Abstract

In this paper a new metric space, called a Normalized Kolmogorov Metric Space, is defined. To evaluate its utility, a block detector is built within this metric space to be used in the automatic classification of finite length data patterns which are corrupted by additive noise. Stationary noise sources with arbitrary probability density functions can be accommodated. Extensive computer simulation results of a typical character recognition problem are presented. Similarities and extensions to earlier work on fuzzy filters are pointed out.

KEYWORDS: Normalized Kolmogorov Metric Space, Distance Measure, Non-Gaussian Noise, Block Detector, Fuzzy Filter

1. Introduction

In this paper we consider the construction of a block detector, for the automatic classification of finite length data patterns corrupted by additive noise, in a new metric space called a Normalized Kolmogorov Metric Space. Distance, as measured in this metric space, is a function of the probability density function of the noise corrupting the data pattern. The results presented are important from two perspectives. First, it presents a technique (albeit suboptimal) for handling non-Gaussian noise sources in situations when the likelihood ratio may not be mathematically tractable. Secondly, it provides an extension to earlier published work on Fuzzy Filters.

The presentation is organized into four sections. The first section, which deals with the optimal block detection of data patterns provides an overview in which important concepts are delineated. The second section formally introduces the Normalized Kolmogorov Metric Space. To evaluate the utility of the Normalized Kolmogorov Metric Space the character recognition problem is considered in the third section. Extensive computer simulation results for probability of error versus signal to noise power are presented and are compared with maximum likelihood recognition and earlier work on Fuzzy Filters. The fourth section summarizes the results obtained.


A digital communication system is one which communicates a discrete number (symbol) chosen from a finite set (alphabet). One of the most common modulation techniques for digital communication is pulse amplitude modulation (PAM) where each unique symbol is mapped to a pulse of a fixed unique amplitude.

A model of a basic PAM system is shown in Figure 1. A sequence of N data symbols, \( \mathbf{a} = (a_1, a_2, \ldots, a_N) \), each drawn independently from an alphabet of equally likely values \( \{0, 1, \ldots, m-1\} \), are transmitted at a rate \( R = 1/T \) symbols per second, through a channel. If the bandwidth of the channel is sufficient, the transmitted pulses are degraded only by the additive Gaussian noise. Therefore the received signal is given by:

\[
\mathbf{y} = \mathbf{a} + \mathbf{n}
\]

where \( \mathbf{a} \) is a vector of data symbols and \( \mathbf{n} \) is a vector of white Gaussian noise samples of variance \( \sigma_n^2 = N_0/2 \). The task of block detector (i.e., receiver) is to operate on the noisy signal vector \( \mathbf{y} \) and make a simultaneous decision on all N data digits so as to produce an estimate \( \hat{\mathbf{a}} \) of the sequence of transmitted symbols \( \mathbf{a} \). Minimizing the probability of error \( P(e) \) of this decision, i.e., that \( \mathbf{a} \) is different than \( \hat{\mathbf{a}} \), is accomplished by a maximum likelihood estimate when each of the \( m^N \) possible \( \mathbf{a} \) are equally likely.

To find the maximum likelihood estimate of \( \mathbf{a} \) is Gaussian noise the log-likelihood ratio indicates that we choose the \( \hat{\mathbf{a}} \) for which:

\[
\Gamma(\hat{\mathbf{a}}) = (\mathbf{y} - \hat{\mathbf{a}})^T \mathbf{N}^{-1}(\mathbf{y} - \hat{\mathbf{a}})
\]

is minimal, where \( \mathbf{N} = \mathbf{E} (\mathbf{n} \mathbf{n}^T) \) is the covariance matrix for the noise samples. If each of the noise samples is independent then \( \mathbf{N} = N_0/2 \mathbf{I} \), where \( \mathbf{I} \) is the identity matrix, therefore \( \Gamma(\hat{\mathbf{a}}) \) becomes:

\[
\Gamma(\hat{\mathbf{a}}) = \| \mathbf{y} - \hat{\mathbf{a}} \|^2 = (\mathbf{y} - \hat{\mathbf{a}})^T (\mathbf{y} - \hat{\mathbf{a}}) = \sum_{i=1}^{N} (y_i - \hat{a}_i)^2
\]

* Only substitution errors are considered (vis-à-vis insertion and deletion errors).
Figure 1: Communications Channel Model

In Gaussian noise, other distance measures are possible with a subsequent degradation in probability of error performance. These may be justified by other considerations such as hardware complexity and throughput of the detector. The performance degradation may even be desirable in pseudo-error rate monitoring.

One such example would be the use of an absolute value distance measure $|y_1 - y_2|$. This distance measure is optimal, in a maximum likelihood sense, for Laplacian noise. However, the simplicity of hardware implementation may tempt one to consider its use in a Gaussian noise environment.

In fact, the distance between the received signal and any of the possible estimates $\hat{x}$ can be measured by any convenient metric. A large number of metrics have been suggested in the pattern recognition literature, each having particular advantages and drawbacks. Some of the more important metrics used as interclass distance measures are: Minkowski, City Block, Euclidean, Chebyshev, Quadratic, Nonlinear.

Establishing the likelihood ratio always provides the appropriate "distance measure" required. However, for non-Gaussian noise sources the analysis may not be mathematically tractable. Despite this fact, the main criticism to applying one of the above mentioned distance measures indiscriminately is that they are not closely related to the error probability.

3. The Normalized Kolmogorov Metric

A more appropriate metric which reflects the local probability structure of data can be obtained by using more sophisticated probabilistic distance measures. For the two class separability problem, the following measures have been suggested:

- Chernoff
- Bhattacharyya
- Matchita
- The Divergence
- Patrick-Fisher
- Lissack-Fu
- Kolmogorov Variational Distance

All these measures provide an upper (lower) bound to the error probability. Of particular interest is the Kolmogorov variational distance, illustrated in Fig. 2, which is closely related to the Bhattacharyya coefficient.

At this point, we would like to introduce a new metric space derived from the Kolmogorov Variational Distance, for a two hypothesis problem (i.e., binary alphabets).

\[
\rho = \int_{-\infty}^{+\infty} |p(x|H_1)p(x|H_2)| dx
\
\rho = 1 - \rho
\]

Fig. 2: The Kolmogorov Variational Distance

Consider the Metric Space $\mathcal{M}(\Omega, \rho)$, where the set $\Omega = \{\omega\}$ is the set of stationary random variable $\omega$, $\rho(\omega|H_1)$, $\rho(\omega|H_2)$ and $\rho$ the distance measure is defined as:

\[
\rho_{H_1}(\psi) = \frac{\int_{\psi}^{\infty} p(x|H_1) dx - \int_{-\infty}^{\psi} p(x|H_2) dx}{\int_{-\infty}^{\infty} p(x|H_1) dx - \int_{-\infty}^{\infty} p(x|H_2) dx}
\]

\[
\rho_{H_2}(\psi) = 1 - \rho_{H_1}(\psi)
\]

$\psi$ is an element of the sample space $\Omega$, the subscript on $\rho$ indicates "with respect to", i.e., the distance from the received sample to the mean value (signal component) under $H_j$. This definition relates the distance of a sample from a specified hypothesis to its contribution to the probability of error. Note that the distance measure is the distribution function of the error probability for the maximum likelihood (ML) receiver. A distance of 0.2, for example indicates a miss of 20% relative to the ML probability of error.

It can be shown that $\rho_{H_1}(\psi)$ satisfies the three metric space axioms, however the proof is omitted for brevity. The distance measure always has a finite range between 0 and 1. For example, equiprobable hypotheses $H_1, H_2$ with Gaussian conditional densities of mean $\omega_1, \omega_2$ and standard deviation $\sigma_1, \sigma_2$, respectively, would have the following distance definition $\rho$:

\[
\rho_{H_1}(\psi) = \frac{Q(\frac{\psi - \psi_1}{\sigma}) - 2Q(\frac{\psi_1 - \psi_2}{2\sigma})}{2Q(\frac{\psi_1 - \psi_2}{2\sigma})}, \quad \forall \psi < \frac{\omega_1 + \omega_2}{2}
\]

\[
= 1 - \frac{Q(\frac{\psi - \psi_1}{\sigma})}{2Q(\frac{\psi_1 - \psi_2}{2\sigma})}, \quad \forall \psi > \frac{\omega_1 + \omega_2}{2}
\]

This particular metric exhibits the property that samples separated by a Euclidean distance of more than 2.0 are essentially the same distance (within 0.1%) using the new metric space.

Development of the Normalized Kolmogorov Metric Space is based on a priori knowledge of the form of the conditional density function. By the principal of maximum entropy, it is reasonable to
choose a normal density when the mean and variance are the only known parameters. (Several methods are available for estimating these parameters.) When parametric estimation is not sufficient it becomes necessary to estimate the density function by direct functional approximation. An iterative technique employed with an appropriate training sequence can estimate the true pdf to an arbitrary degree of closeness⁶.

4. Computer Simulation Results

In order to evaluate the utility of the proposed metric a computer program was written to classify alphanumeric characters corrupted by noise.

A printed alphanumeric character, represented in 7 by 5 picture elements (pixels) as shown in Figure 3(a), can be viewed as a rectangular matrix. The numerical value of each element of this 7 by 5 matrix is either zero or one corresponding to white and black, respectively. By concatenating the rows of the 7 by 5 matrix a data vector \( \mathbf{y} \) of the form described in Section 2 can be generated. Zero mean, white Gaussian noise was added to the data vector \( \mathbf{y} \) producing the vector \( \mathbf{y}^* \). Adding noise to a standard mask effectively distorts the alphanumeric character, as in Figure 3(b), when the numerical value of each \( y \) element is a real number corresponding to the gray level.

![Fig. 3: A Sample 5x7 Pixel Character Pattern](image)

(a) Example of ordinary (clear) alphanumeric pattern

(b) Example of noisy alphanumeric pattern

There are 37 alphanumeric characters in the dictionary of standard masks. Figure 4 shows the font of these characters. Our decision policy in classifying a vector \( \mathbf{y} \) is: choose the \( \mathbf{y}_i \) (one of 37) for which \( \sum_{i=1}^{37} \rho_k(y_i) \) is minimum, where \( \rho_k \) is defined by equation 2.

Probability of block error versus signal to noise ratio simulation results are shown in figure 5 along with a performance curve expected by a maximum likelihood bit by bit detector operating with the same noise model.

Since the Normalized Kolmogorov Metric Space always has a finite range, between 0 and 1, the Normalized Kolmogorov Metric Space can be used as an unambiguous grade of membership definition within the context of fuzzy set theory⁸. This motivated a comparison with earlier work on Fuzzy Filters by Wang et.al.,⁹ on a similar character recognition problem. Their block detector utilized only fuzzy and boolean operators in a normalized Euclidean metric space. The fuzzy filter perform-

![Fig. 4: The Complete Dictionary of Standard Masks](image)

![Fig. 5: P(e) vs. SNR, averaged over the whole character set for noise with a Gaussian probability density function](image)

5. Discussion and Conclusions

The concept of a Normalized Kolmogorov Metric Space has been introduced. It presents a technique (albeit suboptimal) for handling non-Gaussian noise sources in situations when the likelihood ratio may not be mathematically tractable. Since distance within this metric space always has a finite range, it can be accurately accommodated in digital hardware using a finite number of bits, specified by limits on quantization error.
We have successfully demonstrated the utility of the construct in a typical character recognition problem. Since the Normalized Kolmogorov Metric Space by definition always has a finite range between 0 and 1, it can be used as an unambiguous grade of membership definition within the context of fuzzy set theory. In fact, we have demonstrated performance improvements over previous work on fuzzy filters using noise sources with various probability density functions, on a similar character recognition problem.

Though suboptimal in a maximum-likelihood sense, preliminary work in using the Viterbi algorithm\(^1\) in a Normalized Kolmogorov Metric Space environment indicates that, the proposed metric space is important for its robust performance in burst noise and fading environments.

REFERENCES


