Conf-931245--1

Fusion Rule Estimation in Multiple Sensor Systems with Unknown Noise Distributions †

Nageswara S.V. Rao
Center for Engineering Systems Advanced Research
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6364
nrao@plato.epm.ornl.gov

To be presented at
Indo-US Workshop on Distributed Signal and Image Integration Problems, December 16-18, 1993, Pune, India.

†Research sponsored by the Engineering Research Program of the Office of Basic Energy Sciences, of the U.S. Department of Energy, under Contract No. DE-AC05-840R21400 with Martin Marietta Energy Systems, Inc. In addition, this work is also partially funded by National Science Foundation under grant No IRI-9108610.

Fusion Rule Estimation in Multiple Sensor Systems with Unknown Noise Distributions †

Nageswara S.V. Rao
Center for Engineering Systems Advanced Research
Oak Ridge National Laboratory
Oak Ridge, Tennessee 37831-6364
nrao@plato.epm.ornl.gov

Abstract

A system of N sensors $S_1, S_2, \ldots, S_N$ is considered; corresponding to an object with parameter $x \in \Re^d$, sensor S_i yields output $y^{(i)} \in \Re^d$ according to an unknown probability distribution $p_i(y^{(i)}|x)$. A training l-sample $(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)$ is given where $y_i = (y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(N)})$ and $y_i^{(j)}$ is the output of S_j in response to input x_i. The problem is to estimate a fusion rule $f : \Re^{Nd} \mapsto \Re^d$, based on the sample, such that the expected square error

$$I(f) = \int [x - f(y^{(1)}, y^{(2)}, \ldots, y^{(N)})]^2 p(y^{(1)}, y^{(2)}, \ldots, y^{(N)}|x) p(x) dy^{(1)} dy^{(2)} \ldots dy^{(N)} dx$$

is to be minimized over a family of fusion rules Λ based on the given l-sample. Let $f_* \in \Lambda$ minimize $I(f)$; f_* cannot be computed since the underlying probability distributions are unknown. Using Vapnik's empirical risk minimization method, we show that if Λ has finite capacity, then under bounded error, for sufficiently large sample, f_{emp} can be obtained such that

$$P[I(f_{emp}) - I(f_*) > \epsilon] < \delta$$

for arbitrarily specified $\epsilon > 0$ and δ, $0 < \delta < 1$. We identify several computational methods to obtain f_{emp} or its approximations based on neural networks, radial basis functions, wavelets, non-polynomial networks, and polynomials and splines. We then discuss linearly separable systems to identify objects from a finite class where f_{emp} can be computed in polynomial time using quadratic programming methods.

†Research sponsored by the Engineering Research Program of the Office of Basic Energy Sciences, of the U.S. Department of Energy, under Contract No. DE-AC05-840R21400 with Martin Marietta Energy Systems, Inc. In addition, this work is also partially funded by National Science Foundation under grant No IRI-9108610.

1 Introduction

It has been realized by many researchers that there are fundamental limitations on the capabilities of single sensor systems in a number of application areas such as robotics (see special issue on sensor data fusion edited by Brady [6] and the book by Abidi and Gonzalez [1]). Diverse information from many different sensors can often be used to overcome the limitations of a single sensor through coordinated interpretation, where the information from the individual sensors is to be suitably "fused". Research efforts on several aspects of theory and application of data fusion methods are very extensive; a recent survey could be found in Luo and Kay [24]. Also in some systems, a number of sensors of same kind are employed for fault tolerance, and in several others multiple sensors are required to achieve the given task.

A successful operation of a multiple sensor system critically depends on the methods that combine the outputs of the sensors. The problem of obtaining a fusion rule has been the focus of extensive research over the past decades. In general, the design, implementation and computational issues in multiple sensor systems are considerably more challenging than their counterparts in single sensor systems, for the issues due to the distributed information processing are seemingly absent in the latter (Tsitsiklis and Athans [45]). A number of issues related to this problem have been studied under the framework of distributed sensor networks (Barnett [2], Lesser et al. [23], Wesson et al. [50]); see Abidi and Gonzalez [1] and Iyengar et al. [19] and references therein for some recent works. Comprehensive treatments on specialized topics such as spatial reasoning (Kak and Chen [20]) also exist.

We consider the problem of inferring a suitable rule using training examples, where the errors introduced by various individual sensors are unknown and not controllable, e.g. a robot system equipped with sensors. Here, the choice of the sensors has been made and the system is available, and the fusion rule for the system has to be obtained; this aspect has to be contrasted with the general areas of team decision problems (e.g. Radner [31]) and distributed detection (e.g. Tsitsiklis and Athans [45]) where the individual elements as well as the fuser are to be designed to achieve an overall goal. This paradigm is applicable to several existing robot systems equipped with sensors whose error probabilities are either unknown or hard to estimate, but, several objects with known features can be sensed using the system. If the sensor error distributions are known, several cases of this problem have been solved. Some of the earlier work in this direction is due to Chow [10]. This problem is also related to the group decision models studied extensively in political economy (for example see Grofman and Owen [15]); some of the early majority methods of combining the outputs of probabilistic Boolean elements date back to 1786 under the name of Condorcet jury models. The distributed detection problem based on probabilistic formulations have been extensively studied (Chair and Varshney [8], Demirbas [13], Reibman and Nolte [37], Sadjadi [39], Tenney and Sandell [41], Thomopoulos et al. [43, 44]); see Thomopoulos [42] for a summary of a number of results on this problem. Many of these sensor integration techniques are based on maximizing the a posteriori probabilities of hypotheses under a suitable probabilistic model. However, in situations where the probability distributions are unknown or difficult to estimate (or compute) such methods are ineffective. One alternative is to estimate the distribution based on a sample; as illustrated in general by Vapnik [47], sometimes (depending on the classes of probability distributions) the problem of estimating the distributions is more difficult than the subsequent problem of estimating a dependence (such as a pattern classification rule). This property holds for several pattern recognition and regression estimation problems [47].

We consider a general framework based on a multiple sensor system with unknown sensor noise characteristics. But the system is available so that readings corresponding to known parameters can be obtained. In this context, we address the problem of inferring the fusion rule based on a set of training data with only a limited assumptions made on the noise. We formulate this problem as a special case of empirical risk minimization problem of Vapnik [47]. We show that a fusion rule close to "optimal rule" can be inferred under the condition of finite capacity of the fusion rules and boundedness of the errors. This result is existential in that the computational problem of computing such approximation rule could be computationally intractable. Then we illustrate a system for which the computational problem is polynomial-time solvable. From a learning point of view, the paradigm of training to obtain a fusion rule has been applied in the context of neural networks (for example see Huntsberger [18]), but no performance guarantees based on finite samples are available. We are interested in obtaining the bounds for the error based on a finite sample so that system can be trained with the required number of examples.

Consider a system of N sensors $S_1, S_2, \ldots, S_N$ such that corresponding to an object with parameter $x \in \Re^d$, sensor S_i yields output $y^{(i)} \in \Re^d$ according to an unknown probability distribution $p_i(y^{(i)}|x)$. A training l-sample $(x_1, y_1), (x_2, y_2), \ldots, (x_l, y_l)$ is given where $y_i = (y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(N)})$ and $y_i^{(j)}$ is the output of S_j in response to input x_i. The problem is to estimate a fusion rule $f : \Re^{Nd} \mapsto \Re^d$, based on the sample, such that $f(y^{(1)}, y^{(2)}, \ldots, y^{(N)})$ "closely" approximates x. More precisely, we consider the expected square error

$$I(f) = \int [x - f(y^{(1)}, y^{(2)}, \ldots, y^{(N)})]^2 p(y^{(1)}, y^{(2)}, \ldots, y^{(N)}|x) p(x) dy^{(1)} dy^{(2)} \ldots dy^{(N)} dx \quad (1.1)$$

which is to be minimized over a family of fusion rules Λ based on the given l-sample. Let $f_* \in \Lambda$ minimize $I(f)$; f_* cannot be computed since the underlying probability distributions are unknown. Furthermore, since no restrictions are placed on the underlying distributions, it will not be possible to to infer f_* (with probability one) based on a finite sample.

Now consider that the empirical estimate

$$I_{emp}(f) = \frac{1}{l} \sum_{i=1}^{l} [x_i - f(y_i^{(1)}, y_i^{(2)}, \ldots, y_i^{(N)})]^2 \quad (1.2)$$

is minimized by $f = f_{emp} \in \Lambda$. Using Vapnik's empirical risk minimization method, we show that if Λ has finite capacity, then under bounded error, or bounded relative error for sufficiently large sample

$$P[I(f_{emp}) - I(f_*) > \epsilon] < \delta \quad (1.3)$$

for arbitrarily specified $\epsilon > 0$ and δ, $0 < \delta < 1$. Thus this approach yields a fusion rule whose "error" is bounded by an arbitrarily specified precision ϵ with arbitrarily specified confidence $1 - \delta$ given sufficiently large sample. We estimate the required sample size in terms of ϵ and δ and the parameters of Λ. This estimation directly follows from the principle of empirical risk minimization of Vapnik [47].

The problem of computing the required hypothesis f_{emp} is computationally hard for several cases of Λ. For example, if Λ is the set of feedforward neural networks, the corresponding computational problem is called the *loading problem* which in NP-hard (Blum and Rivest [4]). Even for much simpler cases of a single sensor with Boolean output, several of these computational problems are NP-complete (Pitt and Valiant [29]). Also, since no restrictions are placed on the information of the sensors, even if there is no probabilistic error in the sensors, several multisensor fusion problems are NP-complete (Tsitsiklis and Athans [45], Rao [32], Rao et al. [33]). We then identify several approximation methods for computing f_{emp} based on neural networks, radial basis functions, wavelets, non-polynomial networks, and polynomials and splines.

We then consider a special class of linearly separable systems, where the associated computational problem can be solved (exactly) as a quadratic programming problem with positive semidefinite constraint matrix; hence this problem can be solved in polynomial time.

Because of the distribution-free nature of the results, this work is related to the Probably and Approximately Correct (PAC) learning results (Valiant [46], Natarajan [26], Blumer et al. [5]). In this vein, the present problem is a generalization of the N-learners problem that deals with the problem of combining a system of PAC learners (Rao et al. [35], Rao and Oblow [34]).

The organization of this paper is as follows. We present some preliminaries in Section 2. In Section 3, we propose a solution method based on the empirical risk minimization methods of Vapnik [47]. Then, in Section 4, we discuss the class of linearly separable systems.

2 Preliminaries

In this section, we present some basic definitions from Vapnik [47]. For family $\{A_\gamma\}_{\gamma\in\Gamma}$, $A_\gamma \subseteq A$, and for a finite set $\{a_1, a_2, \ldots, a_n\} \subseteq A$ we define

$$\Pi_{\{A_\gamma\}}(\{a_1, a_2, \ldots, a_n\}) = \{\{a_1, a_2, \ldots, a_n\} \cap A_\gamma\}_{\gamma\in\Gamma}.$$

We maximize this quantity with respect to the set $\{a_1, a_2, \ldots, a_n\}$ to obtain

$$\Pi_{\{A_\gamma\}}(n) = \max_{a_1, a_2, \ldots, a_n} |\Pi_{\{A_\gamma\}}(\{a_1, a_2, \ldots, a_n\})|.$$

The following is critical identity established in [47].

$$\Pi_{\{A_\gamma\}}(n) = \begin{cases} 2^n & \text{if } n \leq h \\ < 1.5\frac{n^h}{h!} & \text{if } n > h \end{cases}$$

Notice that for a fixed h, the right hand side increases exponentially with n until it reaches h and then varies as a polynomial in n with fixed power h. This quantity h is called the *Vapnik-Chervonenkis dimension* of the family of sets A_γ; it can also be alternatively defined as the largest size h of a set $\{a_1, a_2, \ldots, a_n\} \subseteq A$ that can be subdivided in all possible ways into two classes by means of sets A_γ. Formally,

$$VC\_dim(\{A_\gamma\}) = \max_n \{\Pi_{\{a_1, a_2, \ldots, a_n\}}(n) = 2^n\}.$$

The VC_dim plays a very critical role in the convergence of empirical measures of sets to their actual measures in that its finiteness is both necessary and sufficient for the convergence. This property has been extensively used in PAC learning in various learnability results [5].

For a set of functions, the *capacity* is defined as the largest number h of pairs (x_i, y_i) that can be subdivided in all possible ways into two classes by means of rules of the form

$$\{\Theta[(x - f(y,\alpha))^2 + \beta]\}_{\alpha,\beta}$$

where

$$\Theta(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{f } z < 0 \end{cases}$$

Formally, the capacity of $\{f(y,\alpha)\}_{\alpha\in\Lambda}$ is the Vapnik-Chervonenkis dimension of the set of indicator functions

$$\{\Theta[(x - f(y,\alpha))^2 + \beta]\}_{(\alpha,\beta)\in\Lambda\times\Re}.$$

In terms of the convergence of empirical measures to expectations, the capacity of set of functions plays a role similar to that of the Vapnik-Chervonenkis dimension for family of sets.

3 Empirical Risk Minimization

In this section we summarize various results from Vapnik [47] that yield solutions to several formulations of the fusion rule estimation problem. To solve for a rule $f_* \in \Lambda$ that minimizes the expected error in (1.1), we instead minimize the empirical error in (1.2) to obtain a best empirical estimate f_{emp}. The closeness of f_{emp} to f_* is specified by the parameters *precision* ϵ and *confidence* δ in condition (1.3) referred to as the (ϵ, δ)-condition; condition (1.3) can also be written as

$$P[I(f_{emp}) - I(f_*) < \epsilon] > 1 - \delta.$$

In order to ensure the (ϵ, δ)-condition, two types of conditions are to be satisfied [47]:

(a) the capacity of $\{f_\alpha\}_{\alpha\in\Lambda}$ must be bounded;

(b) the error $I(.)$ must be bounded, i.e., $\sup_{x,y,\alpha}(x - f(y,\alpha))^2 \leq \tau$ or the relative error must be bounded as follows for some $p > 1$

$$\sup_{\alpha} \frac{\left[\int (x - f(x,\alpha))^{2p} P(x,y)dxdy\right]^{1/p}}{\int (x - f(x,\alpha))^2 P(x,y)dxdy} < \tau.$$

First we illustrate a very simple case where both x and f_α take values from $\{0,1\}$.

Theorem 3.1 *Consider that x and f_α take values from $\{0,1\}$.*

(i) Given l samples, we have

$$P\left[I(\alpha_{emp}) - I(\alpha^*) > 2\kappa\right] < 9\frac{(2l)^h}{h!}e^{-\kappa^2 l/4}$$

where h is the capacity of Λ.

(ii) *If the hypothesis space is finite in that* $\{f(y,\alpha)\} = \{f(y,\alpha_1), f(y,\alpha_2), \ldots f(y,\alpha_M)\}$, *given* l *examples, we have*

$$P\left[I(\alpha_{emp}) - I(\alpha^*) > 2\kappa\right] < 2Me^{-2\kappa^2 l}.$$

Parts (i) and (ii) of this theorem directly follows from Theorem 6.1 and 6.7 of Vapnik [47] respectively. This special case deals with PAC learning formulation, which is also referred to as the pattern recognition problem in [47]. The bounds of this theorem can be sharpened in several cases of PAC learning (see Blumer et al. [5]). In Part (i), notice that the upperbounds on the right hand side are products of two main factors: first one is l^h and the second one is $e^{-\kappa^2 l/4}$. For a fixed value of h, the latter will be decreasing with the sample size l; if l is chosen large enough the right hand side can be made equal to δ.

An example of infinite hypothesis class can be given by the set of all neural net works with a fixed number of nodes, where $f(y,\alpha)$ stands for a feedforward neural network with connection weight vector α.

We now consider the general case. First we consider that the error is bounded by τ in Theorem 3.2, and in Theorem 3.3 the relative error is bounded.

Theorem 3.2 *Consider that the error is bounded as* $\sup_{x,y,\alpha}(x - f(y,\alpha))^2 \le \tau$.

(i) *Then given* l *examples, we have*

$$P[I(\alpha_{emp}) - I(\alpha^*) \ge 2\tau\kappa] \le 9\frac{(2l)^h}{h!}e^{-\kappa^2 l/4}.$$

Equivalently, with probability at least $1-\eta$, *we have*

$$I(\alpha_{emp}) \le I(\alpha^*) + 4\tau\sqrt{\frac{h\left(\ln\frac{2l}{h}+1\right) - \ln\frac{\eta}{9}}{l}}.$$

(ii) *If the hypotheis space is finite in that* $\{f(y,\alpha)\} = \{f(y,\alpha_1), f(y,\alpha_2), \ldots f(y,\alpha_M)\}$. *Then given* l *examples, we have*

$$P\left[I(\alpha_{emp}) - I(\alpha^*) > 2\tau\kappa\right] < 18Mle^{-\kappa^2 l/4}.$$

The Part (i) and (ii) of this theorem directly follow from Theorem 7.1 and 7.3 of Vapnik [47] respectively.

Theorem 3.3 *Consider that the relative error be bounded such that for some* $p > 1$ *we have*

$$\sup_{\alpha}\frac{\left[\int(x - f(x,\alpha))^{2p}P(x,y)dxdy\right]^{1/p}}{\int(x - f(x,\alpha))^2 P(x,y)dxdy} < \tau.$$

(a) *If* $p > 2$, *we have*

$$P\left\{\frac{I(\alpha) - I_{emp}(\alpha)}{I(\alpha)} > \tau a(p)\kappa\right\} < 24le^{-\kappa^2 l/4}$$

where

$$a(p) = \left[\frac{(p-1)^{p-1}}{2(p-2)^{p-1}}\right]^{1/p}.$$

(b) If $1 < p \leq 2$, we have

$$P\left\{\frac{I(\alpha) - I_{emp}(\alpha)}{I(\alpha)} > \tau V_p(\kappa)\right\} < 24le^{-\kappa^2 l^{2-(2/p)}/4}$$

where

$$V_p(\kappa) = \kappa\left[1 - \frac{\ln\kappa}{p^{1/(p-1)}(p-1)}\right]^{\frac{p-1}{p}}.$$

This theorem directly follow from Theorem 7.6 of Vapnik [47].

Note that the results of this section are mainly existential in nature; they do not yield computational methods to either represent f_α or to compute the required α_{emp}; this problem is addressed in the next section. However, they provide very strict guidelines for the conditions under which this empirical estimation procedure is a viable option.

3.1 Computational Problems

Now the basic computational problem of last section is to solve for α_{emp} satisfying

$$I_{emp}(\alpha_{emp}) = \min_{\alpha \in \Lambda}\{\frac{1}{l}\sum_{i=1}^{l}[x_i - f(y_i, \alpha)]^2\}.$$

We now consider finitely representable hypothesis classes, where $\Lambda = \Re^d$ or $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_d)$. The hypothesis classes too numerous to be so represented are unlikely to be of use in practical implementations.

We now briefly discuss some existing methods that are applicable to several classes of Λ. In most cases, we fix a class of Λ and then identify a "network" structure with "connection weights" vector that corresponds to α. Typically α_{emp} is obtained using an incremental algorithm that starts with some initial weight vector and updates it in response to the accuracy of the present weight. This approach is motivated by a wide availability of results on such networks (neural networks being some of the most popular ones) and their incremental algorithms. Typically the problem of computing an exact solution α_{emp} is computationally intractable, but many practically useful implementations of the incremental algorithms are available.

(a) *Feedforward neural networks:* Each f_α is represented by an artificial neural network with at least one hidden layer and with a finite number of nodes. As shown in Cybenko [12], such networks can approximate continuous functions with arbitrary levels c precision (see also Hornik [16], Hornik et al. [17], Barron [3]). Here α corresponds the connection weight vector. The problem of computing α_{emp} is NP-hard in general. An approximation to α_{emp} can be computed using the well-known backpropagation algorithm (Werbos [49] and Rumelhart et al. [38]). Convergence properties of such algorithm have been studied by White [51] (also see White [52]), Stankovic and Milsosavljevic [40], Finnoff [14] and Nedeljkovic [27].

(b) *Radial basis functions:* The radial basis networks with suitably chosen non-linear hidden layers (Broomhead and Lowe [7], Chen et al. [9]) can be used in the computation of f_{emp}; also there are a number of learning algorithms that can be applied in this case.

(c) *Wavelet-based expansion:* Wavelets and related methods have become the focus of attention of a number of researchers (Chui [11]). Zhang and Benveniste, [53] proposed networks of wavelets (in a manner analogous to neural network) which can approximate arbitrary continuous maps; each network is characterized by a finite real vector that corresponds to the dilation and translation operations. They also propose an algorithm similar to the back propagation algorithm that can be used to compute an approximation to α_{emp}.

(d) *Networks of non-polynomial Units:* In a general treatment, Leshno et al. [22] showed that finite networks of non-polynomial units can be used to approximate the arbitrary continuous maps. Although no algorithms to compute the required connection weights are available, backpropagation style algorithms can be designed in several cases.

(e) *Classes of polynomials and splines:* The classes of polynomial and splines have been originally considered by Vapnik [47]. The approximation properties of these classes are fairly limited (Powell [30]), but in cases where Λ is adequately represented by these classes, these methods can be effective. Note that the networks based on these classes are do not satisfy the approximation properties enjoyed by the other above classes (Leshno et al. [22]).

In some cases, it is possible to compute an approximation of f_* directly using a stochastic approximation algorithm so that the (ϵ, δ) condition is satisfied (Rao et al. [36]).

4 Linearly-Separable Systems

Consider a detection system consisting of finite set of objects $O = \{O_1, O_2, \cdots, O_n\}$. When an object is detected in the workspace, each sensor outputs a random vector which corresponds to an "error-free" vector corrupted by noise of an unknown probability distribution. Given a vector of sensor readings y correspondings to an unknown object, we are required to identify the object. Note that this is a special case of formulation of last section where x takes n distinct values.

We call the detection system (O, S) to be *linearly separable* if for each $O_i \in O$, there corresponds a known interval $[a^{(i)}, b^{(i)}]$, with positive and finite $a^{(i)}, b^{(i)} \in \Re$ as follows: (a) the intervals of the distinct objects are disjoint, and (b) there exists a vector $\alpha \in \Re^d$ with all finite components, such that

$$\alpha_*^T y \in [a^{(i)}, b^{(i)}], \qquad \text{for } i = 1, 2, \ldots n$$

where y denotes the sensor readings corresponding to O_i, when no errors are present in sensing. Informally, if the sensor errors are ideally assumed to be zero, there exists a hyperplane α_* that maps the distinct obstacles into disjoint intervals (under the scalar product operation).

In a *learning phase*, an object corresponding to the interval $[a, b]$ chosen according to an unknown distribution is placed in the workspace and the corresponding sensor reading y is recorded; the vector (a_i, b_i, y_i) is the ith *example*.

4.1 Empirical Estimation

We consider the loss function

$$Q(\alpha, y, a, b) = [(a - \alpha^T y)(b - \alpha^T y)]_+$$

where for $z \in \Re$, we have

$$[z]_+ = \begin{cases} z & \text{if } z \geq 0 \\ o & \text{if } z < 0 \end{cases}$$

Notice that if there are no sensor errors, we have $Q(\alpha_*, y, a, b) = 0$; in the presense of errors, $Q(\alpha_*, y, a, b) = 0$ if $\alpha_*^T y \in [a, b]$ and increases quadratically away from $[a, b]$. Now we have the expected loss given by

$$\begin{aligned} I(\alpha) &= \sum_{(a,b)} \int_y [(a - \alpha^T y)(b - \alpha^T y)]_+ p(a, b|y) dy \\ &= \int_{(a,b),y} [(a - \alpha^T y)(b - \alpha^T y)]_+ p(a, b, y) dy da db \end{aligned}$$

where $p(a, b, y) = p(a, b|y)p(y)$ is the joint density of (a, b, y). Now based on (1.1), the corresponding empirical functional is given by

$$I_{emp}(\alpha) = \frac{1}{l} \sum_{j=1}^{n} [(a_i - \alpha^T y_i)(b_i - \alpha^T y_i)]_+.$$

Assuming that each component of the sensor reading is bounded, we have $\alpha_*^T y \leq c$, for some constant c. We have $[(a - \alpha^T y)(b - \alpha^T y)]_+ \leq \tau$ where $\tau = c(a^{(\max)} + b^{(\max)})$ and $a^{(\max)} = \max_i \{a^{(i)}\}$. We can apply Theorem 3.2 to this case, and thus we are posed with the minimization of $I_{emp}(\alpha)$. First, the VC-dimension of set of all α is Nd since they constitute hyperplanes $\Re^{Nd}$ [5]. Now the capacity of Λ is upperbounded by the VC-dimensional of set of polygonal regions formed by three hyperplanes; then the capacity of Λ in this case is upperbounded by $3Nd + 1$ [5, 48].

4.2 Quadratic Programming Problem

We solve the problem of computing I_{emp} in two steps:

(a) We first check if I_{emp} can be made zero by solving the linear programming problem by checking for α_1 such that $a_i \leq \alpha_1^T y_i$ and $\alpha_1^T y_i \leq b_i$, for $i = 1, 2, \ldots, l$. This problem of linear inequalities can be solved using the ellipsoid algorithm [28] with a time complexity of $O(N^2 d^2 L)$ where $L = 2lNd + \log \lceil P \rceil$, P is the product of the nonzero coefficients in input specification of the problem. If no such α_1 exists, then $I_{emp}(\alpha) > 0$ for all α, then we solve this problem in the next step.

(b) Then, we use the interior ellipsoid method in [25] to minimize I_{emp}

$$I_{emp}(\alpha) = \frac{1}{l} \sum_{i=1}^{l} [(a_i - \alpha^T y_i)(b_i - \alpha^T y_i)]$$

$$= \frac{1}{l}\sum_{i=1}^{l}\alpha^T y_j y_j^T \alpha - \frac{1}{l}\sum_{i=1}^{l}[(a_j + b_j)\alpha^T y_j] + \frac{1}{l}\sum_{i=1}^{l} a_j b_j$$
$$= \alpha^T Q\alpha + z^T\alpha + c$$

where z is Nd-vector given by $z = -\frac{1}{l}\sum_{j=1}^{l}[(a_j + b_j)y_j]$, $c = \frac{1}{l}\sum_{i=1}^{l} a_j b_j$, and Q is $Nd \times Nd$ symmetric matrix given by $Q = \frac{1}{l}\sum_{i=1}^{l} y_i y_i^T$, where $y_i y_i^T$ is the outer product. Here Q is symmetric and positive definite matrix. The time complexity of minimizing $I_{emp}(\alpha)$ is polynomial in terms of size of the sample l [21].

We summarize the above results in the following theorem.

Theorem 4.1 *Given* $0 \leq \epsilon, \delta \leq 1$, $\tau = c(a^{(\max)} + b^{(max)})$ *for* $a^{(\max)} = \max_i\{a^{(i)}\}$ *and* $\alpha^T y \leq c$, *if* l *satisfies the inequality*

$$\delta(l, \epsilon) \geq \frac{9(2l)^{3Nd+1}}{(3Nd+1)!} e^{-\frac{\epsilon^2}{16\tau^2}l}$$

we can compute a fusion rule α_{emp} *in polynomial time (in* l, d, N*), such that*

$$P\{|I(\alpha_*) - I(\alpha_{emp})| \geq \epsilon\} \leq \delta$$

where α_* *is an optimal fusion rule of the given linearly separable multiple sensor system.*

5 Conclusions

We considered a general framework based on a system of multiple sensors with unknown noise characteristics. But the system is available so that readings corresponding to objects of known parameters can be obtained. In this context, we addressed the problem of inferring a fusion rule based on a set of training data with only a limited assumptions on the noise. We formulated this problem as a special case of empirical risk minimization problem of Vapnik [47]. We showed that a fusion rule close to "optimal rule" can be inferred under the conditions of finite capacity of the class of fusion rules and boundedness of the errors. This result is existential in that the computational problems associated could be intractable. We identified several computational methods to obtain f_{emp} or its approximations based on neural networks, radial basis functions, wavelets, non-polynomial networks, and polynomials and splines. Then we illustrated a system for which the computational problem is polynomial-time solvable.

The proposed methods are applicable only if suitable samples are available. If the underlying probabilities are available, then other methods are more likely to be effective. Future research directions include (a) identification of classes of Λ based on the specific properties of the system (b) analysis of the effect of the approximation methods for computing f_{emp} on the ϵ and δ, and (c) investigation of constructive methods that directly attempt to compute an approximation to f_* in stead of computing an approximation to f_{emp}.

References

[1] M. A. Abidi and R. C. Gonzalez, editors. *Data Fusion in Robotics and Machine Intelligence.* Academic Press, New York, 1992.

[2] J. A. Barnett. DSN problems-an overview. In *Proc. DSN Workshop*, pages 37–40, 1978. CMU.

[3] A. R. Barron. Universal approximation bounds for superpositions of a sigmoidal function. *IEEE Transactions on Information Theory*, 39(3):931–945, 1993.

[4] A. L. Blum and R. L. Rivest. Training a 3-node neural network is NP-complete. *Neural Networks*, 5:117–127, 1992.

[5] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth. Learnability and the vapnik-chervonenkis dimension. *Journal of the Association of Computing Machinery*, 36(4):929–965, 1989.

[6] J. M. Brady. Foreword. *Int. J. Robotics Research*, 7(5):2–4, 1988. Special Issue on Sensor Data Fusion.

[7] D. S. Broomhead and D. Lowe. Multivariable functional interpolation and adaptive networks. *Complex Systems*, 2:321–355, 1988.

[8] Z. Chair and P.K. Varshney. Optimal data fusion in multiple sensor detection systems. *IEEE Trans. Aerospace Electronic Syst.*, 22(1):98–101, 1986.

[9] S. Chen, S. A. Billings, and P. M. Grant. Recursive hybrid algorithm for non-linear system identification using radial basis function networks. *International Journal of Control*, 55(5):1051–1070, 1992.

[10] C. K. Chow. Statistical independence and threshold functions. *IEEE Trans. Electronic Computers*, EC-16:66–68, 1965.

[11] C. K. Chui. *An Introduction to Wavelets.* Academic Press, New York, 1992.

[12] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of Contols, Signals, and Systems*, 2:303–314, 1989.

[13] K. Demirbas. Distributed sensor data fusion with binary decision trees. *IEEE Trans. Aerospace Electronics Syst.*, 25(5):643–649, 1989.

[14] W. Finnoff. Diffusion approximations fot the constant learning rate backpropagation algorithm and resistence to local minima. In S. T. Hanson, J. D. Cowan, and C. L. Giles, editors, *Advances In Neural Information Processing Systems*, volume 5, pages 459–466. Morgan Kaufmann Publishers, San Mateo, California, 1993.

[15] B. Grofman and G. Owen, editors. *Information Pooling and Group Decision Making.* Jai Press Inc., Greenwich, Connecticut, 1986.

[16] K. Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4:251–257, 1991.

[17] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[18] T. L. Huntsberger. Data fusion: A neural networks implementation. In M. A. Abidi and R. C. Gonzalez, editors, *Data Fusion in Robotics and Machine Learning*, pages 507–535. Academic Press, New York, 1992.

[19] S. S. Iyengar, R. L. Kashyap, and R. B. Madan. Special issues on distributed sensors networks. *IEEE Trans. Syst. Man Cybernetics*, 1991. forth coming.

[20] A. Kak and S. Chen, editors. *Spatial Reasoning and multi-sensor fusion.* Morgan Kaufman Pub. Inc., 1987.

[21] S. Kapoor and P. Vaidya. Fast algorithms for convex quadratic programming and multicommodity. In *Proceedings of the 18th Annual ACM Sysposium on Theory of Computing*, pages 147–159, 1986.

[22] M. Leshno, V. Ya. Lin, A. Pinkus, and S. Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6:861–867, 1993.

[23] V. R. Lesser, D. D. Corkill, J. Pavlis, L. Leikowitz, E.Hdlicka, and R. Brooks. A high level simulation test bed for cooperative distributed problem solving. In *Proc. DSN Workshop*, pages 247–270, 1982. MIT Lincoln Lab.

[24] R. C. Luo and M. G. Kay. Data fusion and sensor integration: State-of-the-Art 1990s. In M. A. Abidi and R. C. Gonzalez, editors, *Data Fusion in Robotics and Machine Learning*, pages 7–135. Academic Press, New York, 1992.

[25] N. Megiddo, editor. *Progress in Mathematical Programming.* Springer-Verlag, New York, 1989.

[26] B. K. Natarajan. *Machine Learning: A Theoretical Approach.* Morgan Kaufmann Pub. Inc., San Mateo, California, 1991.

[27] V. Nedeljkovic. A novel multilayer neural networks training algorithm that minimizes the probability of classification error. *IEEE Transactions on Neural Networks*, 4(4):650–659, 1993.

[28] C. H. Papadimitriou and K. Steiglitz. *Combinatorial Optimization Algorithms and Complexity.* Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1982.

[29] L. Pitt and L. G. Valiant. Computational limitations on learnining from examples. *Journal of the Association for Computing Machinery*, 35(4):965–984, 1988.

[30] M. J. D. Powell. *Approximation Theory and Methods.* Cambridge University Press, 1981.

[31] R. Radner. Team decision problems. *Annals of Mathematical Statistics*, 33:857–881, 1962.

[32] N. S. V. Rao. Computational complexity issues in synthesis of simple distributed detection networks. *IEEE Transactions on Systems, Man and Cybernetics*, 21(5):1071–1081, 1991.

[33] N. S. V. Rao, S. S. Iyengar, and R. L. Kashyap. Computational complexity of distributed detection problems with information constraints. *Computers and Electrical Engineering*, 19(6):445–451, 1993.

[34] N. S. V. Rao and E. M. Oblow. Majority and location-based fusers for system of pac learners. *IEEE Transactions on Systems, Man and Cybernetics*, 24(4), 1994. to appear.

[35] N. S. V. Rao, E. M. Oblow, C. W. Glover, and G. E. Liepins. N-learners problem: Fusion of concepts. *IEEE Transactions on Systems, Man and Cybernetics*, 24(1), 1994. to appear.

[36] N. S. V. Rao, V. R. R. Uppuluri, and E. M. Oblow. Stochastic approximation algorithms for classes PAC learning problem. Technical report, Oak Ridge National Laboratory, 1993. under preparation.

[37] A. R. Reibman and L.W. Nolte. Optimal detection and performance of distributed sensors systems. *IEEE Trans. Aerospace Electronics Syst.*, AES-23(1):24–30, 1987.

[38] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning internal representations by error propagation. In D. E. Rumelhart and J. L. McClelland, editors, *Parallel and Distributed Processing: Explorations in the Microstructures of Cognition*. MIT Press, Cambridge, MA, 1986.

[39] F. A. Sadjadi. Hypothesis testing in a distributed environment. *IEEE Trans. Aerospace Electronics Syst.*, AES-22(2):134–137, 1986.

[40] S. Stankovic and M. Milosavljevic. Training of multilayer perceptrons by stochastic approximation. In V. Milutinovic and P. Antognelli, editors, *Neural Networks, Concepts, Applications and Implementations*, volume IV, pages 201–239. Prentice-Hall, Englewood Cliffs, NJ, 1991.

[41] R. R. Tenney and N. R. Sandell. Detection with distributed sensors. *IEEE Trans. Aerospace Electronics Syst.*, AES-17(4):501–510, 1981.

[42] S. C. A. Thomopoulos. Theories in distributed decision fusion. In A. H. Lewis and H. E. Stephanou, editors, *Distributed Intelligence Systems*, pages 195–200. Pergamon Press, New York, 1992.

[43] S. C. A. Thomopoulos, R. Viswanathan, and B. K. Bougoulias. Optimal decision fusion in multiple sensor systems. *IEEE Trans. Aerospace Electronics Syst.*, AES-23(5), 1987.

[44] S. C. A. Thomopoulos, R. Viswanathan, and B. K. Bougoulias. Optimal and suboptimal distributed decision fusion. *IEEE Trans. Aerospace Electronics Syst.*, 25(5), 1989.

[45] J. N. Tsitsiklis and M. Athans. On the complexity of decentralized decision making and detection problems. *IEEE Transactions on Automatic Control*, AC-30(5):440–446, 1985.

[46] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[47] V. Vapnik. *Estimation of Dependences Based on Empirical Data.* Springer-Verlag, New York, 1982.

[48] R. S. Wenocur and R. M. Dudley. Some special Vapnik-Chervonenkis classes. *Discrete Mathematics*, 33:313–318, 1981.

[49] P. J. Werbos. Backpropagation through time:What it does and how to do it. *Proceedings of the IEEE*, 78(10):1550–1560, 1990.

[50] R. Wesson, F. Hayes-Roth, J. W. Burge, C. Stasz, and C. A. Sunshine. Network structures for distributed situation assessment. *IEEE Trans. Systems, Man andCybernetics*, SMC-11(1):5–23, 1981.

[51] H. White. Some asymptotic results for learning in single hidden layer feedforward network models. *Journal of American Statistical Association*, 84:1008–1013, 1989.

[52] H. White. *Artificial Neural Networks.* Blackwell Publishers, Cambridge, Massachusetts, 1992.

[53] Q. Zhang and A. Benveniste. Wavelet networks. *IEEE Transactions on Neural Networks*, 3(6):889–898, 1992.

END

DATE
FILMED
3 / 7 / 94