2011 50th IEEE Conference on Decision and Control and
European Control Conference (CDC-ECC)
Orlando, FL, USA, December 12-15, 2011

Mixed-Initiative Nested Classification by Optimal Thresholding

Baro Hyun, Mariam Faied, Pierre Kabamba, Anouck Girard

Abstract— **The purpose of this paper is to demonstrate that having two classifiers, a trichotomous classifier (true, false, or *unknown*) with workload-independent performance that turns over the data classified as unknown to a binary classifier (true or false) with workload-dependent performance, gives superior classification performance (lower probability of misclassification) compared to a single dichotomous classifier. We relate the classifier's performance to the inherent difficulty of the classification task at hand (classifiability), and compare the performance of different classifiers.**

I. Introduction

We consider a team composed of a workload-independent trichotomous classifier and a workload-dependent dichotomous classifier (*mixed-initiative* team). The team is structured in a *nested* architecture, that is: first the primary, workload-independent, trichotomous classifier examines the classification task, and if the task is classified as unknown, the secondary, workload-dependent, dichotomous classifier is called upon.

A. Overview

Classification is an act of allocating an entity into a category on the basis of its properties [1]. Both humans and machines are capable of making classification decisions under various circumstances. For example, in military operations such as the Air Force's Intelligence, Surveillance, and Reconnaissance (ISR) mission, human operators make classification decisions by inspecting some visual data (is the object of interest in the photo a threat or not?) while machine classifiers also make such decisions by intelligent algorithms such as Automatic Target Recognition (ATR) [2].

Humans and machines in such decision making, however, show different characteristics. As machines are generally good at repetitive tasks, the performance of machine classifiers is relatively consistent compared to that of humans. On the other hand, the performance of human operators can be affected by numerous external or internal factors such as workload [3]. However, humans at the peak of their performance generally outperform machines because humans are generally better at recognizing patterns than machines.

Classification performance is not only dependent on the classifier's ability, but also on the difficulty of the task at hand. Therefore, it is important to account for an inherent measure of difficulty and relate such a measure with the performance of a classifier.

The research was supported in part by the United States Air Force grant FA 8650-07-2-3744.

The authors are with the department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, USA {bhyun, mfaieda, kabamba, anouck}@umich.edu

In this study, we propose a novel architecture that utilizes a workload-independent and a workload-dependent classifier team by recognizing the complementary aspects of the two heterogeneous classifiers. We begin with studying the mechanisms of machine classifiers by formalizing a problem of thresholding. We study two cases: when a classifier has two options (dichotomous) or three options (trichotomous); and provide both analytical solutions and numerical results. Then, we introduce the concept of classifiability, which denotes the inherent difficulty of the classification task at hand, quantified by the optimal performance of a dichotomous classifier, and provide numerical examples relating classifiability to different classifier's performance. Finally, we formulate a nested classification architecture that utilizes a mixed-initiative team, provide numerical examples on the performance, and compare it to that of a single machine classifier.

B. Literature review

The problem of classification has attracted much attention, from the statistics to the robotics communities.

A classifier is a subject able to perform classification tasks. A good classifier is able to recognize the properties of an entity (pattern recognition, novelty detection), knows how to extract the key information among the properties (data analysis), and is consistent in its performance given the same information (consistency). In [4], the complementary abilities in data analysis of humans and computers are discussed with, as an example, the game of chess. The author discusses the definition of intelligent data analysis, such as pattern recognition, and unintelligent ones such as over-refined data schemes, and points out that the quality of classification is determined by two aspects: how likely the classification is to be incorrect, and how quickly the result is obtained.

In statistics, the problem of classification typically appears as a regression problem. For standard regression methodologies, see [5] and the references therein. In [1], the problem of static classification is posed and the solution is investigated analytically.

Thresholding is a particular method of classification and is ubiquitous in many fields that range from statistics, decoding theory, and image processing [6], [7], [8], [9]. For a thorough review on the state-of-the-art image thresholding techniques, see [6] and the references therein.

Research over the past few decades has suggested a number of statistical operator models for certain types of tasks [10], [11], [12], [13]. The Yerkes-Dodson law [14] relates the arousal and performance of a human. The pioneering work of [3] suggests a model of the human decision-

maker, which is compared to experimental results for human subjects performing a task at a computer-graphics terminal and the notion of workload as the control variable for supervision. This work suggests a definition of mental workload, which can vary person-by-person, and a notion of expert-novice operators differing by their productivity.

In [15], [16] a human operator is considered as a state-dependent queueing process where the state is a task arriving at a deterministic rate and optimal control policies are provided such that the queue does not overflow. A number of human-machine interaction strategies are proposed in [17] for a single operator-multiple heterogeneous vehicles scenario. In [18], a decision support system for sequential visual search tasks is presented while the effectiveness of the system is validated by human-subject experiments. It is shown that the human operator performance improves under the decision support system with automated algorithmic aids.

While much work has been done on the problem of classification, there has been relatively less attention given to the study of classification structures, utilizing multiple heterogeneous classifiers with different characteristics, from the standpoint of controls.

C. Original contributions

The original contributions of this work are as follows:

i. We propose a novel classifier architecture that uses a trichotomous classifier with workload-independent performance that turns over the data classified as unknown to a binary classifier with workload-dependent performance.
ii. We demonstrate that the novel classifier architecture gives superior classification performance (the probability of misclassification) compared to a single dichotomous classifier.
iii. We relate the classifier's performance to the inherent difficulty of the classification task at hand (classifiability), and compare the performances between different classifiers.

D. Organization

The organization of the paper is summarized as follows. First we introduce the theoretical background of this work in Sec. II. The problem of thresholding is formalized in Sec. III. In Sec. IV, we introduce classifiability and relate the classifiability to the classifier's performance. In Sec. V, we formalize mixed-initiative nested classification and provide numerical solutions. The conclusion and future work are provided in Sec. VI.

II. Theoretical background

A. Classifiers

A decider D is a deterministic mapping defined on a set of data into truth values, i.e.,

$$D : \{\text{data}\} \to \{T, F\}.$$

A classifier C is a decider with the domain of the mapping being a specific realization of a random variable. While both decider and classifier are deterministic mappings, the difference between them is that the latter accounts for the *randomness* of the data being classified.

Processing of the data requires two abilities: recognizing truth out of truth (rate of true positives) and falsehood out of falsehood (rate of true negatives). These abilities are characterized by two independent parameters, σ_T and σ_F, respectively. Note that these parameters are entries in the confusion matrix in signal detection theory [19].

B. Probabilistic modeling

Collecting information and making classification decisions are generally based on some measurements, and these measurements are typically obtained through imperfect sensors. Since these imperfect sensors introduce uncertainties in the measurement, e.g., sensor noise, the characteristics of the measurements are random. Therefore, we use probabilistic modeling, rather than deterministic, to investigate the relationship between information and classification performance.

Let X be a discrete random variable that denotes the category of objects of interest that can take two realizations: either T or F.[1] There is a probability associated to the event that X be one of the realizations, given as

$$P(X = T) = u,\ P(X = F) = 1 - u, \tag{1}$$

where $u \in [0,\ 1]$. We denote u as the *prior probability* and it represents the proportion of T objects among the objects of interest.

Let Y be a discrete random variable that denotes the object property that can take two realizations $Y \in \{Y_1, Y_2\}$. Y_1 represents the sensor measuring a property from an F object while Y_2 represents the sensor measuring a property from a T object. For instance, Y_2 can be the profile of a gun from a picture taken from the broadside view of a tank (threat) while Y_1 can be the wheels or the windshield from a picture taken from an automobile (friend).

Note that the number of realization of Y can be more than two, but we restrict our modeling for simplicity and clarity.

The likelihood of the object property given the object category is modeled by conditional probabilities. For two-option object categories and two-option object properties, the conditional probabilities are given as,

$$\begin{aligned} P(Y = Y_1 | X = F) &= \sigma_F, \\ P(Y = Y_2 | X = T) &= \sigma_T, \\ P(Y = Y_2 | X = F) &= 1 - \sigma_F, \\ P(Y = Y_1 | X = T) &= 1 - \sigma_T, \end{aligned} \tag{2}$$

where $\sigma_F, \sigma_T \in [0.5,\ 1]$ parameterize the conditional probabilities. When $\sigma_{(\cdot)} = 0.5$ the sensor is as bad as a pure guess, while when $\sigma_{(\cdot)} = 1$ the sensor is perfect. Note that the range $\sigma_{(\cdot)} \in [0,\ 0.5]$ describes the same phenomenon as $\sigma_{(\cdot)} \in [0.5,\ 1]$, but in a perverse manner.

[1] Note that "T" and "F" can be interpreted as "True" and "False", respectively, or as "Threat" and "Friend". The subsequent theory does not require choosing an interpretation.

C. Maximum likelihood classification

The maximum likelihood classification, also known as *likelihood-ratio rule* [20], is a decision rule based on posterior probabilities.

Definition 1. Bayes' rule
Bayes' rule gives the posterior probability of X given Y. For instance, given $Y = Y_1$ the posterior probability of $X = T$ is

$$P(X = T|Y = Y_1) = \frac{P(Y = Y_1|X = T)P(X = T)}{P(Y = Y_1)}. \quad (3)$$

Note that $P(Y = Y_1)$ can be computed by following the theorem of total probability [21].

Definition 2. Likelihood-ratio rule
Let $O_s \in \{T, F\}$ be a decision variable that follows the likelihood-ratio rule, i.e.,

$$O_s = \begin{cases} T & \text{if } \frac{P(X=T|Y=Y_1)}{P(X=F|Y=Y_1)} > 1 \\ F & \text{if } \frac{P(X=T|Y=Y_1)}{P(X=F|Y=Y_1)} \leq 1. \end{cases} \quad (4)$$

Let $f_{Y_0} \in [0, \infty]$ denote the ratio of the posterior probabilities such that,

$$f_1 = f_{Y=Y_1} = \left(\frac{1-\sigma_T}{\sigma_F}\right)\left(\frac{u}{1-u}\right), \quad (5a)$$

$$f_2 = f_{Y=Y_2} = \left(\frac{\sigma_T}{1-\sigma_F}\right)\left(\frac{u}{1-u}\right). \quad (5b)$$

Let $\delta_{O_{s0}} : \mathscr{R} \to \{0, 1\}$ such that

$$\delta_T(f) = \delta_{O_s=T}(f) = \begin{cases} 1 & \text{if } f > 1 \\ 0 & \text{if } f \leq 1, \end{cases} \quad (6a)$$

$$\delta_F(f) = \delta_{O_s=F}(f) = \begin{cases} 1 & \text{if } f \leq 1 \\ 0 & \text{if } f > 1. \end{cases} \quad (6b)$$

Then, the conditional probabilities of O_s given Y are,

$$P(O_s = T|Y = Y_2) = \delta_T(f_2), \quad (7a)$$

$$P(O_s = T|Y = Y_1) = \delta_T(f_1), \quad (7b)$$

$$P(O_s = F|Y = Y_2) = \delta_F(f_2), \quad (7c)$$

$$P(O_s = F|Y = Y_1) = \delta_F(f_1). \quad (7d)$$

D. Probability of misclassification

The classification performance is quantified by the probability of misclassification. Specifically, the lower the probability of misclassification, the better the classification performance.

Definition 3. Probability of misclassification
The probability of misclassification is the sum of probabilities of two faulty outcomes: false positive and false negative.

$$P_m^1 = P(O_s = T \wedge X = F) + P(O_s = F \wedge X = T) \quad (8)$$

Although we have considered the generic case of equal weights for the two outcomes, there can be different weights associated with the outcomes depending on the strategic objective of the classifier.

Assessing the probability of misclassification yields

$$\begin{aligned} P_m^1 = {} & P(O_s = T \wedge X = F|Y = Y_2)P(Y = Y_2) \\ & + P(O_s = T \wedge X = F|Y = Y_1)P(Y = Y_1) \\ & + P(O_s = F \wedge X = T|Y = Y_2)P(Y = Y_2) \\ & + P(O_s = F \wedge X = T|Y = Y_1)P(Y = Y_1), \end{aligned} \quad (9)$$

by the theorem of total probability. Assuming that the classification is unbiased, we can relax the expression by conditional independence, i.e., $P(O_s = O_{s0} \wedge X = X_0|Y = Y_0) = P(O_s = O_{s0}|Y = Y_0) \cdot P(X = X_0|Y = Y_0)$. Substituting Eq. (7) yields,

$$\begin{aligned} P_m^1 = {} & P(O_s = T|Y = Y_2)P(X = F \wedge Y = Y_2) \\ & + P(O_s = T|Y = Y_1)P(X = F \wedge Y = Y_1) \\ & + P(O_s = F|Y = Y_2)P(X = T \wedge Y = Y_2) \\ & + P(O_s = F|Y = Y_1)P(X = T \wedge Y = Y_1). \end{aligned}$$

Finally,

$$\begin{aligned} P_m^1 = {} & \delta_T(f_2)(1-\sigma_F)(1-u) + \delta_T(f_1)\sigma_F(1-u) \\ & + \delta_F(f_2)\sigma_T u + \delta_F(f_1)(1-\sigma_T)u. \end{aligned} \quad (10)$$

E. Yerkes-Dodson law

Unlike machine classifiers, human operator performance is subject to various human factors, such as workload, fatigue, boredom, stress, etc. Here, we model the human as a workload-dependent classifier. The workload-dependency is depicted by the Yerkes-Dodson law [14] that states that there is an optimal region of workload that allows humans to exhibit a maximum performance. Figure 1 illustrates the concept.

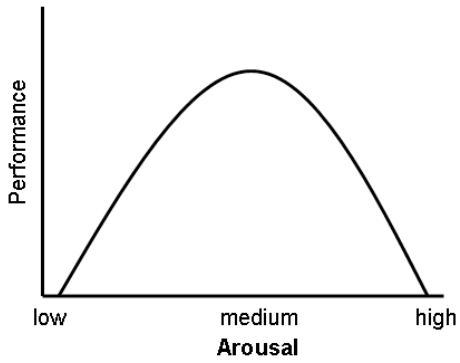


Fig. 1. Illustration of the Yerkes-Dodson law

Note that the Yerkes-Dodson law is not a definitive rule, meaning that depending on human subjects and situations, the performance-workload relationship may exhibit a different trend.

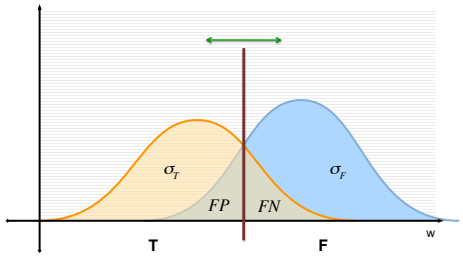
III. Problem formulation

A. The problem of thresholding

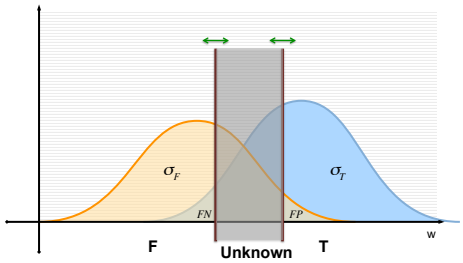
Assume that a property, $w \in \mathscr{R}$, can be measured from a population of objects of interest where the population comprises two disjoint sub-populations, T and F. Each subpopulation is characterized by its own distribution of w.

Assume that the two distributions are distinct such that if a proper threshold is applied, a classifier can distinguish one sub-population from another based on a measurement of w. Once a threshold is determined, measurement values on one side of the threshold are labeled as originating from a T object while properties on the other side are labeled as originating from an F object.

We consider two types of workload-independent classifiers: 1. the classification decision is based on two options (dichotomous), 2. the decision is based on three options (trichotomous). Figure 2 illustrates the concept of such classifiers.



(a) Dichotomous



(b) Trichotomous

Fig. 2. Concept of dichotomous and trichotomous thresholding

1) Dichotomous thresholding: We assume that the distribution of the measurable property w in each sub-population is a Gaussian probability density function (pdf),

$$p_T \sim \mathcal{N}(m_T, s_T^2), \tag{11a}$$
$$p_F \sim \mathcal{N}(m_F, s_F^2), \tag{11b}$$

where $m_{(\cdot)}$ is the mean and $s_{(\cdot)}$ is the standard deviation of the distribution. For the distinctness of the two distributions, we further assume that $m_T < m_F$ without loss of generality. Let $\tau \in \mathscr{R}$ be the threshold variable. For a classifier that uses thresholding, the rates of true positives and negatives are evaluated as:

$$\sigma_T = \int_{-\infty}^{\tau} a_T e^{-(w+b_T)^2/c_T^2} dw, \tag{12a}$$
$$\sigma_F = \int_{\tau}^{\infty} a_F e^{-(w+b_F)^2/c_F^2} dw, \tag{12b}$$

where $a_i = 1/\sqrt{2\pi s_i^2}$, $b_i = -m_i$, and $c_i = \sqrt{2s_i^2}$ with $i \in \{T,F\}$.

The cost function is the probability of misclassification, i.e.,

$$\begin{aligned} P_m^1 =& \delta_T(f_1)\sigma_F(1-u) + \delta_T(f_2)(1-\sigma_F)(1-u) \\ &+ \delta_F(f_1)(1-\sigma_T)u + \delta_F(f_2)\sigma_T u. \end{aligned} \tag{13}$$

The objective is to determine the optimal threshold that minimizes the probability of misclassification, i.e,

$$\min_{\tau} P_m^1.$$

2) Trichotomous thresholding: Dichotomous classification corresponds to the classical propositional logic where a proposition can either be *true* or *false*. Now, allowing a third status, trichotomous classification corresponds to ternary logic where a proposition can be *unknown* in addition to true and false. The reason we allow the unknown status is that there are cases when dichotomous machine classifiers are unsatisfactory. For example, the distributions of the sub-populations may not be easily distinguishable. Trichotomous classification can be formalized by extending the notion of dichotomous classification using two thresholds.

Let $\tau_1 \in \mathscr{R}$ and $\tau_2 \in \mathscr{R}$ be the threshold variables such that the cumulative probability distributions are

$$\sigma_T = \int_{-\infty}^{\tau_1} a_T e^{-(w+b_T)^2/c_T^2} dw, \tag{14a}$$
$$\sigma_F = \int_{\tau_2}^{\infty} a_F e^{-(w+b_F)^2/c_F^2} dw, \tag{14b}$$

with $\tau_1 \leq \tau_2$ where $a_i = 1/\sqrt{2\pi s_i^2}$, $b_i = -m_i$, and $c_i = \sqrt{2s_i^2}$ with $i \in \{T,F\}$. Let us define the range on w between the two thresholds where the classifier is unable to decide as *the region of indecision*, i.e., $[\tau_1, \tau_2]$.

Let P be a pre-specified probability of misclassification that is determined by the mission specification. The objective is to determine the optimal thresholds that minimize the size of the region of indecision, i.e.,

$$\min_{\tau_1,\tau_2} |\tau_2 - \tau_1|,$$

subject to an equality constraint,

$$P_m^1 = P.$$

IV. Classifiability

The fundamental difficulty of a classification task is determined by the nature of the distributions that are to be classified. Given two undistinguishable distributions, e.g., two Gaussian distributions with identical mean and variance, it is impossible to make the classifier's performance better than a pure guess because the task itself is *unclassifiable*. Recognizing this, we use the term *classifiability* to quantify the fundamental difficulty of the classification task at hand.

Definition 4. Classifiability
Classifiability is quantified as the reciprocal of the minimal probability of misclassification performed by a dichotomous classifier in a logarithmic scale, i.e.,

$$\text{Classifiability} = \log \frac{1}{P_m^{1*}}, \tag{15}$$

where $P_m^{1^*}$ denotes the minimal probability of misclassification of a dichotomous classifier. Note that the measure is the *information content* defined by Shannon [23].

Note that as the distance between the means increases, the classifiability of the task increases. On the other hand, if the distance between the means is zero ($|m_T - m_F| = 0$), the classifiability reaches its minimum, $\log\left(\frac{1}{0.5}\right)$.

V. Mixed-initiative nested classification

As stated in the introduction, we consider a mixed-initiative nested classification where two heterogeneous classifiers are composed in a nested architecture. Figure 3 shows the concept. We assume the following:

i. The workload-independent classifier and the workload-dependent classifier examine the task independently.
ii. The workload of the secondary classifier is determined by the probability of being called upon by the primary classifier.

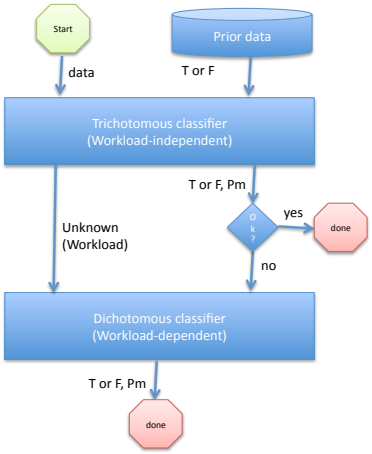


Fig. 3. Concept of mixed-initiative nested classification

3) Workload-independent trichotomous classifier: Let τ_1 and τ_2 be the threshold variables. Then, the cumulative probability distributions for Gaussian distributions are

$$\sigma_{T_1} = \int_{-\infty}^{\tau_1} a_T e^{-(w+b_T)^2/c_T^2} dw \tag{16a}$$

$$\sigma_{F_1} = \int_{\tau_2}^{\infty} a_F e^{-(w+b_F)^2/c_F^2} dw \tag{16b}$$

where $a_i = 1/\sqrt{2\pi s_i^2}$, $b_i = -m_i$, and $c_i = \sqrt{2s_i^2}$ with $i \in \{T, F\}$. Note that $\sigma_{T_1}, \sigma_{F_1} \in [0, 1]$.

The region of indecision of the primary trichotomous classifier, i.e., $[\tau_1, \tau_2]$, determines the workload applied to the secondary classifier. We define a workload variable, $W \in [0, 1]$, with 0 indicating idle and 1 indicating fully loaded. Let $f_i(w) = a_i e^{-(w+b_i)^2/c_i^2}$ with $i \in \{T, F\}$, then the workload variable is defined as

$$W = \int_{\tau_1}^{\tau_2} u f_T(w) + (1-u) f_F(w) dw. \tag{17}$$

Note that the range of W is $[0, 1]$ for any τ_1 and τ_2. We assume that the workload variable is normalized such that the maximum value is unity when the workload-dependent classifier is fully loaded.

4) Workload-dependent dichotomous classifier: The classification performance of a human operator is modeled as follows. Recognizing the concavity of the curve, we model the Yerkes-Dodson law (Fig. 1) as a quadratic function of the workload as,

$$\sigma_{T_2} = -(4\sigma_T^* - 2)W^2 + (4\sigma_T^* - 2)W + 0.5, \tag{18}$$

$$\sigma_{F_2} = -(4\sigma_F^* - 2)W^2 + (4\sigma_F^* - 2)W + 0.5, \tag{19}$$

where $\sigma_{(\cdot)}^* \in [0.5, 1]$ determines the maximum of $\sigma_{(\cdot)}$. If $\sigma_T^* = \sigma_F^*$, the peak performances of a human operator classifying true positives and true negatives are equally good.

5) Probability of misclassification for two classifiers: The probability of misclassification is a sum of contributions of two faulty outcomes: false positives and false negatives. By the theorem of total probability, the probability of misclassification includes all possible cases of misclassification by the two classifiers. Here we only state the formula and exclude the derivation due to space limitations. The probability of misclassification for two classifiers is

$$P_m^2 = \bar{\sigma}_1^T R_2 \bar{\sigma}_2, \tag{20}$$

where

$$\bar{\sigma}_i = \begin{bmatrix} \sigma_{F_i} & 1-\sigma_{F_i} & 1-\sigma_{T_i} & \sigma_{T_i} \end{bmatrix}^T, \; i = 1, 2$$

$$R_2 = \begin{bmatrix} \delta_T(f_{1,1})(1-u) & \delta_T(f_{1,2})(1-u) & 0 & 0 \\ \delta_T(f_{2,1})(1-u) & \delta_T(f_{2,2})(1-u) & 0 & 0 \\ 0 & 0 & \delta_F(f_{1,1})u & \delta_F(f_{1,2})u \\ 0 & 0 & \delta_F(f_{2,1})u & \delta_F(f_{2,2})u \end{bmatrix},$$

with

$$f_{1,1} = \left(\frac{1-\sigma_{T_1}}{\sigma_{F_1}}\right)\left(\frac{1-\sigma_{T_2}}{\sigma_{F_2}}\right)\left(\frac{u}{1-u}\right),$$

$$f_{1,2} = \left(\frac{1-\sigma_{T_1}}{\sigma_{F_1}}\right)\left(\frac{\sigma_{T_2}}{1-\sigma_{F_2}}\right)\left(\frac{u}{1-u}\right),$$

$$f_{2,1} = \left(\frac{\sigma_{T_1}}{1-\sigma_{F_1}}\right)\left(\frac{1-\sigma_{T_2}}{\sigma_{F_2}}\right)\left(\frac{u}{1-u}\right),$$

$$f_{2,2} = \left(\frac{\sigma_{T_1}}{1-\sigma_{F_1}}\right)\left(\frac{\sigma_{T_2}}{1-\sigma_{F_2}}\right)\left(\frac{u}{1-u}\right).$$

The global objective of the nested team architecture is to minimize the probability of misclassification by choosing the threshold variables for the primary trichotomous classifier, i.e.,

$$\min_{\tau_1, \tau_2} P_m^2,$$

subject to inequality constraints,

$$\sigma_{T_1} \geq 0.5, \tag{21a}$$

$$\sigma_{F_1} \geq 0.5, \tag{21b}$$

$$\sigma_{T_1} \leq 1, \tag{21c}$$

$$\sigma_{F_1} \leq 1, \tag{21d}$$

$$\tau_1 \leq \tau_2. \tag{21e}$$

This formalism allows the two classifiers to have the same goal, although the mechanism behind how each classifier functions is completely different. Also, due to the inequality constraints, the formulation does not allow the trichotomous classifier to experience perverse behavior, i.e., $\sigma_{(\cdot)} \in [0, 0.5]$.

6) Numerical solutions: We solve the optimization problem by using the MATLAB *fmincon* command. Figure 4 illustrates the performance comparison between the dichotomous classifier and the mixed-initiative nested classifiers with different initialization of the threshold variables shown in a logarithmic scale. Note that the search space has multiple local minima, so that depending on the initial condition the performance of the mixed-initiative nested classifiers can be different. It is clear that while the performance of the nested classifiers is sensitive to the initialization of the threshold variables, it is no worse than the dichotomous classifier regardless of the initialization as shown in Fig 4. Note that the performances of both dichotomous and mixed-initiative classifiers are linearly decreasing functions (in logarithmic scale) with respect to the classifiability.

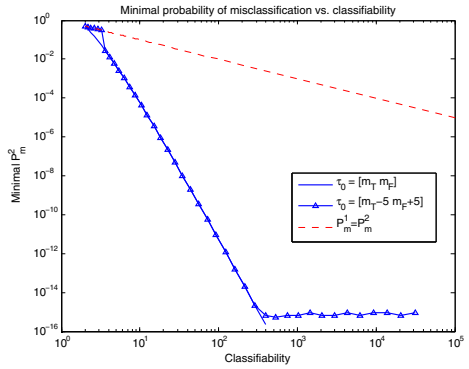


Fig. 4. Comparison of dichotomous and mixed-initiative thresholding performance

Investigating the analytical properties of the solution for mixed-initiative nested classifiers is left as future work.

VI. Conclusion & Future Work

In this paper, we have proposed a novel classifier architecture that uses a trichotomous classifier with workload-independent performance that turns over the data classified as unknown to a binary classifier with workload-dependent performance. We demonstrate that the novel classifier architecture gives superior classification performance (the probability of misclassification) than a single dichotomous classifier, relate the classifier's performance to the inherent difficulty of the classification task at hand (classifiability), and compare the performances between different classifiers.

So far, we have studied a case when a scalar measurable quantity w is provided. We expect to extend this work by introducing multi-dimensional measurable quantities, such that the decision variable for thresholding is no longer a choice of a scalar value, but a choice of a multi-dimensional surface. Additionally, as one of the assumptions that we made is that the knowledge of the distributions of w is provided by calibration, future work will address the case when the distributions are partially given. Finally, investigating the case when unreliable prior information is given is left as future work.

References

[1] S.S. Gupta and L.-Y. Leu, "On a classification problem: ranking and selection approach," *Technical Report #89-27C,* Department of Statistics, Purdue University, 1989.

[2] M. Pachter, P. Chandler, S. Darbha, "Optimal control of an ATR module equipped MAV-Human operator team," *Cooperative Control,* Edwin Elgar Publishers, Eds. Pardalos, Murphey, Grundel and Prokopyev, 2007.

[3] T.B. Sheridan, "Dynamic decisions and work load in multitask supervisory control," *IEEE Transaction on Systems, Man, and Cybernetics,* Vol. SMC-10, No. 5, May. 1980.

[4] D.J. Hand, "Intelligent data analysis: issues and opportunities," *Intelligent Data Analysis,* 2, 1998, p67-79.

[5] M.H. Kutner, C.J. Nachtsheim, J. Neter, W. Li, "Applied linear statistical models," *McGraw-Hill Irwin,* 5th edition, 2005.

[6] C.-I Chang, Y. Du, J. Wang, S.-M. Guo, P.D. Thouin, "Survey and comparative analysis of entropy and relative entropy thresholding techniques," *IEEE Proc. Visual Image Signal Process,* Vol. 153, No. 6, Dec. 2006.

[7] J. Kittler and J. Illingworth, "Minimum error thresholding," *Pattern Recognition,* Vol. 19, No. 1, p. 41-47, 1986.

[8] Y.-J. Wu, M.D. Alston, P.M. Chau, "Dynamic adaptation of quantization thresholds for soft-decision Viterbi decoding with reinforcement learning neural network," *Journal of VLSI Signal Processing,* Vol. 6, p. 77-84, 1993.

[9] P. Reynaud-Bouret and V. Rivoirard, "Near optimal thresholding estimation of a Poisson intensity on the real line," *Electronic Journal of Statistics,* Vol. 4, p. 172-238, 2010.

[10] W. Edwards, "Optimal strategies for seeking information: models for statistics, choice reaction times, and human information processing," *Journal of Mathematical Psychology,* 1965.

[11] M. Stone, "Models for choice-reaction time," *Psychometrika,* vol. 25, No. 3, Sep. 1960.

[12] P.M. Fitts, "Cognitive aspects of information processing: 3. Set for speed versus accuracy," *Journal of Experimental Psychology,* vol. 71, No. 6, 1966.

[13] R.W. Pew, "The speed-accuracy operating characteristics," *Acta Psychologica,* 30 Attention and performance 3, 1969.

[14] R.M. Yerkes and J.D. Dodson, "The relation of strength of stimulus to rapidity of habit-formation," *Journal of Comparative Neurology and Psychology,* vol. 18, pp.459-482,1908

[15] K. Savla, T. Temple and E. Frazzoli, "Human-in-the-loop vehicle routing policies for dynamic environments," *In IEEE Conference on Decision and Control,* Cancun, Mexico, pages 1145-1150, December 2008.

[16] K. Savla and E. Frazzoli, "A Dynamical Queue Approach to Intelligent Task Management for Human Operators," *Proceedings of the IEEE,* Submitted, 2010.

[17] C.E. Nehme, "Modeling human supervisory control in heterogeneous unmanned vehicle systems," Ph.D Dissertation, Department of Aeronautics and Astronautics, MIT, 2009.

[18] L.F. Bertuccelli, N.W.M. Beckers, M.L. Cummings, "Developing operator models for UAV search scheduling," *In Proceedings of AIAA Guidance, Navigation, and Control Exhibit,* Toronto, Canada, 2010.

[19] Scharf, L.L., "Statistical signal processing (detection, estimation, and time series analysis)," *Addison-Wesley,* 1991.

[20] A. Pete, K.R. Pattipati, and D.L. Kleinman, "Methods for fusion of individual decisions," In *Proceedings of the American Control Conference,* 1991.

[21] S. Thrun, W. Burgard and D. Fox, "Probabilistic Robotics," *The MIT Press* 2005.

[22] B. Hyun, P. Kabamba, M. Faied, A. Girard, "On the independence of information and classification performance," *IEEE Conference on Decision and Control,* Orlando, FL, 2011. Submitted.

[23] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.