Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2014, Article ID 572124, 12 pages
http://dx.doi.org/10.1155/2014/572124

Research Article

Generalized Framework for Similarity Measure of Time Series

Hongsheng Yin,[1] Honggang Qi,[2] Jingwen Xu,[2,3] William N. N. Hung,[4] and Xiaoyu Song[5]

[1] *China University of Mining and Technology, Xuzhou 221116, China*
[2] *University of Chinese Academy of Sciences, Beijing 100049, China*
[3] *Institute of Electrical Engineering, Chinese Academy of Sciences, Beijing 100190, China*
[4] *Synopsys Inc., Mountain View, CA 94043, USA*
[5] *Portland State University, Portland, OR 97207, USA*

Correspondence should be addressed to Honggang Qi; hgqi@ucas.ac.cn

Received 9 July 2014; Revised 10 October 2014; Accepted 13 October 2014; Published 3 November 2014

Academic Editor: Yan Liang

Currently, there is no definitive and uniform description for the similarity of time series, which results in difficulties for relevant research on this topic. In this paper, we propose a generalized framework to measure the similarity of time series. In this generalized framework, whether the time series is univariable or multivariable, and linear transformed or nonlinear transformed, the similarity of time series is uniformly defined using norms of vectors or matrices. The definitions of the similarity of time series in the original space and the transformed space are proved to be equivalent. Furthermore, we also extend the theory on similarity of univariable time series to multivariable time series. We present some experimental results on published time series datasets tested with the proposed similarity measure function of time series. Through the proofs and experiments, it can be claimed that the similarity measure functions of linear multivariable time series based on the norm distance of covariance matrix and nonlinear multivariable time series based on kernel function are reasonable and practical.

1. Introduction

A complex system typically needs to be described with multiple state variables. These state variables can be obtained by experimental observations and instrumental measures. With these state variables, a set of discrete multivariate time series (MTS) can be constructed. Mathematically, MTS is expressed in matrix form $\mathbf{X} = (X_1, X_2, \ldots, X_m)$. Its samples are $x_i(t)$, x_i $(i = 1, 2, \ldots, m;\ t = 1, 2, \ldots, n)$, where m and n denote the number of observation variables and the number of observation samples of each observation variable, respectively. Obviously, $\mathbf{X}$ is a $n \times m$ matrix. If $m \geq 2$, the matrix $\mathbf{X}$ represents MTS. Otherwise, if $m = 1$, $\mathbf{X}$ is simplified to a univariate time series (UTS), which is a special case of MTS and can be denoted as a n-dimensional vector $x(n)$. Time series theory is widely applied in various fields such as electricity, finance, medical, multimedia, meteorology and hydrology, scientific research, and industrial control. Discovery of the hidden information and operating regularity in a time series is a research hotpot in data mining and knowledge discovery. The research on time series includes clustering, classification, similarity search, feature extraction, trend forecasting, and decision support. Similarity measures a fundamental research topic on time series theory.

Most of the existing research on the similarity measure of time series is focused on UTS. The common measure functions are L_p-norms $(p = 1, 2, \infty)$ [1, 2], DTW [3–5], longest common subsequence (LCSS) [6], edit distance on real sequence (EDR) [7], edit distance with real penalty (ERP) [8], spatial assemble distance (SpADe) [9], DISSIM [10], swale [11], and TQuEST [12]. With further research on UTS, we can expect more new methods will be proposed in the future. However, few researches focus on the similarity measure of MTS. Yang and Shahabi [13, 14] calculated the similarity of MTS with the extended Frobenius norm. Xu et al. [15, 16] measured the MTS similarity based on information theoretic learning framework.

The contributions in the research on framework for similarity measure of time series are also less. Liu and Jiang [17] proposed a concept of similarity of time series through

analyzing the geometric relation of Euclidean distance of time series in high-dimensional space, which describes the similarity relation between two UTS using both similarity function and transform constraint function to establish an exact concept for similarity of time series.

From existing UTS and MTS literatures, the similarity measure methods of time series are undefined and nonstandardized which makes the research very difficult. Thus, it is very necessary to establish a general concept for uniform similarity of time series. In this paper, we give the definition of similarity measure function of UTS with vector norm and proved that the similarity functions defined in the form of norms are equivalent in original UTS space and linear transformed space. Moreover, we also extend the definitions of similarity of time series and present a uniform theory of similarity measure based on set theory, metric space theory, operator theory, matrix theory, and kernel method. The uniform theory based on distance of vector/matrix norm can be used for measuring the similarity of time series in both original and transformed spaces, whether the time series are univariable or multivariable and the linear transform or the nonlinear transform. The theory analysis and experimental results show that the definition of distance of vector/matrix norm is equivalent in original and transformed spaces.

The rest of this paper is organized as follows. In Section 2, it is proved that the vector norm based definitions of similarity functions are equivalent in UTS original and linear transformed spaces. In Section 3, the theory that matrix norm is used for defining the similarity function of linear MTS is discussed. Then, the theory that kernel function is used for defining the similarity function of nonlinear MTS is discussed in Section 4. In Section 5, all similarity functions proposed in this paper are discussed and analyzed. It is proved that the norm based definition of similarity functions for measuring the similarity of UTS is equivalent in time domain and Fourier transform or wavelet transform domain. Also, the similarity function of linear MTS defined based on covariance matrix norm distance and the similarity function of nonlinear MTS defined based on kernel function are analyzed in this section. The experimental results are shown in Section 6. Finally, we conclude this paper in Section 7.

2. Similarity Measure of Linear Univariable Time Series

For the original data of linear univariable time series, whether they are recoded manually or sampled automatically, it is assumed that they all satisfy Shannon theorem of information theory without any distortion and without considering the data dimension. The linear univariable time series is denoted as set $\mathbf{A}$. Considering two UTS samples $x, y \in \mathbf{A}$, wherein x is the time series to be observed (observed time series) and y is the time series to be referenced (referenced time series). They are represented in vector form $\mathbf{X} = (X_1, X_2, \ldots, X_m)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$, $m \geq n$. To measure the similarity of x and y, the subseries with the same dimension as y needs to be extracted from x. All n-dimension successive subseries of x compose a set $\mathbf{B}$, $\mathbf{B} = (X_i, X_{i+1}, \ldots, X_{i+n-1} \mid i = 1, 2, \ldots, m - n + 1)$, in which each element represents a n-dimension successive time series. The ith element in set $\mathbf{B}$ is denoted as $x_i^{(b)}$, $x_i^{(b)} = (X_i, X_{i+1}, \ldots, X_{i+n-1})$, which is a n-dimension vector.

The similarity measure between observed time series x and referenced time series y is that the subseries of $x, x_i^{(b)}, \forall x_i^{(b)} \in \mathbf{B}$, and the given referenced series y are measured in the similarity measure function $\text{Sim} = f(x_i^{(b)}, y)$ with the threshold $\varepsilon > 0$.

Definition 1. If the similarity function of time series $\text{Sim} = f(x_i^{(b)}, y) \leq \varepsilon$, the observed subseries $x_i^{(b)}$ and referenced series y are similar.

According to the literatures [1, 2], there are two cases of time series similarities, which are as follows: (1) if $m = n$ and $\text{Sim} = f(x_i^{(b)}, y) \leq \varepsilon$, x and y are exactly matching; (2) if $m > n$, $\exists x_i^{(b)} \in \mathbf{B}$, and $\text{Sim} = f(x_i^{(b)}, y) \leq \varepsilon$, the subseries $x_i^{(b)}$ and y are subseries matching. Obviously, exactly matching is a special case of subseries matching. In this paper, subseries $x_i^{(b)}$ and similarity measure function $\text{Sim} = f(x_i^{(b)}, y)$ are defined for unifying the above two cases of similarity matching of time series and measuring the similarity of $x_i^{(b)}$ and y in the same dimension conveniently. The following definitions on similarity of UTS are all based on the unified similarity.

2.1. Common Similarity Measure Functions. In current research, the distance function of time series $x_i^{(b)}$ and y, $d(x_i^{(b)}, y)$, is commonly used as the similarity measure function $\text{Sim} = f(x_i^{(b)}, y)$. The common distance metrics are as follows [17].

(1) Euclidean distance:

$$\text{Sim} = f\left(x_i^{(b)}, y\right) = d\left(x_i^{(b)}, y\right) = \left[\sum_{k=1}^{n}\left(x_{ik}^{(b)} - y_k\right)^2\right]^{1/2}. \quad (1)$$

(2) City-block distance:

$$\text{Sim} = f\left(x_i^{(b)}, y\right) = d\left(x_i^{(b)}, y\right) = \sum_{k=1}^{n}\left|x_{ik}^{(b)} - y_k\right|. \quad (2)$$

(3) Chebyshev distance:

$$\text{Sim} = f\left(x_i^{(b)}, y\right) = d\left(x_i^{(b)}, y\right) = \sum_{k}\left|x_{ik}^{(b)} - y_k\right|. \quad (3)$$

(4) Minkowski distance:

$$\text{Sim} = f\left(x_i^{(b)}, y\right) = d\left(x_i^{(b)}, y\right) = \left[\sum_{k=1}^{n}\left(x_{ik}^{(b)} - y_k\right)^p\right]^{1/p}. \quad (4)$$

(5) Cosine distance:

$$\text{Sim} = f\left(x_i^{(b)}, y\right) = d\left(x_i^{(b)}, y\right) = 1 - \frac{\sum_{k=1}^{n}\left(x_{ik}^{(b)} - y_k\right)}{\sqrt{\sum_{k=1}^{n}\left(x_{ik}^{(b)}\right)^2 \sum_{k=1}^{n}\left(y_k\right)^2}}. \quad (5)$$

(6) Correlation distance:

$$\begin{aligned}\text{Sim} &= f\left(x_i^{(b)}, y\right) = d\left(x_i^{(b)}, y\right) \\ &= 1 - \left(\left(\sum_{k=1}^{n}\left[\left(x_{ik}^{(b)} - \frac{1}{n}\sum_{k=1}^{n}x_{ik}^{(b)}\right)\left(y_{ik}^{(b)} - \frac{1}{n}\sum_{k=1}^{n}y_{ik}^{(b)}\right)\right]\right)\right. \\ &\quad \left.\times\left(\sqrt{\sum_{k=1}^{n}\left(x_{ik}^{(b)} - \frac{1}{n}\sum_{k=1}^{n}x_{ik}^{(b)}\right)^2\sum_{k=1}^{n}\left(y_{ik}^{(b)} - \frac{1}{n}\sum_{k=1}^{n}y_{ik}^{(b)}\right)^2}\right)^{-1}\right).\end{aligned} \tag{6}$$

(7) Mahalanobis distance

$$\begin{aligned}\text{Sim} &= f\left(x_i^{(b)}, y\right) = d\left(x_i^{(b)}, y\right) \\ &= \left[\left(x_k^{(b)} - y\right)V^{-1}\left(x_k^{(b)} - y\right)^T\right]^{1/2}.\end{aligned} \tag{7}$$

In formula (7), matrix V is the covariance matrix of $x_i^{(b)}$ and y. Mathematically, it can be proved that the distance formulas (1)–(3) are special cases of formula (4) with $p = 2, 1, \infty$. In these distance functions, the Euclidean distance is frequently used in practice. Obviously, given the same dimension, the more the similarity between two vectors, the smaller the values of similarity measure functions (1)–(7), and vice versa.

2.2. Relevant Concepts of Time Series Transform. Since time series $x_i^{(b)}$ and y are high-dimension data, $x_i^{(b)}, y \in \mathbf{X} \subset \mathbb{R}^n$, straightforward analysis and processing of similarity of time series need huge computation burden which is unacceptable on both time and space complexities. Although measuring the similarity of two time series is intuitively straightforward, the result may be not very accurate. Thus, the time series $x_i^{(b)}$ and y need to be properly transformed. The transformed time series are denoted as $x_i^{(b)}$ and y', and $\mathbf{X}'$ denotes the transformed space, $x_i^{(b)'}, y' \in \mathbf{X}' \subset \mathbb{R}^n$, and their transform factor is $\mathbf{T}$. The transform should be lossless or lossy within a very small margin so that no or less accuracy loss of data is introduced by the transform. Through the space transform, the dimension of data is greatly reduced so that the data can be processed with lower complexity. Accordingly, it is very important to select a proper transform for solving this problem. For notational convenience the relevant definition is given as follows.

Definition 2. Let $(\mathbf{X}, \rho)$ and $(\mathbf{X}', \rho')$ be two measure spaces. If there exists a mapping $\mathbf{T}$ from $\mathbf{X}$ to $\mathbf{X}'$, and $\forall x, y \in \mathbf{X}$, $\rho'(\mathbf{T}x, \mathbf{T}y) = \rho(x, y)$, then it can be said that $\mathbf{X}$ and $\mathbf{X}'$ are isometric, and $\mathbf{T}$ is the isometric mapping from $\mathbf{X}$ to $\mathbf{X}'$.

Definition 3. Let $\mathbf{T}$ be an operator (mapping) of the normed linear space from $\mathbf{X}$ to $\mathbf{X}'$. If the following hold:

(1) additive $\mathbf{T}(x + y) = \mathbf{T}x + \mathbf{T}y$ $(x, y \in X)$,

(2) homogeneity $\mathbf{T}(\alpha x) = \alpha\mathbf{T}x$,

then it can be said that $\mathbf{T}$ is a linear operator from $\mathbf{X}$ to $\mathbf{X}'$.

Definition 4. $S(\mathbf{X}, \mathbf{X}')$ represents the all bounded linear operators of the normed linear space from $\mathbf{X}$ to $\mathbf{X}'$. Let $\mathbf{T}, \mathbf{T}_1 \in S(\mathbf{X}, \mathbf{X}')$ and let α be arbitrary number. If $\forall x \in \mathbf{X}$ and the following hold:

(1) additive $(\mathbf{T}_1 + \mathbf{T})x = \mathbf{T}_1 x + \mathbf{T}x$,

(2) homogeneity $(\alpha\mathbf{T})x = \alpha(\mathbf{T}x)$,

then it can be said that $S(\mathbf{X}, \mathbf{X}')$ is a linear space.

Definition 5. Let $\mathbf{X}$ and $\mathbf{X}'$ be two normed linear spaces and let $\mathbf{T} \in S(\mathbf{X}, \mathbf{X}')$ be the linear operator from $\mathbf{X}$ to $\mathbf{X}'$. If $\|x\| = \|\mathbf{T}x\|'$ and $\forall x \in \mathbf{X}$, then it can be said that $\mathbf{X}$ and $\mathbf{X}'$ are isometric and $\mathbf{T}$ is the norm preserving isomorphic mapping from $\mathbf{X}$ to $\mathbf{X}'$.

2.3. Definition of Similarity Relation in Set Theory

Definition 6. Let R be a relation in set $\mathbf{A}$. If $\forall x \in \mathbf{A}$ and $(x, x) \in R$, then it can be said that relation R is reflexive. If $\forall x, y \in \mathbf{A}$, $(x, y) \in R$, and $(y, x) \in R$, then it can be said that relation R is symmetrical. If $\forall x, y, z \in \mathbf{A}$, $(x, y) \in R$, $(y, z) \in R$, and there is $(x, z) \in R$, then it is said that relation R is transitive. If a relation R is both reflexive and symmetrical, it is a similarity relation.

It should be noted that the similarity relation does not have the transitive property. For example, father and son are similar, and mother and son are also similar, but father and mother are very possible to be dissimilar. According to the above analysis of similarity of time series, the similarity measure functions are proposed based on the similarity relation of set theory.

2.4. Extended Similarity Definition of Time Series. According to the above discussion, the similarity measure function of time series not only satisfies the similarity relation in set theory, but also should be uniformed in the original space $\mathbf{X}$ and the transformed space $\mathbf{X}'$. Objectively, the straightforward similarity measure of two time series is more accurate than nonstraightforward similarity measure in which the geometric triangle inequality is used for the similarity measure through the third-party time series. Thus, Definition 1 is extended as follows.

Definition 7. Given a linear operator T from $\mathbf{X}$ to $\mathbf{X}'$ and $x_i^{(b)}, y, z \in \mathbf{X} \subset \mathbb{R}^n$, $x_i^{(b)'}, y', z' \in \mathbf{X}' \subset \mathbb{R}^n$ are the transforms of $x_i^{(b)}, y, z$ by the operator T. When the similarity measure function $\text{Sim} = f(*, *) \leq \varepsilon$ ($\varepsilon > 0$ is a given threshold of similarity), the time series $x_i^{(b)}$, y are similar in the constraints of ε. Meanwhile, the similarity function $\text{Sim} = f(*, *)$ should satisfy the following.

(1) Reflexive symmetrical positive definiteness, that is, $\text{Sim} = f(x_i^{(b)}, y) = f(y, x_i^{(b)}) = f(x_i^{(b')}, y') = f(y', x_i^{(b)'}) \geq 0$, if and only if $x_i^{(b)} = y$, $x_i^{(b')} = y'$, and the equation $\text{Sim} = 0$ holds.

(2) Geometric triangle inequality:

$$f\left(x_i^{(b)}, z\right) \le f\left(x_i^{(b)}, y\right) + f(y, z),$$
$$f(x_i^{(b)'}, z') \le f(x_i^{(b)'}, y') + f(y', z').$$

The generalized similarity definition in different spaces is given by Definition 7. The key problem is how to find the proper operator $\mathbf{T}$ and similarity measure function $\mathrm{Sim} = f(*, *)$.

Lemma 8. *Any two norms of finite-dimension linear space are equivalent.*

Proof. Let $\|x\|_1$ and $\|x\|_2$ be two norms of linear space $\mathbf{X}$ and let $\{e_1, e_2, \ldots, e_n\}$ be a linear basis of $\mathbf{X}$. If $\forall x \in \mathbf{X}$ and $x = \sum_{k=1}^{n} a_k e_k$, then there are positive integers A_1, A_2, B_1, B_2, which satisfy $A_1 \|x\|_1 \le \left\{\sum_{k=1}^{n} |a_k|^2\right\}^{1/2} \le B_1 \|x\|_1$ and $A_2 \|x\|_2 \le \left\{\sum_{k=1}^{n} |a_k|^2\right\}^{1/2} \le B_2 \|x\|_2$. Thus, $A_1 \|x\|_1 \le B_2 \|x\|_2 \le (B_1 B_2 / A_2) \|x\|_1$; that is, $(A_1/B_2) \|x\|_1 \le \|x\|_2 \le (B_1/A_2) \|x\|_1$. Finally, it can be proved that $\|x\|_1$ and $\|x\|_2$ are equivalent. □

Theorem 9. *Let* $\mathbf{T}$ *be the norm preserving isomorphic mapping from* $\mathbf{X}$ *to* $\mathbf{X}'$. *The time series* $x_i^{(b)}$, $y \in \mathbf{X} \subset \mathbb{R}^n$ *are transformed as* $x_i^{(b)'}$, $y' \in \mathbf{X}' \subset \mathbb{R}^n$ *by* $\mathbf{T}$. $\mathbf{X}$ *has the usual definition of the norm* $\|\cdot\|$, *with which the distance* $(\mathbf{X}, \rho)$ *can be derived by the norm. Similarly, the distance* $(\mathbf{X}', \rho')$ *also can be derived by the norm* $\|\cdot\|$. *Then, consider the following.*

(1) *In space* $\mathbf{X}$, *formulae (1), (2), (3), and (4) represent that the similarity measure function* $\mathrm{Sim} = f(x_i^{(b)}, y)$ *is equivalent.*

(2) *In spaces* $\mathbf{X}$ *and* $\mathbf{X}'$, *the similarity measure functions* $\mathrm{Sim} = f(*, *)$ *represented by* $\left\|x_i^{(b)} - y\right\|$ *and* $\left\|x_i^{(b)'} - y'\right\|$ *are equivalent.*

Proof. (1) Since formulae (1), (2), (3), and (4) are 2-norm, 1-norm, ∞-norm, and p-norm of the vector difference, respectively, according to Lemma 8, formulae (1), (2), (3), and (4) can be easily proved by norm axioms that they satisfy Definition 7. Thereby it is proved.

(2) It may be known from Definitions 3 and 5 that $\left\|x_i^{(b)} - y\right\| = \left\|(x_i^{(b)} - y)\right\| = \left\|\mathbf{T}x_i^{(b)} - \mathbf{T}y\right\|' = \left\|x_i^{(b)'} - y'\right\|$, and then it is easily proved by norm axioms that they satisfy Definition 7. □

3. Similarity Measure of Linear Multivariable Time Series

In previous sections, we mentioned the definitions of similarity of univariable time series. The signal variable time series is represented in vector form mathematically. The more similar the two vectors, the shorter the distance between them. The distance of two identical vectors should be zero. Thus, the similarity function of univariable time series should be defined based on vector norm. Multivariable time series is represented in matrix mathematically. So, the similarity function of multivariable time series should be defined based on matrix norm.

Definition 10. Matrix norms: if the real function $f(\mathbf{A}) = \|\mathbf{A}\|$, where $\mathbf{A}$ is any matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, satisfies the following:

(1) positive definiteness: $\|\mathbf{A}\| \ge 0$ and $\|\mathbf{A}\| = 0$, if and only if $\mathbf{A} = 0$,

(2) homogeneity: for any real number α, $\|\alpha\mathbf{A}\| = |\alpha| \|\mathbf{A}\|$,

(3) triangle inequality: for $\forall \mathbf{A} \in \mathbb{R}^{m \times n}$, $\|\mathbf{A} + \mathbf{B}\| \le \|\mathbf{A}\| + \|\mathbf{B}\|$,

(4) compatibility: for $\forall \mathbf{A} \in \mathbb{R}^{m \times n}$, $\|\mathbf{AB}\| \le \|\mathbf{A}\| \cdot \|\mathbf{B}\|$,

then it is said that $\|\mathbf{A}\|$ is a matrix norm in $\mathbb{R}^{m \times n}$, that is, the norm of matrix $\mathbf{A}$.

3.1. Commonly Used Matrix Norm. The commonly used matrix norms are given as follows:

(1) $\|\mathbf{A}\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{m} |a_{ij}|$, the maximum sum of absolute element of each column of matrix $\mathbf{A}$, and $\|\mathbf{A}\|_1$ is also called the column norm of $\mathbf{A}$;

(2) $\|\mathbf{A}\|_\infty = \max_{1 \le i \le m} \sum_{j=1}^{n} |a_{ij}|$, the maximum sum of absolute element of each row of matrix $\mathbf{A}$, and $\|\mathbf{A}\|_\infty$ is also called the row norm of $\mathbf{A}$;

(3) $\|\mathbf{A}\|_2 = \sqrt{\lambda_{\max}(\mathbf{A}^T\mathbf{A})}$, wherein $\mathbf{A}^T$ denotes matrix transpose of $\mathbf{A}$ and $\lambda_{\max}(\mathbf{A}^T\mathbf{A})$ is the maximum of absolute eigenvalue of $\mathbf{A}^T\mathbf{A}$, and $\|\mathbf{A}\|_2$ is also called the 2-norm of $\mathbf{A}$, or spectral norm;

(4) $\|\mathbf{A}\|_F = \left(\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2\right)^{1/2} = [\mathrm{tr}(\mathbf{A}^H\mathbf{A})]^{1/2}$, wherein $\mathbf{A}^H$ is conjugate transpose of $\mathbf{A}$. $\|\mathbf{A}\|_F$ is called the Frobenius norm (F-norm), which is similar to 2-norm in the form of vector, and also is compatible with the vector norm $\|x\|_2$. Its advantage in the norm invariance after F-norm is multiplied by unitary matrix, that is, the following theorem.

Theorem 11. *Let* $\mathbf{A} \in \mathbb{R}^{n \times m}$ *and* $\mathbf{U}, \mathbf{V}$ *be m-rank and n-rank unitary matrices, respectively, and then* $\|\mathbf{UA}\|_F^2 = \|\mathbf{A}\|_F^2 = \|\mathbf{AV}\|_F^2$.

Proof. $\|\mathbf{UA}\|_F^2 = t_r[(\mathbf{UA})^H(\mathbf{UA})] = t_r[\mathbf{A}^H\mathbf{U}^H\mathbf{UA}] = t_r[\mathbf{A}^H\mathbf{A}] = \|\mathbf{A}\|_F^2$. According to definition, we know that $\|\mathbf{A}\|_F = \left\|\mathbf{A}^H\right\|_F$. With the above results, we have $\|\mathbf{AV}\|_F^2 = \left\|\mathbf{AV}^H\right\|_F^2 = \left\|\mathbf{V}^H\mathbf{A}^H\right\|_F^2 = \left\|\mathbf{A}^H\right\|_F^2 = \|\mathbf{A}\|_F^2$.

Same as vector norm, the equivalence of matrix norms has the following similar conclusions. For $\forall \mathbf{A} \in \mathbb{R}^{n \times m}$, any matrix norms of $\mathbf{A}$ are equivalent. Mathematically, in the above four kinds of norms, $\|\mathbf{A}\|_1$ and $\|\mathbf{A}\|_\infty$ can be easily computed, and $\|\mathbf{A}\|_2$ and $\|\mathbf{A}\|_1$ have better properties and are widely applied. However, $\|\mathbf{A}\|_2$ is more complicated in engineering application and sensitive to the variation of matrix elements. $\|\mathbf{A}\|_F$ can be computed more easily and thus is widely applied. □

3.2. Definition of Similarity Functions of Multiple Variable Time Series. It is assumed that $\mathbf{X} = (X_1, X_2, \ldots, X_m)$ and $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m)$ are MTS, their samples are denoted as $x_i(t)$, $y_i(t)$, $(i = 1, 2, \ldots, m; t = 1, 2, \ldots, n)$, m is the number of observations, and n is the size of observations. Obviously, they are $n \times m$ matrices, which are denoted as $\mathbf{A}(n \times m)$ and $\mathbf{B}(n \times m)$. Their definition of similarity function is as follows:

$$\text{Sim} = f(\mathbf{X}, \mathbf{Y}) = \|\mathbf{A} - \mathbf{B}\|. \tag{8}$$

4. Similarity Measure of Nonlinear Time Series

We are inspired by support vector machines (SVMs), where a classifier can convert the difficult nonlinear classification problem in input space $\mathbf{X} \subset \mathbb{R}^m$ into simple linear classification problem in feature space $\mathbf{H}_k$ using kernel method. The essence is that the difficult classification of unclear similarity of the same class of samples and small difference of different classes of samples are converted from the input space into the feature space $\mathbf{H}_k$ in which the similarity of the same class of samples is enhanced and the difference of different classes of samples is enlarged. Thus, in this paper, the kernel method is introduced for the research of the similarity of nonlinear time series.

In input space $\mathbf{X} \subset \mathbb{R}^m$, the researching time series $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ is a m-dimension multivariable with n samples, and same as the researching series, the reference time series $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$ is a m-dimension multivariable with n samples. In the feature space $\mathbf{H}_k$, the nonlinear mapping of $\mathbf{X}$ and $\mathbf{Y}$ is denoted as $\Phi(\mathbf{X}) = (\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \ldots, \Phi(\mathbf{x}_m))$ and $\Phi(\mathbf{Y}) = (\Phi(\mathbf{y}_1), \Phi(\mathbf{y}_2), \ldots, \Phi(\mathbf{y}_m))$.

Definition 12. In the feature space $\mathbf{H}_k$, the inner product and norm are defined as follows:

$$\langle \Phi(\mathbf{X}), \Phi(\mathbf{Y}) \rangle = \sum_{i=1}^{n} \Phi(\mathbf{x}_i) \Phi(\mathbf{y}_i), \tag{9}$$

$$\|\Phi(\mathbf{X})\|_2 = \langle \Phi(\mathbf{X}), \Phi(\mathbf{Y}) \rangle^{1/2} = \left[\sum_{i=1}^{n} |\Phi(\mathbf{x}_i)|^2 \right]^{1/2}. \tag{10}$$

Definition 13. The similarity function of the researching series $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ and the reference time series $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$ are defined as follows:

$$\text{Sim} = f(\mathbf{X}, \mathbf{Y}) = \|\Phi(\mathbf{X}) - \Phi(\mathbf{Y})\|_2. \tag{11}$$

5. Discussion and Analysis

5.1. Analysis on the Similarity of Linear Transformed Univariable Time Series. According to operator theory, both Fourier transform and wavelet transform are linear operators. Thus, the relevant conclusions on the similarity of time series based on Fourier transform and wavelet transform are obtained as follows.

Lemma 14. *Let $f, g \in L^2(\mathbb{R}^n)$, and then Fourier transforms of f and g, $\widehat{f}(\omega), \widehat{g}(\omega) \in L^2(\mathbb{R}^n)$, have the following properties:*

(1) *invariant inner product* $(\widehat{f}, \widehat{g}) = (f, g)$,

(2) *norm preserving* $\|\widehat{f}\| = \|f\|$.

Proof. Consider the following.

(1)

$$\begin{aligned}
(f, g) &= \int_{-\infty}^{\infty} f^*(t)\, g(t)\, dt \\
&= \int_{-\infty}^{\infty} f^*(t) \left[\int_{-\infty}^{\infty} \widehat{g(\omega)} \exp(j\omega t)\, d\omega \right] dt \\
&= \int_{-\infty}^{\infty} \widehat{g(\omega)} \left[f^*(t) \exp(j\omega t)\, dt \right] d\omega \\
&= \int_{-\infty}^{\infty} \widehat{g(\omega)} \left[f(t) \exp(-j\omega t)\, dt \right]^* d\omega \\
&= \int_{-\infty}^{\infty} \widehat{g(\omega)}\, \widehat{f(\omega)}^*\, d\omega = \left(\widehat{f}, \widehat{g}\right).
\end{aligned} \tag{12}$$

(2) For the above equation, let $f = g$, and then there is $\|\widehat{f}\| = \sqrt{(\widehat{f}, \widehat{f})} = \sqrt{(f, f)} = \|f\|$. □

Inference 1. Time series $x_i^{(b)}, y \in \mathbf{X} \subset \mathbb{R}^n$ are transformed as $x_i^{(b)'}, y' \in \mathbf{X}'$ by discrete Fourier transform (DFT). The similarity functions $\text{Sim} = f(*, *)$ is represented by norms $\|x_i^{(b)} - y\|$ and $\|x_i^{(b)'} - y'\|$ are equivalent.

Proof. According to Theorem 9 and Lemma 14, the results are directly deduced. □

Analogously, the following conclusions are also obtained. Due to the limited space, the proof is not presented here.

Inference 2. Time series $x_i^{(b)}, y \in \mathbf{X} \subset \mathbb{R}^n$ are transformed as $x_i^{(b)'}, y' \in \mathbf{X}'$ by discrete wavelet transform (DWT). The similarity function $\text{Sim} = f(*, *)$ represented by norms $\|x_i^{(b)} - y\|$ and $\|x_i^{(b)'} - y'\|$ is equivalent.

5.2. Practice Computing Method of Similarity of Linear Multivariable Time Series. Let matrices $\mathbf{A}_{n\times m}$ and $\mathbf{B}_{n\times m}$ represent MTS $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m)$ and $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_m)$, and let the covariance matrices between the columns in $\mathbf{A}_{n\times m}$ and $\mathbf{B}_{n\times m}$ be $\text{COV}(\mathbf{A})_{m\times m}$ and $\text{COV}(\mathbf{B})_{m\times m}$. All eigenvalues of covariance matrices $\text{COV}(\mathbf{A})_{m\times m}$ are arranged in descending order $\lambda_{\alpha 1}, \lambda_{\alpha 2}, \ldots, \lambda_{\alpha m}$ and their corresponding standard orthogonal eigenvectors are $\alpha_1, \alpha_2, \ldots, \alpha_m$. Similarly, all eigenvalues of covariance matrices $\text{COV}(\mathbf{B})_{m\times m}$ are arranged in descending order $\lambda_{\beta 1}, \lambda_{\beta 2}, \ldots, \lambda_{\beta m}$, and the corresponding standard orthogonal eigenvectors are $\beta_1, \beta_2, \ldots, \beta_m$. Thus, the eigenvectors of covariance matrices $\text{COV}(\mathbf{A})_{m\times m}$ and $\text{COV}(\mathbf{B})_{m\times m}$ are $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_m]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \ldots, \beta_m]$, respectively, and then their extended similarity function is defined as follows:

$$\text{Sim} = f(\mathbf{X}, \mathbf{Y}) = \|\boldsymbol{\alpha} - \boldsymbol{\beta}\|. \tag{13}$$

According to principle component analysis, physical meaning of formula (13) is equivalent to that the distance (similarity) between the linear orthogonal transforms of $\mathbf{A}_{n\times m}$ and $\mathbf{B}_{n\times m}$ can be measured using the norms in

the transformed space. Yang and Shahabi [13, 14] use the extended Frobenius norm to compute the similarity of MTS which is a special case of formula (13). Formula (13) can be understood by referring to the definition of similarity function of univariable time series transformed by linear operator.

5.3. Practical Computation of Similarity of Nonlinear Multivariable Time Series. Mathematically, it is not difficult to prove that the norm of formula (10) derived by the inner product of formula (9) satisfies the Parallelogram formula: $\|\Phi(\mathbf{X}) + \Phi(\mathbf{Y})\|^2 + \|\Phi(\mathbf{X}) - \Phi(\mathbf{Y})\|^2 = 2(\|\Phi(\mathbf{X})\|^2 + \|\Phi(\mathbf{Y})\|^2)$. Since $\|\Phi(\mathbf{X}) - \Phi(\mathbf{Y})\|_2^2 = \langle\Phi(\mathbf{X}), \Phi(\mathbf{X})\rangle - 2\langle\Phi(\mathbf{X}), \Phi(\mathbf{Y})\rangle + \langle\Phi(\mathbf{Y}), \Phi(\mathbf{Y})\rangle$, it does not need to know the mapping function explicitly. Instead, the kernel function $K(\mathbf{x}, \mathbf{y})$ is used to compute the similarity function defined in formula (11); that is,

$$\text{Sim} = f(\mathbf{X}, \mathbf{Y}) = \sqrt{K(\mathbf{x}, \mathbf{x}) - 2K(\mathbf{x}, \mathbf{y}) + K(\mathbf{y}, \mathbf{y})}. \quad (14)$$

Formula (14) shows that the nonlinear similarity measure of nonlinear time series in input space can be linearly measured in feature space through kernel method. Currently, commonly used kernel functions are as follows:

(1) linear kernel function:

$$K(\mathbf{x}, \mathbf{y}) = \mathbf{x}\mathbf{y}, \quad (15)$$

(2) p-order polynomial kernel function:

$$K(\mathbf{x}, \mathbf{y}) = (\mathbf{x}\mathbf{y} + 1)^p, \quad (16)$$

(3) Gaussian radial basis RBF kernel function:

$$K(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \quad (17)$$

(4) neural network kernel function:

$$K(\mathbf{x}, \mathbf{y}) = \tanh(v(\mathbf{x}\mathbf{y}) + c). \quad (18)$$

6. Experimental Results

In our experiments, we take the nonlinear multivariate time series as a typical example without loss of generality for showing the accuracy of the proposed similarity measurement method. The similarity measurement function of formula (1) is used in the input space, and the similarity measurement function of formula (14) is used in the feature space. The Gaussian radial basis RBF kernel function of formula (17) is used for mapping the nonlinear samples to high-dimension space, which is low complexity compared with polynomial kernel function, especially the high rank kernel function. All samples in the original test and training sets are well mixed. The samples are randomly selected to construct the new test and training sets. Because the KNN classifier can decide a sample belong to the class in which more of k nearest samples are contains. Based on the proposed similarity measure function to verify the equivalence of time series in different spaces, KNN and kernel KNN classifiers are employed to classify the samples in the new sets in input space and feature space, respectively. Each random experiment is carried out 20 times. Then the classification accuracies in the experiments are analyzed and compared. Moreover, the parameter σ of Gaussian radial basis of RBF kernel is optimized in the experiments. The σ is initialized as a small value first. Subsequently, each random experiment for determining each value of σ is repeatedly carried out 20 times. The σ is increasingly adjusted according to the average classification accuracy of each experiment until the optimal value of σ is obtained. In our experiments, five published datasets, Cylinder-Bell-Funnel (CBF), Fish, Face (four) [18], *Iris*, and Wine [19], are tested.

Presently, there were already some researching work on the classification of time series. One of the most common benchmark datasets is CBF [20] which was used by [21–23]. CBF dataset consists of three time series, Cylinder $c(t)$, Bell $b(t)$, and Funnel $f(t)$, which are generated by the following equations:

$$\begin{aligned} c(t) &= (6 + \eta) X_{[a,b]}(t) + \varepsilon(t), \\ b(t) &= \frac{(6 + \eta) X_{[a,b]}(t)(t - a)}{(b - a)} + \varepsilon(t), \\ f(t) &= \frac{(6 + \eta) X_{[a,b]}(t)(b - a)}{(b - t)} + \varepsilon(t), \\ X_{[a,b]} &= \{1, \text{if } a \le t \le b, \text{else } 0\}, \end{aligned} \quad (19)$$

where η and $\varepsilon(t)$ are drawn from a standard normal distribution $N(0, 1)$, a is an integer drawn uniformly from the range $[16, 32]$, and $(b - a)$ is an integer drawn uniformly from the range $[32, 96]$. The three typical curves, Cylinder, Bell, and Funnel, are shown in Figures 1(a)–1(c), respectively. The curve of classification accuracy relative to σ is shown in Figure 1(d) and the classification accuracy rates of KNN and kernel KNN classifiers are also shown in Figure 1(e). It can be seen from these experimental results that the similarity of time series in original and feature spaces, which are measured with the proposed generalized similarity function defined by the distance of vector or matrix norm, is equivalent regardless of whether the time series are linearly or nonlinearly transformed. The same conclusion is obtained from the experimental results in Figures 2, 3, 4, and 5. Moreover, the average classification accuracies of each experiment repeatedly tested 20 times for five datasets are listed in Table 1, which also confirm our conclusion.

7. Conclusion

Based on set theory, metric space theory, operator theory, matrix theory, and kernel method, the definition of similarity of time series is extended and unified theoretically to establish a generalized framework for the similarity measure of time series. In the generalized framework, the similarity of time series is defined as the distance of unified vector/matrix

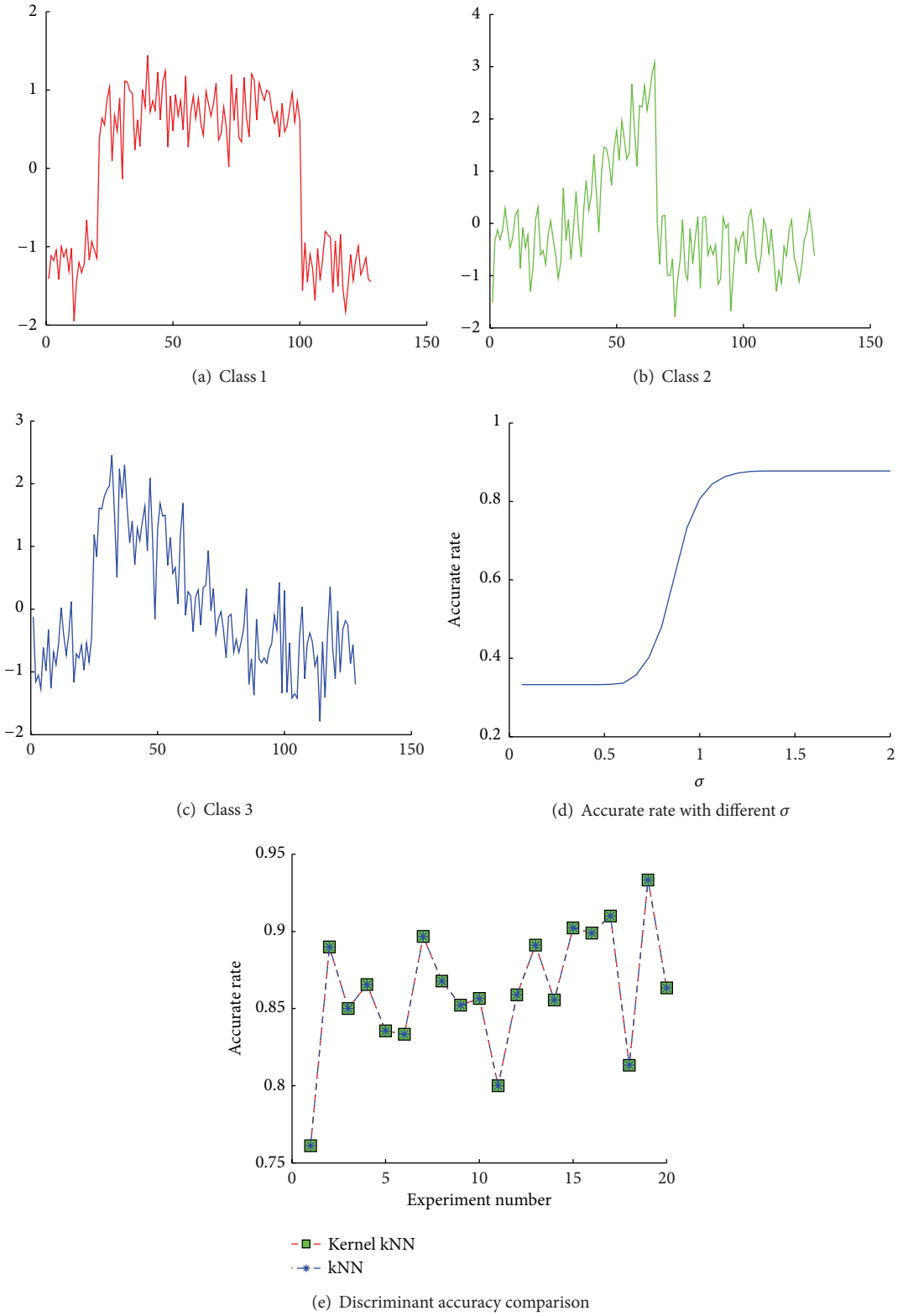


FIGURE 1: Test results of dataset CBF.

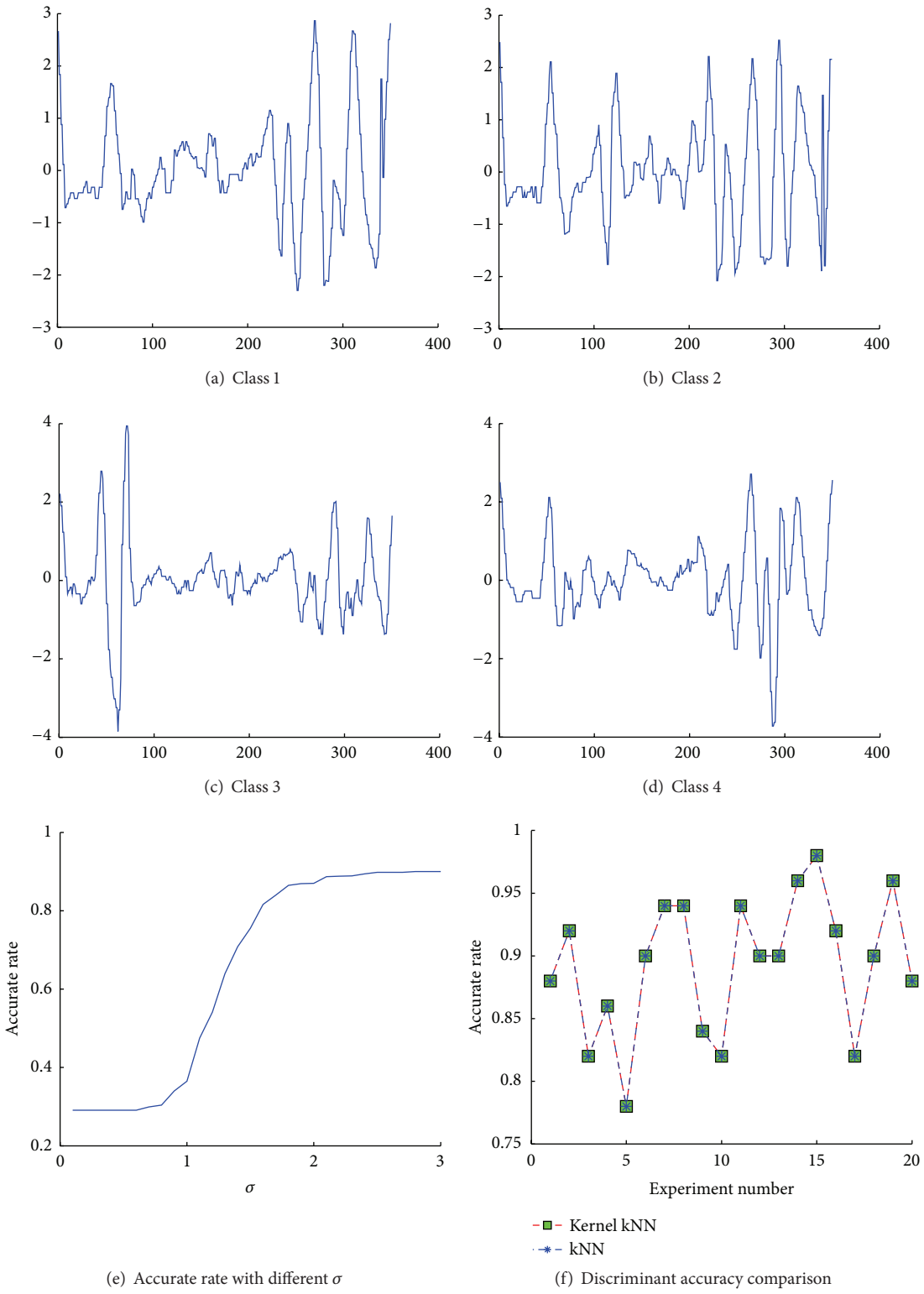


Figure 2: Test results of dataset Face (four).

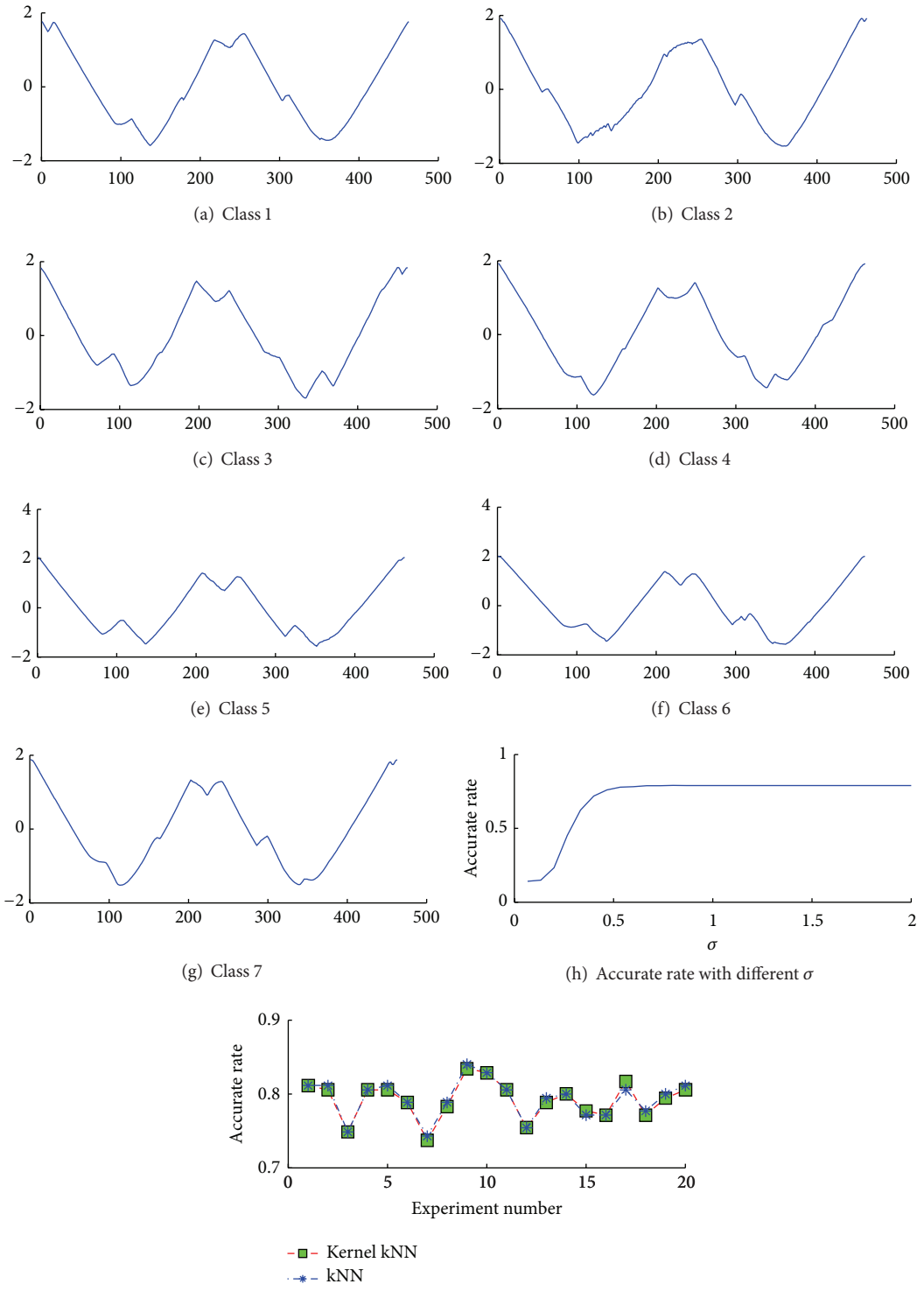
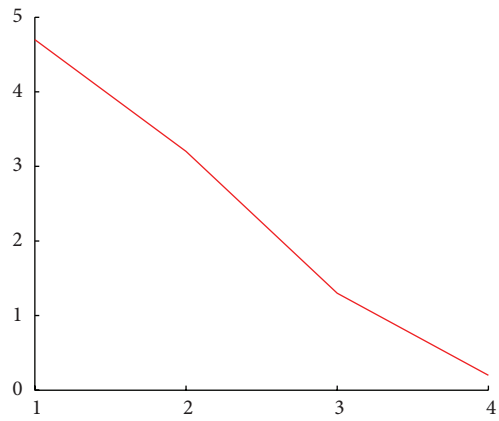
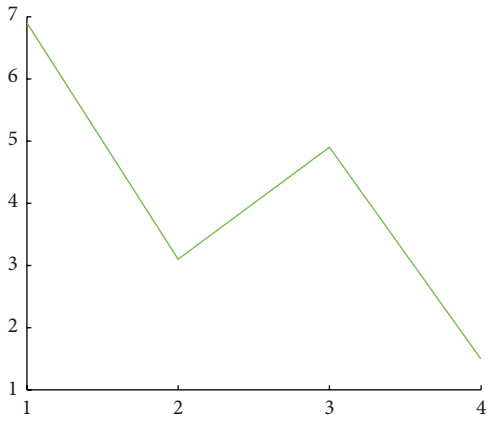


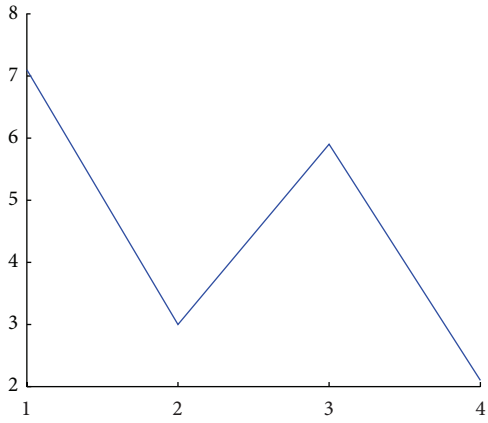
Figure 3: Test results of dataset Fish.



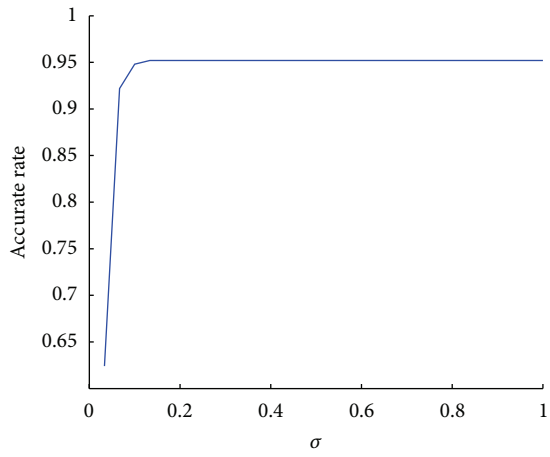
(a) Class 1: *Iris setosa*



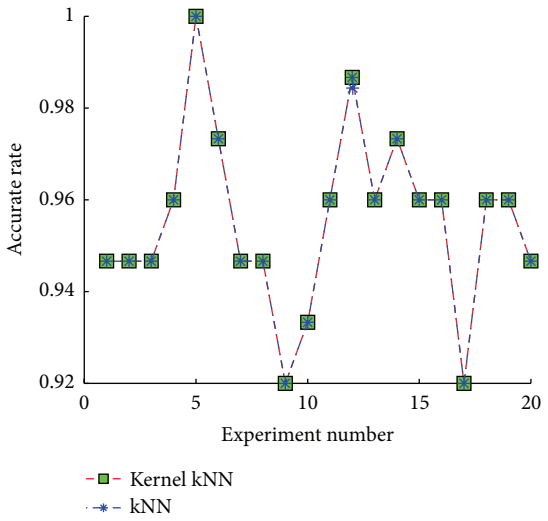
(b) Class 2: *Iris versicolor*



(c) Class 3: *Iris virginica*



(d) Accurate rate with different σ



(e) Discriminant accuracy comparison

Figure 4: Test results of dataset *Iris*.

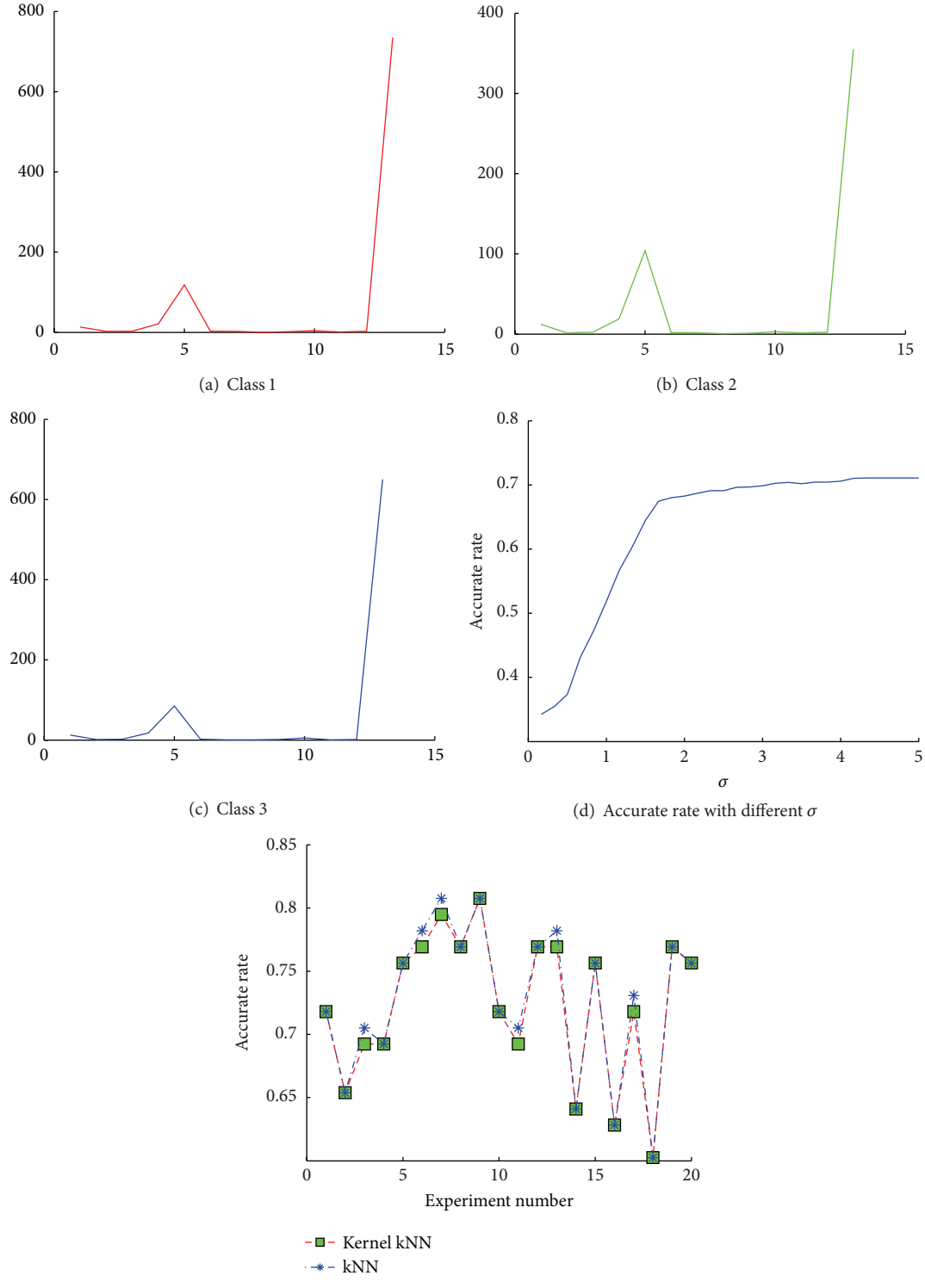


FIGURE 5: Test results of dataset Wine.

norm, which is suitable for both time series of univariable and multiple variables in any linear transformed space or nonlinear transformed space. The proposed similarity definition has been proven to be equivalent in original space and transformed space. The experimental results on some published time series datasets confirm that the theoretical deduction on the generality of similarity measure of time series defined in this paper is right.

Table 1: Discriminant accuracy comparison.

Datasets	Average accurate rate	
	KNN	Kernel KNN
CBF	0.862	0.862
Fish	0.793	0.792
Iris	0.953	0.953
Wine	0.728	0.724
Face (four)	0.893	0.893

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This work is partly supported by the National Natural Science Foundation of China 61379100 and 61472388.

References

[1] C. Faloutsos, M. Ranganathan, and Y. Manolopoulos, "Fast Subsequence matching in time-series databases," in *Proceedings of the ACM International Conference on Management of Data*, vol. 23, pp. 419–429, 1994.

[2] R. Agrawal, C. Faloutsos, and A. Swami, "Efficient similarity search in sequence databases," in *Proceedings of the International Conference on Foundations of Data Organization and Algorithms*, pp. 69–84, Chicago, Ill, USA, 1993.

[3] D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," *Knowledge Discovery and Data Mining Workshop*, vol. 10, no. 16, pp. 359–370, 1994.

[4] E. Keogh and C. A. Ratanamahatana, "Exact indexing of dynamic time warping," *Knowledge and Information Systems*, vol. 7, no. 3, pp. 358–386, 2005.

[5] D. J. Berndt and J. Clifford, "Finding patterns in time series: a dynamic programming approach," in *Advances in Knowledge Discovery and Data Mining*, pp. 229–248, American Association for Artificial Intelligence, Menlo Park, Calif, USA, 1996.

[6] M. Vlachos, G. Kollios, and D. Gunopulos, "Discovering similar multidimensional trajectories," in *Proceedings of the 18th International Conference on Data Engineering*, pp. 673–684, San Jose, Calif, USA, March 2002.

[7] L. Chen, M. T. Ozsu, and V. Oria, "Robust and fast similarity search for moving object trajectories," in *Proceedings of the ACM International Conference on Management of Data (SIGMOD '05)*, pp. 491–502, Baltimore, Md, USA, 2005.

[8] L. Chen and R. T. Ng, "On the marriage of L_p-norms and edit distance," in *Proceedings of the 30th International Conference on Very Large Data Bases*, vol. 30, pp. 792–803, Toronto, Canada, 2004.

[9] Y. Chen, M. A. Nascimento, B. C. Ooi, and A. K. H. Tung, "SpADe: on shape-based pattern detection in streaming time series," in *Proceedings of the 23rd International Conference on Data Engineering (ICDE '07)*, pp. 786–795, Istanbul, Turkey, April 2007.

[10] E. Frentzos, K. Gratsias, and Y. Theodoridis, "Index-based most similar trajectory search," in *Proceedings of the 23rd International Conference on Data Engineering (ICDE '07)*, pp. 816–825, Istanbul, Turkey, April 2007.

[11] M. D. Morse and J. M. Patel, "An efficient and accurate method for evaluating time series similarity," in *Proceedings of the ACM International Conference on Management of Data*, pp. 569–580, 2007.

[12] J. Aßfalg, H.-P. Kriegel, P. Kröger, P. Kunath, A. Pryakhin, and R. Renz, "Similarity search on time series based on threshold queries," in *Advances in Database Technology—EDBT 2006*, vol. 3896 of *Lecture Notes in Computer Science*, pp. 276–294, Springer, Berlin, Germany, 2006.

[13] K. Yang and C. Shahabi, "A PCA-based similarity measure for multivariate time series," in *Proceedings of the 2nd ACM International Workshop on Multimedia Databases (MMDB '04)*, pp. 65–74, Washington, D.C., USA, November 2004.

[14] K. Yang and C. Shahabi, "An efficient k nearest neighbor search for multivariate time series," *Information and Computation*, vol. 205, no. 1, pp. 65–98, 2007.

[15] J.-W. Xu, A. R. Paiva, I. Park, and J. C. Principe, "A reproducing kernel Hilbert space framework for information-theoretic learning," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 5891–5902, 2008.

[16] J.-W. Xu, P. P. Pokharel, A. R. C. Paiva, and J. C. Príncipe, "Nonlinear component analysis based on correntropy," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '06)*, pp. 1851–1855, Vancouver, Canada, July 2006.

[17] S. Liu and H. Jiang, "Study of the conception of the similarity in time series," *Journal of Huazhong University of Science and Technology*, vol. 32, no. 7, pp. 75–79, 2004 (Chinese).

[18] The UCR Time Series Data Mining Archive, http://www.cs.ucr.edu/~eamonn/time_series_data/dataset.zip.

[19] UCI Machine Learning Repository, http://archive.ics.uci.edu/ml/.

[20] N. Saito, "Local feature extraction and its application us ing a library of bases," in *Topics in Analysis and Its Applications: Selected Theses*, pp. 265–451, 2000.

[21] S. Manganaris, *Supervised classiffication with temporal data [Ph.D. thesis]*, Computer Science Department, School of Engineering, Vanderbilt University, 1997.

[22] M. W. Kadous, "Learning comprehensible descriptions of multivariate time series," in *Proceedings of the 16th International Machine Learning Conference*, pp. 454–463, Morgan Kaufmann, 1999.

[23] J. J. R. Diez and C. A. Gonzalez, "Applying boosting to similarity literals for time series classiffication," in *Proceedings of the 1st International Workshop on Multiple Classifier Systems*, pp. 210–219, 2000.