Knowledge-Based Systems 97 (2016) 224–236

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Comparison study of orthonormal representations of functional data in classification

Yinfeng Meng[a,b], Jiye Liang[a,c,*], Yuhua Qian[a,c]

[a] School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China
[b] School of Mathematical Sciences, Shanxi University, Taiyuan 030006, Shanxi, China
[c] Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Taiyuan 030006, Shanxi, China

ARTICLE INFO

Article history:
Received 23 June 2015
Revised 20 November 2015
Accepted 25 December 2015
Available online 11 January 2016

Keywords:
Functional data
Orthonormal representation
Orthonormal basis
Classification

ABSTRACT

Functional data type, which is an important data type, is widely prevalent in many fields such as economics, biology, finance, and meteorology. Its underlying process is often seen as a continuous curve. The classification process for functional data is a basic data mining task. The common method is a two-stage learning process: first, by means of basis functions, the functional data series is converted into multivariate data; second, a machine learning algorithm is employed for performing the classification task based on the new representation. The problem is that a majority of learning algorithms are based on Euclidean distance, whereas the distance between functional samples is L_2 distance. In this context, there are three very interesting problems. (1) Is seeing a functional sample as a point in the corresponding Euclidean space feasible? (2) How to select an orthonormal basis for a given functional data type? (3) Which one is better, orthogonal representation or non-orthogonal representation, under finite basis functions for the same number of basis? These issues are the main motivation of this study. For the first problem, theoretical studies show that seeing a functional sample as a point in the corresponding Euclidean space is feasible under the orthonormal representation. For the second problem, through experimental analysis, we find that Fourier basis is suitable for representing stable functions(especially, periodic functions), wavelet basis is good at differentiating functions with local differences, and data driven functional principal component basis could be the first preference especially when one does not have any prior knowledge on functional data types. For the third problem, experimental results show that orthogonal representation is better than non-orthogonal representation from the viewpoint of classification performance. These results have important significance for studying functional data classification.

1. Introduction

Recent years have witnessed considerable improvements in data acquisition technology and data storage abilities. As a result, it has become imperative to classify individual systems in various research fields based on one or more data series. The underlying process of every data series is an unknown function (continuous curve), called functional data. The classification process for functional data is typically the same as that for their underlying generation functions.

At present, for the classification of functional data, there are two types of commonly used methods. One involves constructing functional classifiers, such as a functional support vector machine (SVM) by means of kernel techniques [3,33,48] and functional logistic regression [5,18,24,43,47,49], and the other is a two-stage classification method [28]. For the second method, in the first stage, usually, functional samples are represented in a finite dimensional functional subspace by means of basis functions; thus, functional data with infinite dimension becomes multivariate data, which consists of coefficients before the basis functions. In the second stage, a classical learning algorithm for finite dimensional data is used. The reason is that the high dimensionality of data series renders many data mining methods ineffective and fragile [8]. This obstacle is sometimes referred to as the "curse of dimensionality" [14]. In most data series mining problems, there is a need for dimensionality reduction and forming new data series representations [27]. It is required that the new representation preserves sufficient information for solving data series mining problems correctly. Once the basis is chosen, the optimal value for the number of basis functions can be derived from the data [48].

* Corresponding author at: School of Computer and Information Technology, Shanxi University, Taiyuan 030006, Shanxi, China. Tel./fax: +86 0351 7018176.

E-mail addresses: mengyf@sxu.edu.cn (Y. Meng), ljy@sxu.edu.cn (J. Liang), jinchengqyh@sxu.edu.cn (Y. Qian).

http://dx.doi.org/10.1016/j.knosys.2015.12.016
0950-7051/© 2016 Elsevier B.V. All rights reserved.

Representing data series in the transformed domain is a common dimensionality reduction approach. Some of the popular transformation techniques are Fourier transform [15,33,53] and wavelet transform [11,16,32,37]. Functional principal component analysis(FPCA) [10,21,29,39,43,46,54–56] is a popular technique that uses statistical methods. Other methods include B-spline functions [1,3,35,59], Mercer kernel transforms [36,38], radial basis functions [4,5,26], etc.

In fact, the representation of functional data is essentially a kind of approximation of itself. In the process of machine learning of functional data, a kind of structured representation using basis functions is used to transform functional data into multivariate data, and then, the distances between functional samples are converted into the Euclidean distances between the corresponding multivariate data. However, the representability of using the corresponding multivariate data to represent functional data, and the rationality of using the distance between the corresponding two multivariate data to replace the distance between two functional samples have not been studied in detail. Therefore, the relationship of different spaces is first introduced, and then the orthonormal representation theory is employed to explain the representability and rationality.

Theoretically, under orthonormal basis, for any two different functional samples, the distance between them can be approximated based on the distance between their low-dimensional representations, which is isomorphic to the corresponding Euclidean distance. At this time, choosing an appropriate orthonormal basis is still a problem. Therefore, three kinds of common orthonormal basis and their differences are considered. The three kinds of orthonormal basis are normal Fourier basis, wavelet basis, and functional principal component basis, the eigenequation of FPCA is derived by means of variational theory.

It is well known that non-orthogonal representation can also represent a functional data series as certain multivariate data. Therefore, it is important to verify if orthogonal basis has a stronger representation ability than non-orthogonal basis for functional data under the same number of basis functions from the viewpoint of classification performance.

In order to verify the representation ability of the above orthonormal basis in classification, the extracted features(the coefficient vector, which consists of coefficients before the basis functions) of the functional data will be used in classification model construction. It has been pointed out in the literature [17] that support vector machine(SVM) and random forest are two preferred classification methods, and thus, LibSVM [12] and RandomForest [9,44] are first used to classify the functional data for three kinds of orthonormal representations. As other choices, logistic regression [29,40], K-nearest neighbor [30,31], and artificial neuron network [34,41] will also be used as classifiers for discriminating functional samples. Based on these classifiers, we shall also compare the classification performance of orthogonal representation with that of non-orthogonal representation.

The main objective of this paper is to explain the rationality behind converting functional samples into corresponding multivariate data that are to be used for training a classifier. At the same time, from the point of view of experiments, we shall explain that among the three basis candidates, Fourier basis is suitable for representing stable signals(especially, periodic functions), wavelet representation can yield better results than Fourier representation for non-stationary signals, and orthonormal basis obtained through functional principal components offers good representation ability for some functional data with complex trend characteristics. Functional principal component analysis (FPCA), in particular, can be the first choice when people do not have any prior knowledge. Furthermore, we also demonstrate that orthogonal basis is indeed better than non-orthogonal basis from the viewpoint of classification performance.

Table 1
The observation form of functional data.

Sample	t_1	t_2	$\cdots$	t_p
X_1	$X_1(t_1)$	$X_1(t_2)$	$\cdots$	$X_1(t_p)$
X_2	$X_2(t_1)$	$X_2(t_2)$	$\cdots$	$X_2(t_p)$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	
X_N	$X_N(t_1)$	$X_N(t_2)$	$\cdots$	$X_N(t_p)$

The remainder of this paper is organized as follows. Some basic concepts of functional data and some approximation theory under orthonormal representation are presented in Section 2. Section 3 describes three kinds of common orthonormal representations for functional data, and in particular, the eigenequation for functional principal component is derived using the variational principle. Section 4 introduces several classification methods including LibSVM, RandomForest, logistic regression, K-nearest neighbor, and artificial neuron network. Furthermore, four classification performance indexes such as the precision, the recall, $F1$ score, and the accuracy are introduced in detail. Section 5 provides numerical studies for feature extraction and classification methods for functional data. In this section, we analyze the classification performance of three different kinds orthonormal basis, point out which kind of orthonormal basis is appropriate to represent what type of functional data, and answer whether orthogonal representation is better than non-orthogonal representation for classifying functional data for the same number of finite basis functions. Section 6 concludes the paper with some remarks and discussions.

2. Orthonormal representation for functional data

2.1. The basic concepts of functional data

Advances in data collection and storage have led to an increased presence of functional data, whose graphical representations are curves, images, or shapes [51]. The observation form of the functional data is also a two-dimensional table, which is shown in Table 1, in which $X_i(t)$ (abbreviated as X_i), $t \in I$, $i = 1, 2, \ldots, N$ is an underlying continuous and smooth function, and $X_i \in L^2(I)$, where $L^2(I)$ is the space of the square-integrable functions defined on the compact set I, $X : I \to \mathcal{R}$, $(\int_I X^2(t)dt)^{1/2} < \infty$, $\mathcal{R}$ is the real number space. At the same time, $L^2(I)$ is a separable Hilbert space with the inner product $< X, Y >= \int_I X(t)Y(t)dt$ and the norm $\|X\|_2 = (\int_I X^2(t)dt)^{1/2}$. $X_i(t_j)$ denotes the observed value for $X_i(t)$ at a discrete point t_j for the ith functional sample.

To understand the $L^2(I)$ space, the relationship among different spaces is first introduced. It is well known that the introduction of the distance is for the purpose of studying the convergence. People, therefore, defined the metric space. In the metric space, the distance between any two elements can be computed. If the concept of completion (any Cauchy sequence is a convergent sequence [58]) is introduced in the metric space, the space will become a complete metric space.

However, the metric space only has a topological structure, which restricts its application area. If a linear operation is introduced to the metric space, a linear normal space [58] can be obtained and the algebraic operation between elements can be carried out. In this case, the distance is transformed into the norm, which combines the metric and the linear operations perfectly. In other words, the linear normal space not only keeps its topological structure but also maintains its algebraic structure. The