Chapter 3

Subspace based transformation for univariate timeseries

In the previous chapter, we discussed the problem of anomaly detection for time series data and categorized the existing techniques based on their methodology and the way they transform the data prior to anomaly detection. In this chapter we discuss a novel transformation technique for univariate time series which uses concepts from the window based scheme discussed in section 2.7.1. We present the motivation behind this transformation and the methodology in the following sections.

3.1 Motivation

In the discussion of window based techniques (section 2.7.1), we suggested that, once the univariate test and training time series are divided into windows, any traditional anomaly detection technique for multivariate data can be applied to these sets of windows. The reason for this application is that if a normal univariate time series is generated by a process, the windows of the normal time series also follow a generative process in the multivariate space. Hence the set of fixed-length windows obtained from a univariate time series can be considered as a multivariate time series. Univariate anomalous time series follow a different process or do not follow any process; hence the windows obtained from them need not follow the same process as the windows obtained from normal time series.

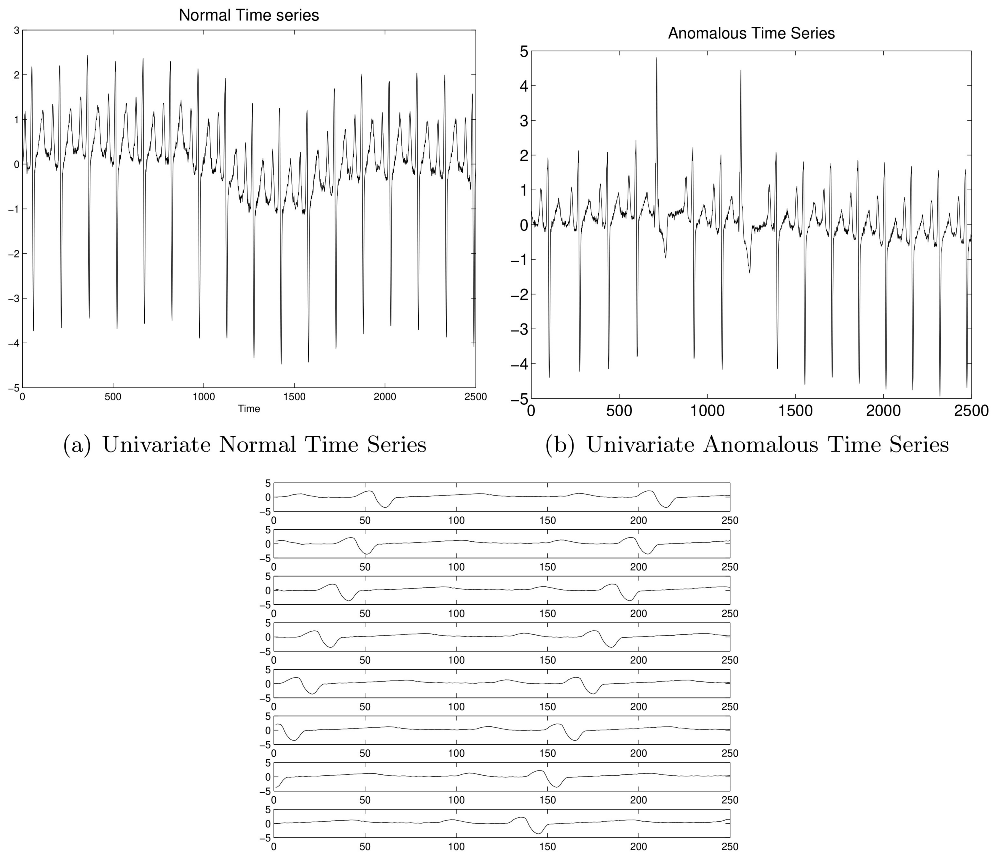
We explain this with an example. Consider the normal time series in Figure 3.1(a). The first few windows of it are shown in Figure 3.1(c) (obtained with parameters: window length $n = 250$ and number of steps (hop) $h = 10$). We consider each of the time instance of a window as a variable, i.e., if the window length is n then the multivariate time series has n variables. In Figure 3.2 we show the time series corresponding to a sample of these variables. Figure 3.2(a) shows the time series corresponding to time instances (variables) 50 , 150 and 200 of the windows obtained from normal time series in Figure 3.1(a). We will refer to these variables as normal variables. Similarly Figure 3.2(b) shows the time series corresponding to same time instances of the windows obtained from the anomalous time series in Figure 3.1(b). As can be seen, the time series of normal variables follow a periodic pattern, while the time series of anomalous variables does not follow the periodic pattern. Hence these windows from univariate time series can be considered as multivariate time series as they follow a specific process in the multivariate space.

With this motivation we proceed to present a brief overview of multivariate time series and anomaly detection techniques in this domain in section 3.2.

3.2 Detecting Anomalies in Multivariate Time Series

Data collected in many application domains consists of multivariate time series. For example, a single aircraft generates different univariate time series, which are considered as a single multivariate time series. Each of these univariate time series correspond to data coming from a single sensor or a switch on the aircraft (73). Similarly the daily network log stored for network intrusion detection depicts a multivariate time series, where each variable measures certain aspects of the time series (74).

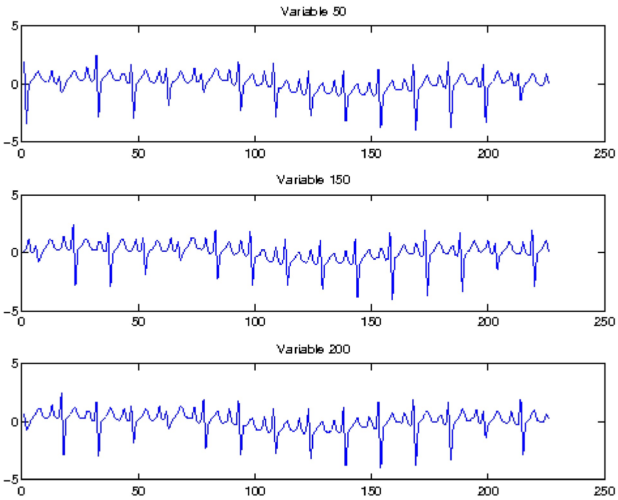
Anomaly detection for multivariate time series data is distinct from traditional anomaly detection for multivariate data as well as anomaly detection for univariate time series data. This is because the anomaly detection techniques for multivariate data analyse only the multivariate aspect of the data, while the techniques for univariate time series data analyse the sequence aspect of each variable independently. Often, the anomaly in a multivariate time series can be detected only by analyzing the sequence of all (or a subset of) variables.



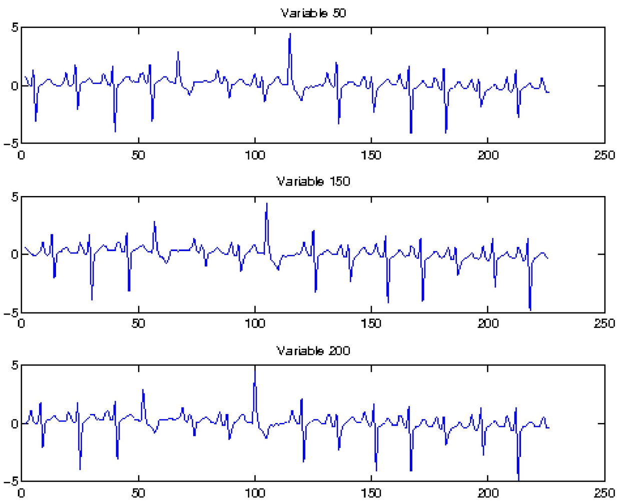
(a) Univariate Normal Time Series (b) Univariate Anomalous Time Series

(c) First Windows

Figure 3.1: Windows of Univariate Time series



(a) Variables of Normal Time Series



(b) Variables of Anomalous Time Series

Figure 3.2: Time series corresponding to normal and anomalous variables

A real life instance of anomalies in multivariate time series data can be found in aviation safety domain (73). During the course of the flight of an aircraft, multiple sensors measure different aspects of the flight, thereby generating a multivariate time series. An fault or accident during the flight of the aircraft can be detected by finding anomalies in the multivariate time series data. To detect such anomalies in multivariate sequences, multivariate aspect as well as the sequence aspect needs to be simultaneously modeled.

A multivariate anomaly detection technique proposed by Chandola (75), is a window based technique which accounts for both the multivariate and sequence aspects of the data while detecting anomalies. The key underlying idea is to reduce a multivariate time series into a univariate time series by exploring the change in the correlation structure of the time series, using subspace monitoring. We discuss subspace monitoring in section 3.2.1 and the anomaly detection technique for multivariate time series using subspace monitoring, WIN_{SS} in section 3.2.2.

3.2.1 Subspace Monitoring for Multivariate Time Series

Subspace monitoring for damage detection and other related areas falls under the broad purview of *statistical process control* (76; 77). Jordan et al (78) proposed the original concept of comparing two subspaces using the angles between their principal components, which was later explained as *canonical correlation* by Hotelling (79). Chandola et al (75) provide a detailed discussion of the angle between subspaces using principal angles.

Comparison of multivariate time series often use subspace based analysis, so as to compare their generative models such as vector AR models (80), ARMA models (81), and linear dynamical systems (76). It has been shown that the modes and modal shapes of vibrating structures (76) coincide with their system eigenstructure. Any change in this structure can be computed using covariance driven identification methods which use the change in subspaces.

The anomaly detection technique proposed by Chandola et al (75) measures the change between two subspaces, S_A and S_B, which is defined as the following (77): *Change in subspace when S_A changes to subspace S_B is equal to the maximum distance between an arbitrary unit vector $\hat{x}$ in S_B and the subspace S_A.* This change δ_{AB} is given as:

$$\delta_{AB} = \sqrt{1 - \lambda_{min}} \tag{3.1}$$

where λ_{min} is the minimum eigen value of $B^T(AA^T)B$. A and B are the m-dimensional basis vectors of subspaces A and B, respectively. For proof see (77).

3.2.2 Converting a Multivariate Time Series to Univariate Time Series

Chandola et al (75) propose a novel anomaly detection technique, WIN_{SS} which consists of two steps. Each training and test multivariate time series is converted into a univariate time series in the first step and the second step involves using an existing univariate anomaly detection technique on the transformed univariate time series to detect anomalies.

Let $S \in \Re^{|S|\times m}$ be a multivariate time series of length $|T|$ and consisting of m variables.

A w length window of S starting at time t is denoted as $W_t = S[t]S[t+1]\ldots S[t+w-1]$. Thus $T - w + 1$ such windows are extracted from S. Consider two successive windows $W_t, W_{t+1} \in \Re^{w\times m}$. V_t denotes the subspace spanned by the top *few*[1] principal components of W_t. Similarly V_{t+1} denotes the subspace spanned by the top *few* principal components of W_{t+1}. Note that W_t and W_{t+1} can have a different number of basis vectors. The change between the two successive subspaces, $\delta_{t,t+1}$ can be defined using 3.1 as:

$$\delta_{t,t+1} = \sqrt{1 - \lambda_{min}} \tag{3.2}$$

where λ_{min} is the minimum eigenvalue of the matrix $V_t^T V_{t-1} V_{t-1}^T V_t$. Thus the multivariate time series S can be transformed into a univariate time series $\delta_{1,2}\delta_{2,3}\ldots$.

The transformed univariate time series captures the dynamics of the subspace structure of a multivariate time series, such that the normal time series are expected to follow similar dynamics, while anomalous time series are different. Any traditional univariate anomaly detection technique can be applied to this transformed time series. Chandola et al (75) use a window based anomaly detection technique in the second step.

[1] Capturing α% of the total variance.

3.3 Transformation

Section 3.2.2 describes how the anomaly in the multivariate space can be captured using the subspace based analysis by exploring the correlation structure of the multivariate time series. With this motivation, we proceed further to apply the subspace analysis to the windows obtained from univariate time series which can be considered as multivariate time series. The process of this subspace based transformation is : converting a univariate time series to a multivariate time series and back to another univariate time series.

3.3.1 Methodology

For a given univariate time series T of length N, $T = t_1 t_2 \ldots t_N$, overlapping windows of length n are created. Let w_i be the i^{th} window which can be represented as $t_i t_{i+1} \ldots t_{i+n-1}$. These windows are stacked and are considered as multivariate time series, W, represented as :

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{N-n+1} \end{pmatrix}$$

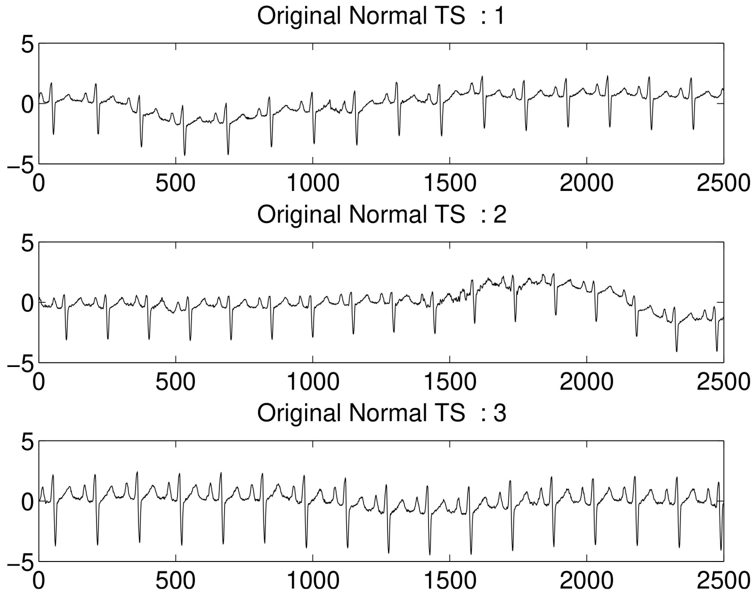
$$= \begin{pmatrix} t_1 & t_2 & \ldots & t_n \\ t_2 & t_3 & \ldots & t_{n+1} \\ \vdots & \vdots & \vdots & \vdots \\ t_{N-n+1} & t_{N-n+2} & \ldots & t_N \end{pmatrix}$$

Each time instance of a window is considered as a variable, so the columns correspond to time series of different variables. The time series corresponding to k^{th} variable is $t_k t_{k+1} t_{k+2} \ldots$ and the length of the multivariate time series is the number of windows extracted. As discussed in section 2.7.1, the windows from a univariate time series can be obtained by sliding one or more steps at a time, also called hop (h). In the above case we considered $h = 1$. If $h < n$, the windows are overlapping. For any general hop h, the time series corresponding to k^{th} variable is $t_k t_{k+h} t_{k+2h} \ldots$.

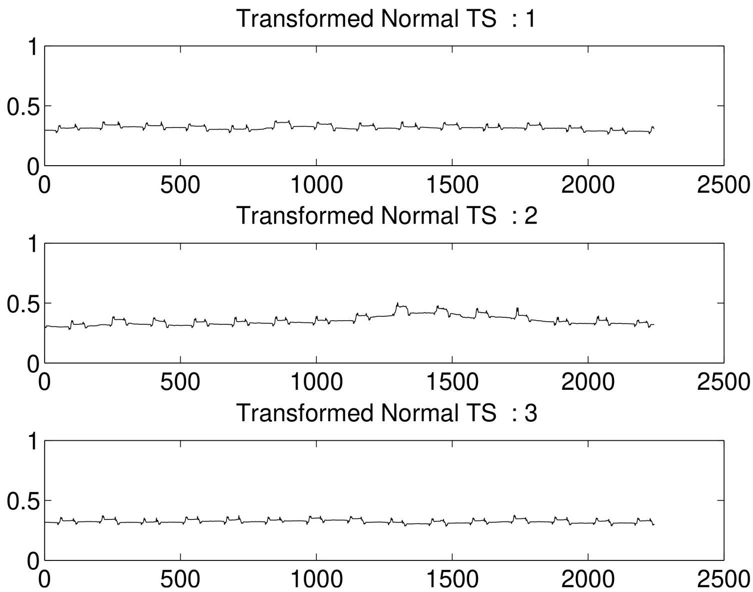
Each training and test time series in the database is converted to a multivariate time series which is further transformed to another univariate time series by using the subspace based transformation described in 3.2.2. This approach explores the change in correlation structure of the variables and the sequence aspect of a multivariate time series. Overlapping windows of a multivariate time series are created and the change in subspaces of consecutive windows is computed to obtain the univariate time series. Once the transformation is applied on all the time series, the existing univariate anomaly detection techniques can be applied.

Figure 3.3 shows a set of univariate normal time series and their corresponding transformations. As can be seen the transformed time series capture the periodicity of the original time series well. That is the angle between consecutive subspaces keeps repeating with similar periodicity.

Figure 3.4 shows a set of univariate anomalous time series and their corresponding transformations. The transformed anomalous time series do not have similar characteristics as transformed normal time series. The change in subspaces corresponding to two normal windows is different from the change in subspaces corresponding to two anomalous windows, which highlights the existence of anomaly. The properties of the transformation are clearly discussed in the Chapter 4.

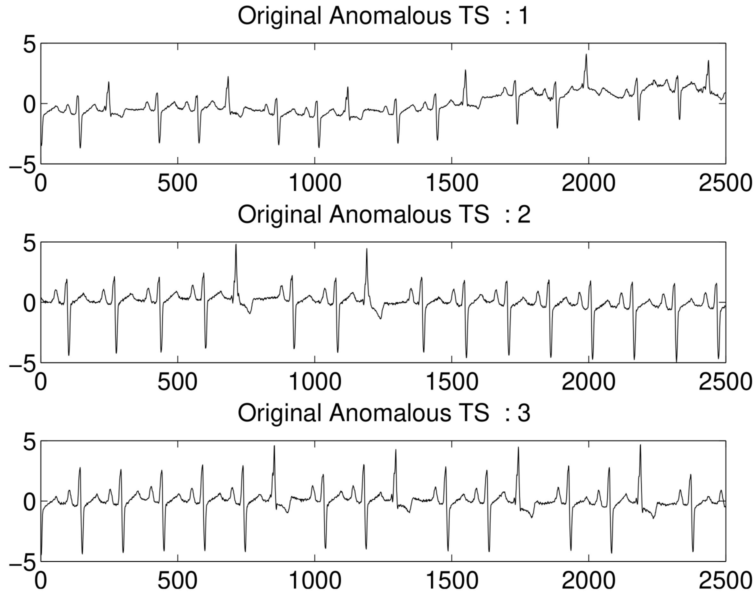


(a) Univariate Normal Time Series

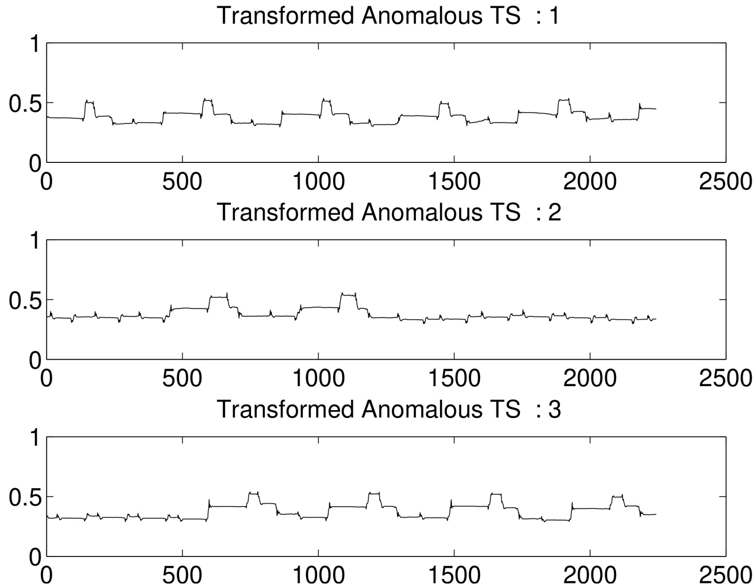


(b) Transformed Univariate Normal Time Series

Figure 3.3: Original and Transformed normal time series



(a) Univariate Anomalous Time Series



(b) Transformed Univariate Anomalous Time Series

Figure 3.4: Original and Transformed Anomalous time series