $i = 1, 2$ and if $F(f_{\mathfrak{T}}) \geq F(f_{c_i})$ for $i = 1, 2$ then we say that $\mathfrak{R}$, $\mathfrak{T}$ are fusors. Furthermore, suppose $F(f_{\mathfrak{R}}) \geq F(f_{\mathfrak{T}})$. Then we

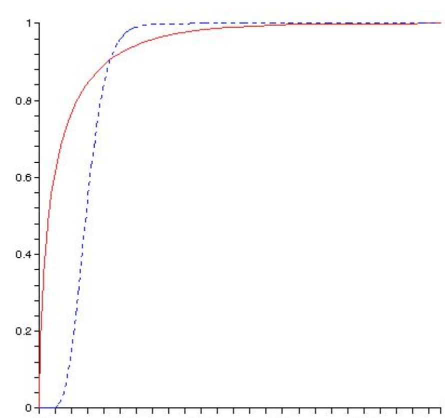



Fig. 7: ROC curves of Competing Fusion Rules

have that $\mathfrak{R} \succeq_{ROC} \mathfrak{T}$. Thus, $\mathfrak{R}$ is the fusor a researcher would select under the given assumptions and data.

## 5 Conclusion

A fusion researcher should have a viable method of comparing fusion rules. It is required to define fusion correctly, and to demonstrate to the scientific community improvements over existing methods. We have shown in this paper that every fusion system can generate a corresponding ROC curve, and under a mild assumption of smoothness of the ROC curve, a Bayes Optimal Threshold (BOT) can be found for each classifier family. Given additional assumptions on the a priori probabilities of a target or non-target, along with given thresholds for $\mathrm{P}_{fp}$ and $\mathrm{P}_{tp}$, a functional can be generated which will yield a real value for each ROC curve. This functional called the ROC functional will generate a partial order of classifier families, fusion rules, and ultimately fusors, which can then be used to select the best fusor from among a finite collection.

Future research in this area will include looking for different functionals which may be of interest to researchers, considering fusion systems with greater than two-class label sets as the end result, and robustness of classifiers and fusors. Also, more research must be done in lessening the assumption of smoothness in the ROC curve since many ROC curves can only be approximated.

## References


[1] S. M. Lane, *Categories for the Working Mathematician, Second Edition*, Springer, New York, 1978.

[2] C. McClarty, *Elementary Categories, Elementary Toposes*, Oxford University Press, New York, 1992.

[3] J. Adámek, H. Herrlich, and G. Strecker, *Abstract and Concrete Categories*, John Wiley and Sons, Inc, New York, 1990.

[4] F. W. Lawvere and S. H. Schanuel, *Conceptual Mathematics, A First Introduction to Categories*, Cambridge University Press, Cambridge, 1991.

[5] D. L. Hall and J. Llinas, *Handbook of Multisensor Data Fusion*, CRC Press, Boca Raton, FL, 2001.

[6] S. G. Alsing, *The Evaluation of Competing Classifiers*. Ph.D. dissertation, Air Force Institute of Technology, Wright-Patterson AFB, OH, March 2000.

[7] I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Dover, Mineola, NY, 2000.