ARTICLE IN PRESS
YACHA:1184

Appl. Comput. Harmon. Anal. • • • (• • • •) • • •–• • •

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

The cumulative distribution transform and linear pattern classification

Se Rim Park [a], Soheil Kolouri [b], Shinjini Kundu [c], Gustavo K. Rohde [d,e,*]

[a] *ECE Department, Carnegie Mellon University, Pittsburgh, PA, 15213, United States*
[b] *HRL Laboratories, Malibu, CA, 90265, United States*
[c] *Medical Scientist Training Program, University of Pittsburgh, PA, 15213, United States*
[d] *Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA, 22908, United States*
[e] *Department of Biomedical Engineering, University of Virginia, Charlottesville, VA, 22908, United States*

A R T I C L E   I N F O

Article history:
Received 23 March 2015
Received in revised form 9 February 2017
Accepted 11 February 2017
Available online xxxx
Communicated by Charles K. Chui

Keywords:
Cumulative distribution transform
Signal classification

A B S T R A C T

Discriminating data classes emanating from sensors is an important problem with many applications in science and technology. We describe a new transform for pattern representation that interprets patterns as probability density functions, and has special properties with regards to classification. The transform, which we denote as the Cumulative Distribution Transform (CDT), is invertible, with well defined forward and inverse operations. We show that it can be useful in 'parsing out' variations (confounds) that are 'Lagrangian' (displacement and intensity variations) by converting these to 'Eulerian' (intensity variations) in transform space. This conversion is the basis for our main result that describes when the CDT can allow for linear classification to be possible in transform space. We also describe several properties of the transform and show, with computational experiments that used both real and simulated data, that the CDT can help render a variety of real world problems simpler to solve.

1. Introduction

Mathematical transforms are useful tools in engineering, physics, and mathematics given that they can often render certain problems easier to solve in transform space. Fourier transforms [21] for example, are well-known for providing simple answers related to the analysis of linear shift-invariant systems. Wavelet transforms, on the other hand, are well suited for detecting and analyzing signal transients (fast changes) [29]. These and other transforms have been instrumental in the design of sampling and reconstruction

* Corresponding author.
E-mail addresses: park@cmu.edu (S.R. Park), skolouri@hrl.com (S. Kolouri), shk71@pitt.edu (S. Kundu), gustavo@virginia.edu (G.K. Rohde).

http://dx.doi.org/10.1016/j.acha.2017.02.002
1063-5203/

algorithms for analog-to-digital conversion, modulation and demodulation, compression, communications, etc., and have found numerous applications in science and technology.

On the other hand, the past few decades have brought about the emergence of ubiquitous, accurate, user friendly, and low cost digital sensing devices. These devices produce a wealth of data about the world we live in, ranging from digital microscopy images of sub-cellular patterns to satellite imagery and detailed telescope images of our universe. The relative ease with which vast amounts of data can be accessed and queried for information has brought about challenges related to 'telling signals apart', or sensor data classification. Examples include being able to distinguish between benign and malignant tumors from medical images [19], between 'normal' and 'abnormal' physiological sensor data (e.g. flow cytometry) [34], identifying people from images of faces or fingerprints [41], identifying biological/chemical threats from resonant optical spectra [16] and others. The high-dimensional nature of the measurements in relation to the number of samples available often makes these problems challenging.

Important practical questions often arise in the process of designing solution to many data classification problems. Examples would be: "Which features should be extracted?", "What classifier should be used?", "How can one model, visualize and understand any discriminating variations in the dataset?", etc. For many applications where optimal feature sets are yet to be discovered, researchers are faced with the task of utilizing a *trial and error* approach that involves testing for different combinations of features [15,25], classifiers [11], kernels [33] in the effort to arriving at a useful solution of the problem. We note that many of the available signal transforms (Wavelet, Fourier, Hilbert, etc.) are linear transforms, and thus offer limited capabilities related to enhancing or facilitating separation in feature (transform) space unless some non-linear operations are performed.

Here we describe a new one-dimensional signal transformation framework, with well defined analysis (forward transform) and synthesis (inverse transform) operations that, for signals that can be interpreted as probability density functions, can help facilitate the problem of recognition. Denoted as the Cumulative Distribution transform (CDT), the CDT can be viewed as a one to one mapping between the space of smooth probability densities and the space of differentiable functions, and therefore by definition retains all of the signal information. We show that the CDT can be computed efficiently, and is able to turn certain types of classification problems linearly separable in the transform space. In contrast to linear data transformation frameworks (e.g. Fourier and Wavelet transforms) which simply consider signal intensities at fixed coordinate points, thus adopting an 'Eulerian' point of view, the idea behind the CDT is to also consider the location of the intensities in a signal, with respect to a chosen reference, in the effort to 'simplify' pattern recognition problems. Thus, the CDT adopts a 'Lagrangian' point of view for analyzing signals. The idea is similar to our work on linear optimal transport [40], and the links will be explicitly elucidated below.

1.1. Signal discrimination problems

Let $\mathbb{P}$ and $\mathbb{Q}$ denote two disjoint classes of functions (signals) within a normed vector space V. The goal in classification is to deduce a functional to 'regress' a given label for each signal [4]. For a binary classification problem, the label of each signal can be considered 0/1 or -1/+1, and the problem of classifying a signal f can be solved by finding a linear functional $T : V \to \mathbb{R}$ and $b \in \mathbb{R}$ such that

$$\begin{aligned} T(f) < b \qquad \forall f \in \mathbb{P}, \\ T(f) > b \qquad \forall f \in \mathbb{Q}. \end{aligned} \tag{1}$$

Below we specifically consider the case when T is a linear classifier in V. For example, for real functions in L^2, one may find w such that $T(f) = \int_V w(x)f(x)dx$. For discrete signal data in countable domain $\mathbb{Z}$ one may find w such that $T(f) = \sum_{k\in\mathbb{Z}} w[k]f[k]$. Thus the goal is to obtain the linear function w and the