On the Stationarity of Multivariate Time Series for Correlation-Based Data Analysis

Kiyoung Yang and Cyrus Shahabi
Computer Science Department
University of Southern California
Los Angeles, CA 90089-0781
[kiyoungy,shahabi]@usc.edu

Abstract

Multivariate time series (MTS) data sets are common in various multimedia, medical and financial application domains. These applications perform several data-analysis operations on large number of MTS data sets such as similarity searches, feature-subset-selection, clustering and classifications. Correlation-based techniques, such as Principal Component Analysis (PCA), have proven to improve the efficiency of many of the above-mentioned data-analysis operations on MTS, which implies that the correlation coefficients concisely represent the original MTS items. In this paper, we propose to utilize the stationarity of the MTS data sets, in order to represent the original MTS items more stably, as well as concisely with the correlation coefficients. That is, before performing any correlation-based data analysis, we first executes the stationarity test to decide whether the MTS item is stationary or not, i.e., whether the correlation is stable or not. Subsequently, for a non-stationary MTS data set, we difference it to render the data set stationary. Even though our approach is general, to focus the discussion we describe our approach within the context of our previously proposed techniques for MTS variable-subset-selection and similarity search. In order to show the validity of our approach, we performed several experiments on four real-world data sets. The results show that the performances of our variable-subset-selection and similarity search techniques have significantly improved in terms of classification accuracy and precision/recall.

1 INTRODUCTION

A time series is a series of observations, $x_i(t); [i = 1, \cdots, n; t = 1, \cdots, m]$, made sequentially through time where i indexes the measurements made at each time point t [32]. It is called a univariate time series (UTS) when n is equal to 1, and a multivariate time series (MTS) when n is equal to, or greater than 2. A UTS item is usually represented in a vector of size m, while each MTS item is typically stored in an $m \times n$ matrix, where m is the number of observations and n is the number of variables (e.g., sensors).

MTS data sets are common in various fields, such as in multimedia, medicine and finance. For example, in multimedia, Cybergloves used in the Human and Computer Interface (HCI) applications have around 20 sensors, each of which generates 50$\sim$100 values in a second [18, 29]. For gesture recognition and video sequence matching using computer vision, several features are extracted from each image continuously, which renders them MTSs [6, 1, 28]. In medicine, Electro Encephalogram (EEG) from 64 electrodes placed on the scalp are measured to examine the correlation of genetic predisposition to alcoholism [36]. Functional Magnetic Resonance Imaging (fMRI) from 696 voxels out of 4391 has been used to detect similarities in activation between voxels in [13].

An MTS item is typically very high dimensional. For example, an MTS item from one of the data sets used in the experiments in Section 4 contains 3000 observations with 64 variables. If a traditional distance metric for similarity search, e.g., Euclidean Distance, is to be utilized, this MTS item would be considered as a 192000 (3000 $\times$ 64) dimensional data. 192000 dimensional data would be overwhelming not only for the distance metric, but also for indexing techniques. To the best of our knowledge, there has been no attempt to index data sets with more than 100000 dimensions/features[1]. Hence, instead of using this high dimensional MTS data set as is, a number of techniques have been proposed to represent the MTS data set concisely for data mining processes, such as classification and similarity search [24, 22, 30, 33, 35]. For example, in [22], an

[1]In [2], the authors employed MVP-tree to index 65536 dimensional gray-level MRI images.

MTS item that contains an EEG signal with 39 channels is decomposed into multiple univariate time series. Each univariate time series is subsequently transformed into 3 autoregressive (AR) coefficients. Hence, each MTS item is represented with 117 (3×39) features, after which feature subset selection is performed. In [30], Principal Component Analysis (PCA) Similarity Factor (S_{PCA}) [20] is employed for the similarity measure between two MTS items. That is, in order to compute the similarity between two MTS items, they first compute the correlation coefficient matries[2] of the two MTS items, and then decompose them via Singular Value Decomposition (SVD) to obtain the principal components. Consequently, they measure how similar the corresponding principal components (PCs) from the two MTS items are.

For MTS analysis, e.g., similarity search and feature subset selection, it has been empirically shown that the correlation information among the variables plays an important role, and the correlation-based techniques, such as PCA, perform well [30, 24, 33]. In [34], each MTS item is represented with the upper triangle elements of the correlation coefficient matrix. The performance of feature subset selection using the correlation coefficients is shown to outperform the one using the AR coefficients which does not consider the correlation information among the variables. In [24], the similarity between two MTS items is defined to be the inner product of the first principal components from the two MTS items with the variances explained by the principal components as weights. The so defined similarity measure has been successfully applied to recognize and segment the multi-attribute motions.

To recapitulate, the correlation-based techniques utilize the correlation coefficients to represent the original MTS items for data mining tasks. The good performance of correlation-based techniques hence implies that the correlation coefficients concisely represents the original MTS items in a dimension-reduced form. Note that, given an MTS item $\mathbf{A}$ of size $m \times n$, it is typically the case that $m \gg n$. For example, an MTS item from one of the data sets used in Section 4 contains 3000 observations, while there are only 64 variables. In general, the size of a correlation coefficient matrix for $\mathbf{A}$ is much smaller than that of $\mathbf{A}$.

In this paper, we propose to utilize the *stationarity* of time series in order to better represent the MTS items with the correlation coefficients. Intuitively, a time series is defined to be *stationary* if the statistical properties of the time series, e.g., the mean and the correlation coefficients, do not change over time. Hence, if an MTS item is found to be stationary, the correlation information of the MTS item does not change over time, which would make the correlation based representations of the original MTS items more *stable*. For example, assume that we are given an MTS item $x_i(t)$ of size $m \times n$, i.e., where $1 \leq i \leq n$ and $1 \leq t \leq m$. From $x_i(t)$, consider two matrices of size $(m-1) \times n$, i.e., $x_i^1(t)$ where $1 \leq t \leq m-1$ and $x_i^2(t)$ where $2 \leq t \leq m$. If $x_i(t)$ is stationary, then the correlation coefficient matrices of $x_i^1(t)$ and $x_i^2(t)$, i.e., $Corr(x_i^1(t))$ and $Corr(x_i^2(t))$ would be statistically *not* different. However, if $x_i(t)$ is *non-stationary*, then $Corr(x_i^1(t))$ and $Corr(x_i^2(t))$ would be statistically different. This example implies that for non-stationary MTS items, the correlation coefficients change *statistically significantly* depending on just one observation out of, e.g., 3000 observations, which is not the case for stationary MTS items. Therefore, we firstly test the stationarity of each MTS items in the database. Subsequently, we determine the stationarity of the MTS data set based on the majority of stationarities of MTS items in the data set. If the MTS data set turns out to be non-stationary, we *stationarize* all the MTS items in the data set before we perform correlation-based data analysis.

In order to evaluate the effectiveness of the proposed approach, we conducted several experiments on four real-world data sets: AUSLAN [18] obtained from UCI KDD repository [15], the Brain Computer Interface (BCI) data set [21], the BCI at the Max Planck Institute (MPI) data set [22] and EEG data set [15]. In order to find out the stationarity of the MTS data set, Johansen's co-integration test [16] has been firstly performed. AUSLAN and BCI are found to be stationary, while BCI MPI and EEG are non-stationary. Even though our approach is general, to focus the discussion we describe our approach within the context of our previously proposed techniques for MTS variable-subset-selection and similarity search. The performances depending on the stationarity of the data set have been compared in terms of the classification accuracy using Corona [34] and the precision/recall using Eros [33]. The results show that the performance improves up to 10% in classification accuracy and up to 24% in precision/recall.

The remainder of this paper is organized as follows. Section 2 discusses the background of our proposed approach. Our proposed approach is presented in Section 3, which is followed by the experiments and results in Section 4. Conclusions and future work are presented in Section 5.

2 BACKGROUND

Our proposed approach utilizes the stationarity of an MTS item. In this section, we briefly describe the stationarity and how to test the stationarity of a time series. For details, please refer to [16, 8, 9, 11].

[2] PCA may employ either the correlation coefficient matrix or the covariance matrix for a given MTS item. We would assume that the correlation coefficient matrix is utilized for PCA, since correlation-based analysis is more frequently encountered [17].

2.1 Stationarity

In this section, we describe the stationarity of a time series. In order to test the stationarity of a time series, the Unit Root test is performed for a univariate time series, and the Co-integration test is utilized for a multivariate time series, which are described in Section 2.2 and in Section 2.4, respectively.

Definition 1 *A time series is said to be strictly stationary if the joint distribution of $X(t_1), \cdots, X(t_n)$ is the same as the joint distribution of $X(t_1+\tau), \cdots, X(t_n+\tau)$ for all $t_1, \cdots, t_n, \tau$, where $X(t)$ denotes the random variable at time t. In other words, shifting the time origin by an amount τ has no effect on the joint distributions, which must therefore depend only on the intervals between $t_1, t_2, \cdots, t_n$. The above definition holds for any value of n [4].*

Intuitively, a time series is *strict-sense stationary* if all of its statistical properties are invariant with respect to shifts of the time origin [25]. Since many of the properties of stationary processes depend only on the structure of the process as specified by its first and second moments [4], the weaker definition of stationarity is generally used, which is defined as follows [4]:

Definition 2 *In practice, it is often useful to define stationarity in a less restricted way than Definition 1. A process is called second-order stationary (or weakly stationary) if its mean is constant and its auto-covariance function depends only on the lag, i.e., τ, so that*

$$E[X(t)] = \mu$$

and

$$Cov[X(t), X(t+\tau)] = \gamma(\tau)$$

If $\tau = 0$, the second-order stationarity implies that both the variance and the mean are constant.

Figure 1 represents examples of stationary and non-stationary univariate time series from BCI MPI data set [22] used in Section 4. The dashed line in the middle is the mean of the time series. As can be seen in Figure 1(a), if the time series is stationary, the values moves irregularly away from its mean, but eventually *reverts* to its mean, while as in Figure 1(b), if the time series is non-stationary, the graph shows some *trend*.

To reiterate, a stationary time series is one whose statistical properties such as mean, variance and correlation, are all constant over time. Hence, the correlation coefficients of stationary time series would be more *stable* than those of non-stationary time series for the concise representation of the original MTS item, which would also affect the performance of the subsequent correlation-based data analysis processes.

2.2 Unit Root Test

For a univariate time series, the Unit Root test is frequently employed for testing the stationarity. The test first poses the *null hypothesis* that the given time series has the unit root, which means that the time series is non-stationary, and tests if the null hypothesis is to be statistically accepted or rejected in favor of *alternative hypothesis* that the given time series is stationary.

The Unit Root test estimates the following autoregressive model

$$x(t) = \rho x(t-1) + \epsilon_t \tag{1}$$

where $x(t)$ is an observation value at time t, and ϵ_t is a sequence of independent normal random variables with mean 0 and variance σ^2 [8]. The time series $x(t)$ converges, as $t \to \infty$, to a stationary time series if $|\rho| < 1$. If $|\rho| = 1$, the time series is not stationary and the variance of $x(t)$ is $t\sigma^2$ [8]. The Unit Root test subsequently tests the following one-sided hypothesis

$$H_0 : \rho = 1$$
$$H_1 : \rho < 1$$

The name, unit root, comes from the fact that the coefficient of $x(t-1)$ is unity, if the time series is non-stationary, and the Unit Root tests, as the name suggests, tests if ρ is unity or not. Note that the autoregressive model in Equation (1) is a simple form; the most common unit root test [27], the augmented Dickey-Fuller (ADF) test, employs an autoregressive model that contains the *trend* and *drift*. For details, please refer to [27].

2.3 Differencing

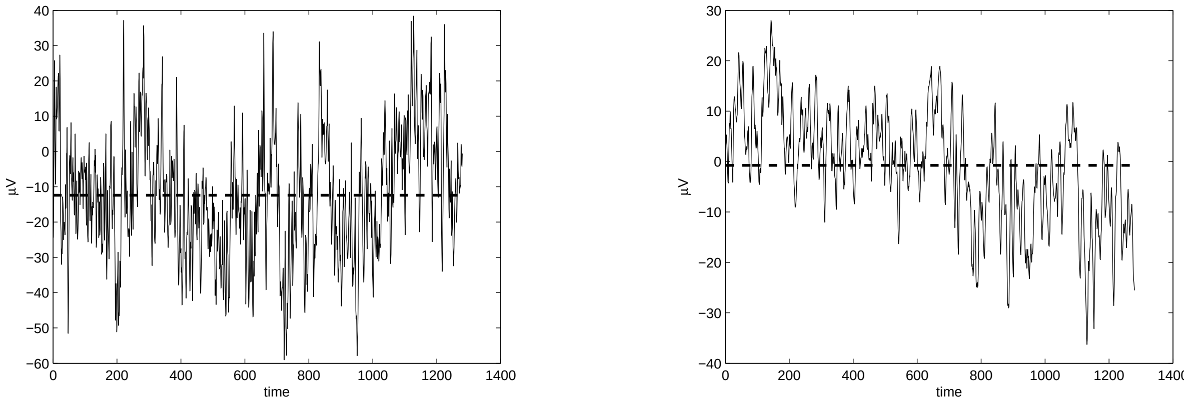
A simple transformation of non-stationary time series into a stationary one can be achieved by taking the first differences of time series $(x(t) - x(t-1))$.

Definition 3 *Differencing is a special type of filtering, which is particularly useful for removing a trend. It is simply to difference a given time series until it becomes stationary [4]. For non-seasonal data, first-order differencing is usually sufficient to attain apparent stationarity, so that the new series $y(1), \cdots, y(m-1)$ is formed from the original series $x(1), \cdots, x(m)$ by*

$$y(t) = x(t+1) - x(t) = \nabla x(t+1)$$

Figure 2 shows that after differencing, i.e., the *trend* is removed, the same non-stationary data as in Figure 1(b) is transformed into a stationary time series.

In Section 4, if the MTS data set is found to be non-stationary, we perform the differencing to render the data set as stationary. Since the duration of each MTS data from



(a) Stationary **(b) Non-stationary**

Figure 1. Examples of stationary and non-stationary univariate time series

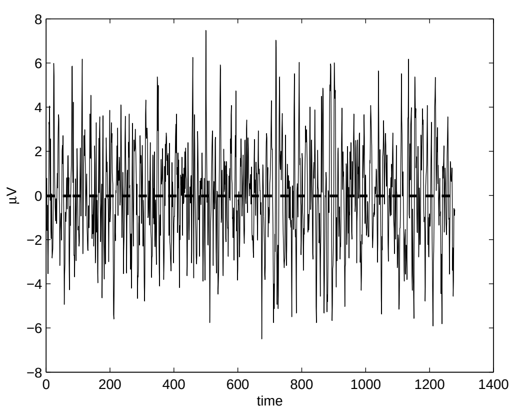


Figure 2. The same non-stationary data as Figure 1(b) after differencing

the MTS data sets used in Section 4 is less than 1 minute, the first-order differencing would be sufficient to stationarize the non-stationary data sets. Occasionally second-order differencing is required using the operator ∇^2, where

$$\nabla^2 x_{t+2} = \nabla x_{t+2} - \nabla x_{t+1} = x_{t+2} - 2x_{t+1} + x_t$$

For data sets with seasonal trends, however, different techniques should be employed. Please refer to [4] for details.

2.4 Co-integration Test

In order to determine the stationarity of a multivariate time series, we need to consider the following two cases:

- If all the univariate time series in an MTS item are stationary, then the MTS item is stationary.
- If some or all of the univariate time series in an MTS item are non-stationary, we need to perform the co-integration test to make sure that the MTS item is non-stationary.

Definition 4 *A univariate time series $x(t)$ is said to be integrated of order d, written I(d), if it needs to be differenced d times to make it stationary. If two series $x_1(t)$ and $x_2(t)$ are both I(d), then any linear combination of the two series will usually be I(d) as well. However, if a linear combination exists for which the order of integration is less than d, say I(d - b), then the two series are said to be co-integrated of order (d, b), written CI(d, b). If this linear combination can be written in the form $\alpha^T x_i(t)$, where $x_i^T(t) = (x_1(t), x_2(t))$, then the vector α is called a co-integrating vector [4].*

That is, co-integration means that one or more linear combinations of non-stationary time series is stationary even

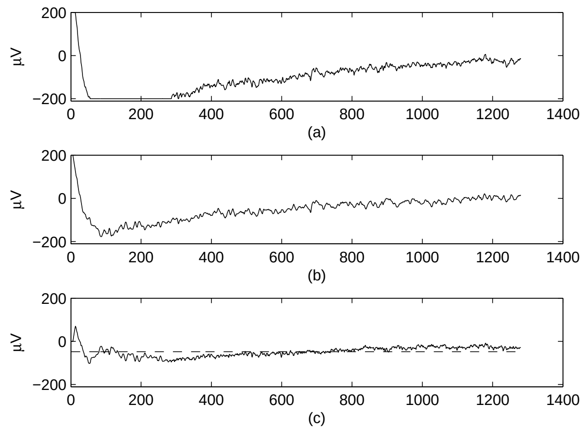


Figure 3. Two non-stationary time series, (a) and (b), that are co-integrated, (c) = (a) - (b).

though individually they are not. If the time series are co-integrated, they cannot move *too far* away from each other [9].

For example, Figures 3(a) and (b) are two non-stationary time series, and Figure 3(c) is Figure 3(a) - Figure 3(b), which is stationary. Hence, Figures 3(a) and (b) are co-integrated and the co-integrating vector α is $(1, -1)$.

There are a number of co-integration tests. As in [9], Johansen's test [16] is chosen since it is based on the well-accepted likelihood ratio principle, and performs better than other methods.

3 THE PROPOSED APPROACH

Before performing any correlation-based data analysis for multivariate time series, we propose to render the data set as stationary, if necessary, as in Algorithm 1. That is, we firstly determine the stationarity for all the MTS items in the data set by performing the Johansen's test. The Johansen's Co-integration test has been performed using the Econometrics Toolbox [23]. Note that we do not utilize the ADF test for each univariate time series, since we are not interested in the stationarity of each UTS in an MTS item. Besides, as in [10], it requires a strategy that deals with three cases to appropriately apply the ADF test to an UTS, which seems to be rather cumbersome. Moreover, it has been shown that ADF test has low power, failing to reject the unit root hypothesis in many cases [26, 7].

As in Line 3 of Algorithm 1, we extract the number of co-integrating relationships, γ, where $1 \leq \gamma \leq n$. As long as γ is equal to the number of variables of an MTS item, n, then we consider the MTS item as stationary (Lines 4∼8). The stationarity of an MTS data set is subsequently determined by the majority of the stationarities of the MTS items. The $sum(H)$ in Line 10 yields the number of non-stationary items in the data set. Hence, if the number of non-stationary items is greater than the half of the number of total items in the data set, the data set is determined to be non-stationary, and each MTS item in the data set is *first-order differenced* into a stationary item.

As described in Section 2.1, if we make sure that the data set is stationary, the original MTS item can be more *stably* represented, as well as *concisely*, with correlation coefficients.

Table 1. Summary of data sets used in the experiments

	AUSLAN	BCI	BCI MPI	EEG
# of variables	22	64	39	64
average length	60	3000	1280	256
# of labels	95	2	2	2
# of items per label	27	139	1000	6980/3908
total # of items	2565	278	2000	10888

Algorithm 1 Determine the stationarity of an MTS data set

Require: MTS data set, N {the number of items in the data set}, n {the number of variables in an MTS item}

1: **for** i = 1 to N **do**
2: $res \leftarrow$ Johansen's test;
3: $\gamma \leftarrow$ extract the number of co-integrating relationships from res;
4: **if** $\gamma = n$ **then**
5: $H(i) \leftarrow 0$; {stationary}
6: **else**
7: $H(i) \leftarrow 1$; {non-stationary}
8: **end if**
9: **end for**
10: **if** $sum(H) > ceil(N/2)$ **then**
11: Stationarize the given MTS data set;
12: **end if**

4 PERFORMANCE EVALUATION

In order to evaluate the effectiveness of the proposed approach, we conducted experiments on four real-world data sets. Even though our approach is general, to focus the discussion we describe our approach within the context of our previously proposed techniques for MTS variable-subset-selection and similarity search. The performances depending on the stationarity of the data set have been compared in terms of the classification accuracy using Corona [34] and the precision/recall using Eros [33].

4.1 Datasets

The experiments have been conducted on four different real-world data sets, i.e., AUSLAN, BCI, BCI MPI and EEG, which are all labeled MTS data sets whose labels are given.

The **Australian Sign Language (AUSLAN) data set** uses 22 sensors on the hands to gather the data sets generated by signing of a native AUSLAN speaker [18]. It contains 95 distinct signs, each of which has 27 examples. In total, the number of signs gathered is 2565. The average length is around 60.

The **Brain Computer Interface (BCI) data set** [21] was collected during the BCI experiment, where a subject had to perform imagined movements of either the left small finger or the tongue. The time series of the electrical brain activity was collected during these trials using 64 ECoG platinum electrodes. All recordings were gathered at 1000Hz. The total number of data items is 278 and the average length is 3000.

The **Brain Computer interface (BCI) data set** at the **Max Planck Institute (MPI)** [22] was gathered to examine the relationship between the brain activity and the motor imagery, i.e., the imagination of limb movements. Eight right handed male subjects participated in the experiments, out of which three subjects were filtered out after pre-analysis. 39 electrodes were placed on the scalp to record the EEG signals at the rate of 256Hz. The total number of items is 2000, i.e., 400 items per subject.

The **Electroencephalogram (EEG) data set** [15] was collected to examine EEG *correlates* of genetic predisposition to alcoholism. This data set contains EEG recordings of control and alcoholic subjects. 64 electrodes placed on the scalp were used to measure the EEG recordings at the rate of 256Hz. Each data item has 256 samples, and the total number of items is 10888[3].

Table 1 shows the summary of the data sets used in the experiments.

4.2 Methods

In order to evaluate our proposed approach, we applied Corona and Eros to the data set with and without differencing. For a stationary data set, we compare the performances of Corona and Eros without differencing to those with differencing, and see how they change. For a non-stationary data set, we compare the performances of Corona and Eros with differencing to those without differencing, and see how much they improve. The Johansen's Co-integration test has been performed with a significance level of 5% using the Econometrics Toolbox [23].

[3]Out of 11057 items, 169 erroneous items that contain variables with all zero values have been removed.

Table 2. Stationarity Test Results

Dataset	Co-integration Test	
	# of stationary items	# of non-stationary items
AUSLAN	1311	792
BCI	233	45
BCI MPI	309	1691
EEG	954	9934

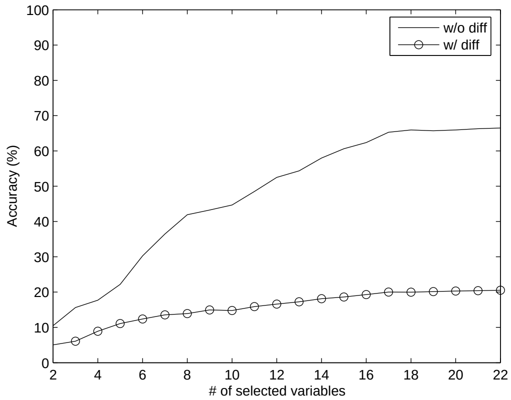
Corona is a supervised features subset selection technique for MTS data sets. Each MTS data is firstly represented as correlation coefficients, which are subsequently transformed into a vector. After training a Support Vector Machine (SVM) on the transformed data, Corona recursively eliminates one variable that contributes the least to the separating hyperplane. For Corona, an SVM with linear kernel was adopted as the classifier. For each subset of variables selected by Corona, we performed 10-fold cross validation [14]. After repeating the cross validation 10 times, the average classification accuracy was computed for each selected subset of variables. SVM classification is completed with *LIBSVM* [3]. *mean* aggregating function has been employed for *Corona*.

Eros is a similarity measure for MTS data sets, which is based on Principal Component Analysis (PCA). Instead of comparing two MTS data represented in matrices element-by-element, Eros computes the similarity between two MTS data by measuring how similarity two corresponding principal components (PCs) are using the aggregated eigenvalues as weight. For Eros, we performed modified leave-one-out kNN search as in [33]. For simplicity, we chose 10 for $maxr$. Hence, in recall-precision graphs, the ranges of x and y axes are from 0.1 to 1.0 and from 0 to 1.0, respectively. Recall that each data set used in the experiments has more than 10 relevant items as shown in Table 1. For example, AUSLAN has 95 labels and each label has 27 items. The recall-precision graph [12] is then plotted, which has been frequently used to measure the performance of Content Based Image Retrieval (CBIR) systems [31, 19] as well as Information Retrieval (IR) systems. *mean* aggregating function on the raw eigenvalues has been used for the weight vector w of *Eros*, again, for simplicity.

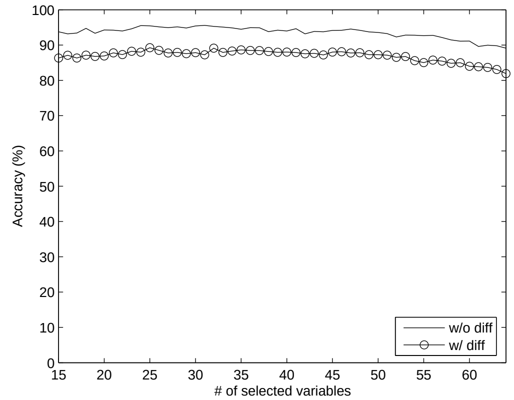
4.3 RESULTS

4.3.1 Stationarity

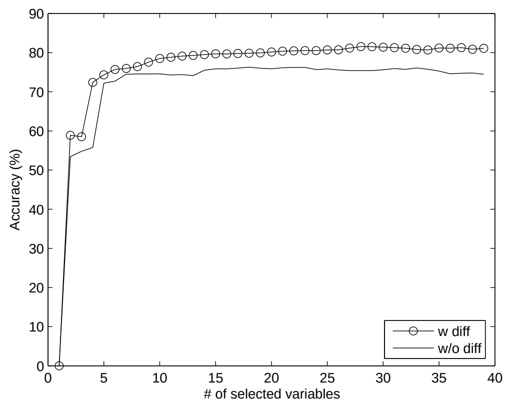
Table 2 summarizes the results of Johansen's Co-integration test on the four data sets. AUSLAN and BCI are determined to be stationary, while BCI MPI and EEG are non-stationary. Note that BCI, BCI MPI and EEG are all collected using the electrodes attached to the scalp, but their stationarities are different. In the literature, it is widely as-



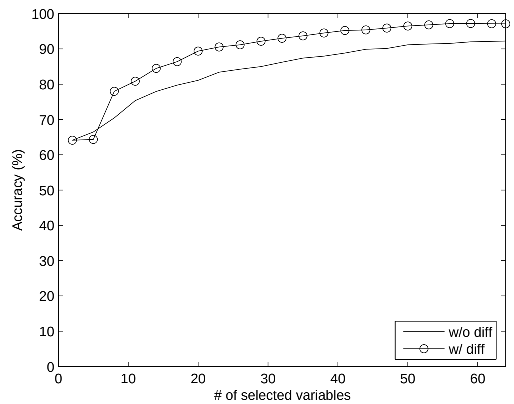
(a) AUSLAN (stationary)



(b) BCI (stationary)



(c) BCI MPI (non-stationary)



(d) EEG (non-statinoary)

Figure 4. Classification Accuracy

sumed that EEG signals are non-stationary. However, in [5], it is suggested that EEG epochs shorter than 12 seconds may be considered stationary. Though all the EEG epochs employed for experiments are shorter than 12 seconds, two out of three data sets turned out to be non-stationary, and the other is stationary.

Note that the average of *rank* of the AUSLAN data set is 18.4698 out of 22. Only 162 items out of 2565 are full rank. Johansen's Co-integration test initially failed for the 2403 non-full-rank items of the AUSLAN data set. In order to make the items full rank, we removed the UTS whose standard deviation is near 0. The result of AUSLAN shown in Table 2 is acquired after this preprocessing.

4.3.2 Accuracy and Precesion/Recall

Figures 4 and 5 depict the classification accuracy using Corona and the precision/recall using Eros, respectively. For AUSLAN, BCI, BCI MPI and EEG, the better performance corresponds to the stationarities of the data sets. That is, if the data set is stationary, the better performance is obtained without differencing, and if the data set is non-stationary, the better performance is achieved with differencing.

The performance improvements between with and without differencing are up to 10% in terms of classification accuracy, and are up to 24% in terms of precision/recall. Hence, our proposed approach to utilizing stationarity is well justified.

Note that if a stationary time series is again differenced, the resultant time series is also stationary. Given this fact, one may be attempted to perform the first-order differencing without firstly testing the stationarity of the time series. However, for the stationary data sets, i.e., AUSLAN and BCI, the performances of Eros and Corona are worse when the *redundant* first-order differencing is performed, which indicates that the test of stationarity is required as a preprocessing step.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed to render the given MTS dataset as stationary, if necessary, before performing correlation-based data analysis, such as PCA. Based on the stationarity, the correlation coefficients represent the original MTS items more *stably* as well as concisely. Empirically, we have shown that if the given dataset is non-stationary, making the dataset stationary improves the performances of the correlation-based data analysis in terms of classification accuracy and precision/recall.

We intend to extend this technique to the stream of data where the determination of the stationarity as well as the subsequent data analysis processes, such as, feature subset selection, can be performed incrementally adjusting itself based on the observations collected thus far.

Acknowledgement

This research has been funded in part by NSF grants EEC-9529152 (IMSC ERC), IIS-0238560 (PECASE) and IIS-0307908, and unrestricted cash gifts from Microsoft. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] J. Alon, S. Sclaroff, G. Kollios, and V. Pavlovic. Discovering clusters in motion time-series data. In *IEEE Computer Vision and Pattern Recognition*, pages 18–20, June 2003.

[2] T. Bozkaya and M. Ozsoyoglu. Indexing large metric spaces for similarity search queries. *ACM TODS*, 24(3), 1999.

[3] C.-C. Chang and C.-J. Lin. Libsvm – a library for support vector machines. http://www.csie.ntu.edu.tw/∼cjlin/libsvm/, 2004.

[4] C. Chatfield. *The Analysis of Time Series: An Introduction*. Chapman & Hall/CRC, 2003.

[5] B. A. Cohen and A. S. Jr. Stationarity of the human electroencephalogram. *Med. Biol. Eng. Comput.*, 15:513–518, 1977.

[6] A. Corradini. Dynamic time warping for off-line recognition of a small gesture vocabulary. In *IEEE RATFG*, pages 82–89, July 2001.

[7] D. N. DeJong, J. C. Nankervis, N. E. Savin, and C. H. Whiteman. Integration versus trend stationary in time series. *Econometrica*, 60(2):423–433, March 1992.

[8] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427–431, June 1979.

[9] D. A. Dickey, D. W. Jansen, and D. L. Thornton. A primer on cointegration with an application to money and income. *Federal Reserve Bulletin, Federal Reserve Bank of St. Louis*, pages 58–78, March/April 1991.

[10] J. Elder and P. E. Kennedy. Testing for unit roots: What should students be taught? *Journal of Economic Education*, 31(2):137–146, 2001.

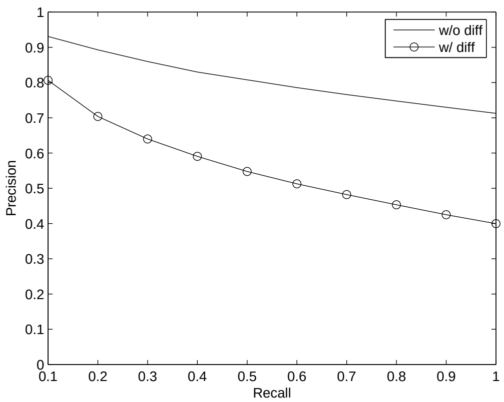
[11] R. F. Engle and C. W. J. Granger. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–276, March 1987.

[12] W. B. Frakes and R. Baeza-Yates. *Information Retrieval: Data Structures and Algorithms*. Prentice-Hall, 1992.

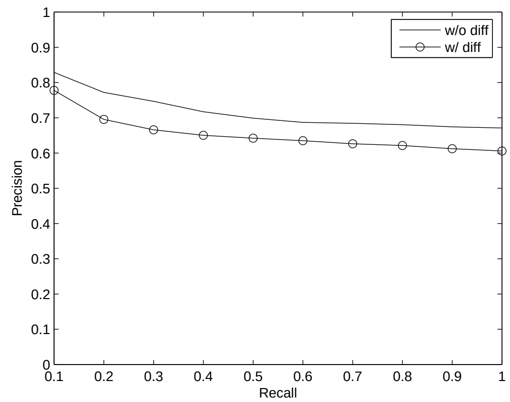
[13] C. Goutte, P. Toft, E. Rostrup, F. A. Nielsen, and L. K. Hansen. On clustering fMRI time series. *NeuroImage*, 9(3):298–310, 1999.

[14] J. Han and M. Kamber. *Data Mining: Concepts and Techniques*, chapter 3, page 121. Morgan Kaufmann, 2000.

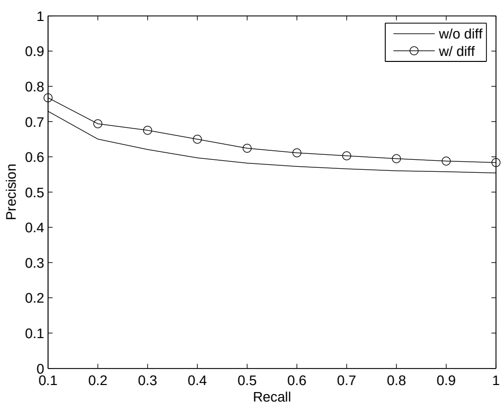
[15] S. Hettich and S. D. Bay. The UCI KDD Archive. *http://kdd.ics.uci.edu*, 1999.



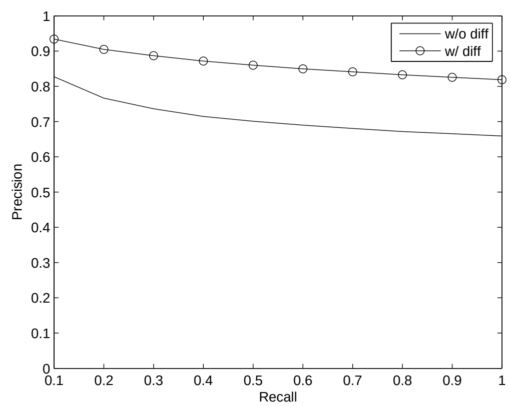
(a) AUSLAN (stationary)



(b) BCI (stationary)



(c) BCI MPI (non-stationary)



(d) EEG (non-stationary)

Figure 5. Precision/Recall

[16] S. Johansen. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, 1995.

[17] I. T. Jolliffe. *Principal Component Analysis*. Springer, 2002.

[18] M. W. Kadous. *Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series*. PhD thesis, University of New South Wales, 2002.

[19] D.-H. Kim and C.-W. Chung. Qcluster: relevance feedback using adaptive clustering for content-based image retrieval. In *Proc. of the ACM SIGMOD Conference*, 2003.

[20] W. Krzanowski. Between-groups comparison of principal components. *JASA*, 74(367), 1979.

[21] T. N. Lal, T. Hinterberger, G. Widman, M. Schröder, N. J. Hill, W. Rosenstiel, C. E. Elger, B. Schölkopf, and N. Birbaumer. Methods towards invasive human brain computer interfaces. In *Advances in Neural Information Processing Systems 17*, pages 737–744. Cambridge, MA, 2005.

[22] T. N. Lal, M. Schröder, T. Hinterberger, J. Weston, M. Bogdan, N. Birbaumer, and B. Schölkopf. Support vector channel selection in BCI. *IEEE Trans. Biomed. Eng.*, 51(6), June 2004.

[23] J. P. LeSage. Econometrics toolbox. http://www.spatial-econometrics.com/, 2005.

[24] C. Li, P. Zhai, S.-Q. Zheng, and B. Prabhakaran. Segmentation and recognition of multi-attribute motion sequences. In *ACM MM '04*, pages 836–843, New York, NY, USA, 2004.

[25] T. K. Moon and W. C. Stirling. *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.

[26] P. Perron. The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6):1361–1401, 1989.

[27] P. C. B. Phillips and Z. Xiao. A primer on unit root testing. *Journal of Economic Surveys*, 12(5):423–469, 1998.

[28] C. Rao, A. Gritai, M. Shah, and T. Syeda-Mahmood. View-invariant alignment and matching of video sequences. In *IEEE ICCV*, pages 939–945, October 2003.

[29] C. Shahabi. AIMS: An immersidata management system. In *VLDB CIDR*, January 2003.

[30] A. Singhal and D. Seborg. Clustering of multivariate time-series data. In *Proc. of the American Control Conference*, volume 5, 2002.

[31] R. O. Stehling, M. A. Nascimento, and A. X. Falcão. A compact and efficient image retrieval approach based on border/interior pixel classification. In *Proc. of the ACM-CIKM Conference*, 2002.

[32] A. Tucker, S. Swift, and X. Liu. Variable grouping in multivariate time series via correlation. *IEEE Trans. Syst., Man, Cybern. B*, 31(2):235–245, 2001.

[33] K. Yang and C. Shahabi. A PCA-based similarity measure for multivariate time series. In *MMDB '04: Proceedings of the 2nd ACM international workshop on Multimedia databases*, pages 65–74, Washington, DC, USA, 2004. ACM Press.

[34] K. Yang, H. Yoon, and C. Shahabi. A supervised feature subset selection technique for multivariate time series. In *International Workshop on Feature Selection for Data Mining: Interfacing Machine Learning with Statistics (FSDM) in conjunction with 2005 SIAM International Conference on Data Mining (SDM'05)*, Newport Beach, CA, April 2005.

[35] H. Yoon, K. Yang, and C. Shahabi. Feature subset selection and feature ranking for multivariate time series. *IEEE Trans. Knowledge Data Eng. - Special Issue on Intelligent Data Preparation*, 17(9), September 2005.

[36] X. L. Zhang, H. Begleiter, B. Porjesz, W. Wang, and A. Litke. Event related potentials during object recognition tasks. *Brain Research Bulletin*, 38(6):531–538, 1995.