Classifiers on Relatively Compact Sets

Irwin W. Sandberg and Ajit T. Dingankar

Abstract— **The problem of classifying signals is of interest in several application areas. Typically we are given a finite number m of pairwise disjoint sets $C_1, \ldots, C_m$ of signals, and we would like to synthesize a system that maps the elements of each C_j into a real number a_j, such that the numbers $a_1, \ldots, a_m$ are distinct. In a recent paper it is shown that this classification can be performed by certain simple structures involving linear functionals and memoryless nonlinear elements, assuming that the C_j are compact subsets of a real normed linear space. Here we give a similar solution to the problem under the considerably weaker assumption that the C_j are relatively compact and are of positive distance from each other. An example is given in which the C_j are subsets of $L_p(a,b),\ 1 \le p < \infty$.**

Keywords— **Signal Classification, Relative Compactness, Neural Networks**

I. Introduction

The problem of classifying signals is of interest in several application areas. (E.g., it arises in connection with both automatic target recognition and symbol detection in digital communication using analog channels.) Typically we are given a finite number m of pairwise disjoint sets $C_1, \ldots, C_m$ of signals, and we would like to synthesize a system that maps the elements of each C_j into a real number a_j, such that the numbers $a_1, \ldots, a_m$ are distinct. In a recent paper [3] it is shown that this classification can be performed by certain simple structures involving linear functionals and memoryless nonlinear elements, assuming that the C_j are compact subsets of a real normed linear space. In Section II we give a similar solution to the problem under the considerably weaker assumption that the C_j are relatively compact (i.e., have compact closure) and are of positive distance from each other. An example is given in which the C_j are subsets of $L_p(a,b),\ 1 \le p < \infty$. This example shows that a very wide class of signals can be classified using the structures described. The signals need not be continuous, in contrast with those of the example given in [3].

II. Classification Theorem

We need a few preliminaries:

Let $C_1, \ldots, C_m$ be subsets of a real normed linear space X, with $m > 1$. Let us say that the collection $\{C_1, \cdots, C_m\}$ is *strongly separated* if each C_i is nonempty and the distance between each pair of sets is positive in the usual sense that

$$0 < \inf\{\, \|a - b\| : a \in C_i\,, b \in C_j\}, \quad i \ne j.$$

We define a classifier on $\cup_i C_i$ to be a functional that for each i assigns a real number a_i to every $x \in C_i$ where $a_1, \cdots, a_m$ are distinct.

The authors are with the Department of Electrical and Computer Engineering, The University of Texas at Austin, Austin, Texas 78712.

Let X^* be the set of bounded linear functionals on X (i.e., the set of bounded linear maps from X to the reals $\mathbb{R}$). Given a compact subset C of X, let Y be any set of continuous maps from X to $\mathbb{R}$ that is dense in X^* on C, in the sense that for each $\phi \in X^*$ and any $\epsilon > 0$ there is a $y \in Y$ such that $|\phi(x) - y(x)| < \epsilon,\ x \in C$. Examples of such sets Y are given in Section II.C.

Finally, let U be any set of continuous maps $u : \mathbb{R} \to \mathbb{R}$ such that given $\sigma > 0$ and any bounded interval $(\beta_1, \beta_2) \subset \mathbb{R}$ there exists a finite number of elements $u_1, \ldots, u_\ell$ of U for which $|\exp(\beta) - \sum_j u_j(\beta)| < \sigma$ for $\beta \in (\beta_1, \beta_2)$.[1]

Our classification result is as follows.

Theorem 1: Let $C_0 = C_1 \cup \ldots \cup C_m$ where $\{C_1, \ldots, C_m\}$ is a strongly-separated collection of relatively compact subsets of X, and let f be a classifier on C_0 that takes the value a_i on $C_i,\ 1 \le i \le m$. Also, let $\epsilon > 0$ be given. Then there exists a positive integer k, real constants $c_1, \ldots, c_k$, elements $u_1, \ldots, u_k$ of U, and elements $y_1, \ldots, y_k$ of Y with C the closure of C_0 such that

$$a_i - \epsilon < \sum_{j=1}^{k} c_j u_j[y_j(x)] < a_i + \epsilon \quad (x \in C_i\,, \quad 1 \le i \le m).$$

Proof: We will use the following Lemma.

Lemma 1 ([3]) Let g be a real-valued continuous map defined on a compact subset K of X. Then given $\epsilon > 0$ there are a positive integer k, real numbers $c_1, \ldots, c_k$, elements $u_1, \ldots, u_k$ of U, and elements $y_1, \ldots, y_k$ of Y with $C = K$ such that

$$|g(x) - \sum_j c_j u_j[y_j(x)]| < \epsilon$$

for $x \in K$.

Continuing with the proof and using the separability of the C_i, we see that f satisfies

$$x_1, x_2 \in C_0, \quad \|x_1 - x_2\| < \delta \Rightarrow |f(x_1) - f(x_2)| = 0$$

for any sufficiently small positive δ, showing that f is uniformly continuous on C_0.

Note that $K := \overline{C}_1 \cup \overline{C}_2 \ldots \cup \overline{C}_m$, (with $\overline{C}_i$ the closure of C_i), being the union of finitely many compact sets, is compact. Since C_0 is dense in K, there exists [2, pp. 99, Problem 13] a unique continuous extension of the functional f from C_0 to K. Let g denote this extension. Using the lemma we obtain the desired result. ■

A. Comments

Theorem 1 shows that the simple structure illustrated in Fig. 1 consisting of elements y_i of a class of functionals,[2]

[1] Of course we can take U to be the set whose only element is $\exp(\cdot)$, or the set $\{u : u(\beta) = (\beta)^n/n!,\ n \in \{0, 1, \ldots\}\}$.

[2] The functionals can be taken to be linear. As mentioned earlier, examples are given in Section II.C.

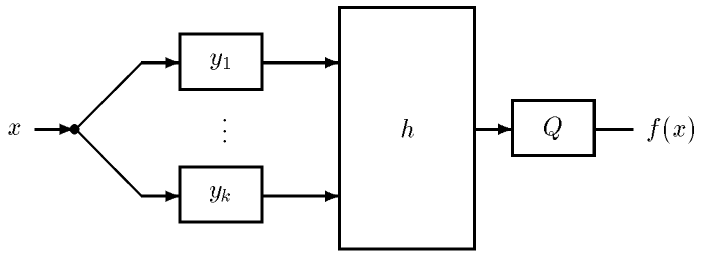

Fig. 1. Architecture of the classifier.

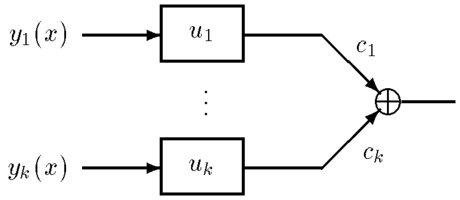

Fig. 2. Neural network implementation of the map h.

a memoryless nonlinear map h, and a quantizer (here a map Q that for each j takes numbers in the interval $(a_j - 0.25\rho, a_j + 0.25\rho)$ into a_j, where $\rho = \min_{i \neq j} |a_i - a_j|)$[3] is capable of implementing any classifier on strongly separated relatively compact sets. The theorem shows also that neural networks of a certain architecture with a single hidden layer may be used to implement the map h. This architecture is illustrated in Fig. 2 where every u_j block represents the evaluation of a real-valued function u_j defined on the reals.[4] As described above, the u_j are drawn from a set U of continuous functions whose finite sums can approximate the exponential function $\exp(\cdot)$ on bounded intervals. In particular, u_j can be taken to be $\exp(\cdot)$ for all j.

B. Example of C_0

Here we prove a theorem which provides a class of examples of relatively compact sets of particular interest concerning the classification of signals. Given $\alpha, \beta, \lambda > 0$, let $A(\lambda, \alpha, \beta)$ denote any set of real-valued functions defined on a fixed finite interval $[a, b]$ such that for every $x \in A(\lambda, \alpha, \beta)$:

1) The number of discontinuities of x is at most λ.
2) For every interval of continuity $\mathcal{I}$ of x,
$$|x(t_1) - x(t_2)| \leq \alpha |t_1 - t_2|, \quad \forall\, t_1, t_2 \in \mathcal{I}.$$
3) $|x(t)| \leq \beta, \quad \forall\, t \in [a, b]$.

Let $L_p(a, b)$, with $1 \leq p < \infty$, denote the usual normed space of real-valued Lebesgue-measurable functions defined on $[a, b]$ that are p^{th} power integrable.

Theorem 2: $A(\lambda, \alpha, \beta)$ is a relatively compact subset of $L_p(a, b)$.

Proof: We use the bound (see the proof of Lemma 3 of [4])

$$\int_a^b |x(t) - x(t+\delta)|^p dt \leq [(2\beta)^p(\lambda + 1) + \alpha(b - a)]|\delta|$$

for $x \in A(\lambda, \alpha, \beta)$ and $\delta \in I\!\!R$ (by an abuse of notation x in the integrand is an extension of x that vanishes outside $[a, b]$). This bound shows that given an $\epsilon > 0$ there exists a $\sigma > 0$ that is independent of x such that

$$\int_a^b |x(t) - x(t+\delta)|^p dt < \epsilon^p$$

for $|\delta| < \sigma$. Also, $A(\lambda, \alpha, \beta)$ is uniformly bounded because

$$\int_a^b |x(t)|^p dt \leq (b - a)\beta^p.$$

Hence, by a theorem due to Riesz [1, p. 44], $A(\lambda, \alpha, \beta)$ is relatively compact. ■

Theorem 2 shows that with $X = L_p(a, b)$, the C_i can be taken to be any strongly separated family in which each

[3] Take ϵ in Theorem 1 to be less than $.25\rho$.

[4] An important problem not considered in this paper is that of actually determining the elements in Figs. 1 and 2.

member is some $A(\lambda, \alpha, \beta)$. This example illustrates that a very wide class of signals can be classified using the structures described in the preceding section. The signals need not be continuous, in contrast with those of the example given in [3].

C. Examples of Y

Let $[a, b]$ and $L_p(a, b)$ be as described in the preceding section. For the case in which $X = L_p(a, b)$, each y_i in Fig. 1 can be taken to be represented by integration over $[a, b]$ after pointwise multiplication by a kernel in the conjugate space, since all linear functionals can be represented in that manner.

The following example of an alternative simpler representation is taken from [5]. For each integrable $x: [a, b] \to I\!R$ and each $\delta > 0$, let the so-called *Steklov function* be defined by

$$x_\delta(t) = \frac{1}{2\delta} \int_{t-\delta}^{t+\delta} x(\tau)\, d\tau, \quad t \in [a, b]$$

where (abusing the notation) x in the integrand is an extension of x that vanishes outside $[a, b]$. Note that x_δ is continuous on $[a, b]$ and that, in particular, it belongs to $L_p(a, b)$. For a given $n \in I\!N$ and $\delta > 0$, we use $\mathcal{S}_{n,\delta}$ to denote

$$\{s: L_p(a, b) \to I\!R \mid \\ s(x) = \sum_{j=1}^{n} c_j\, x_\delta(\alpha_j), \quad c_j \in I\!R, \quad 1 \le j \le n\}$$

in which the α_j are the elements of the uniformly spaced grid of n points $\alpha_j = a + j(b - a)/n, \quad 1 \le j \le n$. The set $\mathcal{S}_{n,\delta}$ is a family of what might be called Steklov functionals of order n and width δ. The class $\mathcal{S}$ of all Steklov functionals is defined by

$$\mathcal{S} = \bigcup_{\substack{n \,\in\, I\!N \\ \delta > 0}} \mathcal{S}_{n,\delta}.$$

The following theorem is proved in [5]. It justifies taking $\mathcal{S}$ as an example of a set of continuous maps from $L_p(a, b)$ to $I\!R$ that is dense in $L_p(a, b)^*$ over compacta. (That is, it justifies the statement that given any compact $C \subset L_p(a, b)$ we can take $Y = \mathcal{S}$.)[5]

Theorem 3: Let $f \in L_p(a, b)^*$ and let C be a compact subset of $L_p(a, b)$. Then for each $\epsilon > 0$ there exists an $s \in \mathcal{S}$ such that

$$|f(x) - s(x)| < \epsilon, \quad x \in C.$$

Note that the elements of $\mathcal{S}$ are simpler in form than the integrals involving a kernel mentioned at the beginning of this section, in that a member of $\mathcal{S}$ is specified by just a finite number of parameters.

[5] Also, Theorem 3 relates material in [3] and [6]. It shows that the main result in [3] yields as a particular case a result similar to one of the principal theorems in [6].

References

[1] L. A. Liusternik and V. J. Sobolev, *Elements of Functional Analysis*, Frederick Ungar Publishing Co.: New York, 1961.

[2] W. Rudin, *Principles of Mathematical Analysis*, McGraw-Hill: New York, 1976.

[3] I. W. Sandberg, "General Structures for Classification," *IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications*, vol. 41, no. 5, pp. 372–376, May 1994.

[4] I. W. Sandberg, "Uniform Approximation and the Circle Criterion," *IEEE Transactions on Automatic Control*, vol. 38, no. 10, pp. 1450–1458, October 1993.

[5] I. W. Sandberg and A. T. Dingankar, "On Approximation of Linear Functionals on L_p Spaces," submitted.

[6] T. Chen and H. Chen, "Approximations of Continuous Functionals by Neural Networks with Application to Dynamical Systems," *IEEE Transactions on Neural Networks*, vol. 4, no. 6, pp. 910-918, November 1993.