MODIFIED EMBEDDING FOR MULTI-REGIME DETECTION IN NONSTATIONARY STREAMING DATA

Evan Kriminger[†], José C. Príncipe[†], and Choudur Lakshminarayan[*]

[†]Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA
[*]Hewlett-Packard Laboratories, Palo Alto, CA, USA
{ekriminger, principe}@ufl.edu, choudur.lakshminarayan@hp.com

ABSTRACT

Many practical data streams are typically composed of several states known as regimes. In this paper, we invoke phase space reconstruction methods from non-linear time series and dynamical systems for regime detection. But the data collected from sensors is normally noisy, does not have constant amplitude and is sometimes plagued by shifts in the mean. All these aspects make modeling even more difficult. We propose a representation of the time series in the phase space with a *modified embedding*, which is invariant to translation and scale. The features we use for regime detection are based on comparing trajectory segments in the modified embedding space with *cross-correntropy*, which is a generalized correlation function. We apply our algorithm to non-linear oscillations, and compare its performance with the standard time delay embedding.

Index Terms— Real time detection, time series embedding, symbolic dynamics, correntropy

1. INTRODUCTION

In modern industrial operations, sensors monitor the state of the system and analytical methods are applied to the streaming data to detect a multitude of anomalous events and warn about impending failures. Detecting changes in the streaming data in real time is one of the fundamental challenges in data stream processing [1]. For example, in oil and gas applications, oil wells are equipped with thousands of sensors and gauges to measure oil flow rates, pressure, and temperature. Factors such as fluid composition, oil viscosity, compressibility, and specific gravity of water induce tremendous variability and produce varying flow regimes.

Fast detection of regime change of non-linear time series by model based methods is difficult in an on-line setting due to their reliance on computationally intensive mathematics [2]. Frequency domain approaches fail as well because they require a window of data to estimate spectral features that cause detection delays [3]. It is also possible to extract time domain features from the time series and make decisions based on their statistics. However, this methodology requires hand crafting by the user and fails when the time series statistics change [4]. On the other hand, *attractor* reconstruction provides a model independent representation of the dynamics that generate the time series [5]. An attractor is a set towards which a system converges over time. In the reconstructed phase space, online regime detection is achieved by comparing incoming trajectory segments to the embedded training set. The comparison is made in

the sense of a similarity measure that compares points in the trajectory at corresponding times. In this paper we use correntropy, which is a generalized correlation function [6].

In the following sections, we present the necessary background on time series embeddings and describe the drawbacks in nonstationary environments. We propose a modification of the traditional time delay embedding to provide translation invariance. We further modify the embedding to provide scale invariance. We call this a "modified embedding" because the representation is no longer an embedding in the strict mathematical sense. Cross-correntropy is introduced, and we present an algorithm for regime detection in the modified embedding space. Finally, we demonstrate the performance of the algorithm on non-linear oscillations.

2. BACKGROUND

An embedding is a map from an m-dimensional manifold to a $(2m+1)$-dimensional Euclidean space, where every point on the original manifold has a unique image in the higher dimensional space. Takens Embedding Theorem [7] provides a means of reconstructing the phase space of a multi-dimensional dynamical system from the time delays of a single series of measurements. Consider a discrete time series with x_n being the value at time n. Then, at each time n, we can build a vector

$$\mathbf{x}^{(n)} = [x_n, x_{n-\tau}, ..., x_{n-2m\tau}],$$

where m is the embedding dimension and τ is the time delay. The limit set of the trajectories (the attractor) is embedded in the manifold on created by the $\mathbf{x}^{(n)}$ values.

One typically determines the time delay by finding the first time lag that produces a local minima in a dependence measure between x_n and $x_{n-\tau}$, such as autocorrelation or mutual information [8]. The selection of τ is flexible and chosen such that the components of $\mathbf{x}^{(n)}$ are not correlated. After the time-delay τ is fixed, the embedding dimension m is estimated by algorithms such as Grassberger-Procaccia, which approximates the *correlation dimension* [9]. The time delay embedding preserves dynamical invariants such as entropy, dimensional, and Lyapunov exponents [8], which are used to analyze the underlying physical system.

In the case of regime detection, we are less concerned with the properties of the system than with changes in the reconstructed trajectories produced by noise, amplitude scaling and shifting means in the input streaming data. One can normalize the data, but if the data is time varying, these normalizations must be implemented online in the test set. Instead of these preprocessing approaches that are normally ad-hoc, this paper includes invariance to these aspects

directly in the embedding framework, with the advantage of fast on-line operation and also of a sound mathematical foundation. The goal of this paper is to provide a modified embedding space in which the points $\mathbf{x}$ and $a\mathbf{x} + b$ are indistinguishable, where a is a scaling factor, and b is the translation.

3. TRANSLATION INVARIANCE

Symbolic dynamics are used to provide accurate representations of reconstructed attractors. In particular, encoding time delay embedding vectors into symbols based on order patterns provides translation invariance [10]. In an m-dimensional space, each point maps to one of $m!$ order patterns. In applications, information loss is substantial due to this encoding. For example, the Lorenz attractor which unfolds in 3 dimensions, has only 6 order patterns. We propose a variant of the order patterns, called *difference patterns*. A difference pattern, $\Delta\mathbf{x}^{(n)}$, at time n, is given by

$$\Delta\mathbf{x}^{(n)} = [x_n - x_{n-2m\tau}, ..., x_{n-(2m-1)\tau} - x_{n-2m\tau}].$$

where every component of $\mathbf{x}^{(n)}$ is subtracted by the last component. The last component of $\Delta\mathbf{x}^{(n)}$ will always be zero, so is removed. Therefore, the original $(2m + 1)$-dimensional point in the reconstructed phase space is mapped to a $2m$-dimensional space. This is analogous to establishing a quantitative order pattern with the $x_{n-2m\tau}$ component as a zero reference point. The translation invariance in the space of order patterns is obvious. Let $y_n = x_n - b$, be a translated time series.

$$\begin{aligned}
\Delta\mathbf{y}^{(n)} &= [y_n - y_{n-2m\tau}, ..., y_{n-(2m-1)\tau} - y_{n-2m\tau}] \\
&= [(x_n - b) - (x_{n-2m\tau} - b), ...] \\
&= [x_n - x_{n-2m\tau}, ...] \\
&= \Delta\mathbf{x}^{(n)}
\end{aligned}$$

Translating the time series does not alter the difference pattern. Similarly to the differencing operation applied to nonstationary time series. However, this representation is no longer an embedding because it is not invertible.

4. SCALE INVARIANCE

To achieve scale invariance, we simply normalize the $\Delta\mathbf{x}^{(n)}$ vectors by their Euclidean norms. The new vectors,

$$\Delta\hat{\mathbf{x}}^{(n)} = \frac{\Delta\mathbf{x}^{(n)}}{||\Delta\mathbf{x}^{(n)}||},$$

are the projections of the $\Delta\mathbf{x}^{(n)}$ onto the unit sphere in $2m$-dimensional space. Consider now a translated and scaled time series $y_n = ax_n - b$. First, we create the difference pattern.

$$\begin{aligned}
\Delta\mathbf{y}^{(n)} &= [ax_n - ax_{n-2m\tau}, ...] \\
&= a\Delta\mathbf{x}^{(n)}
\end{aligned}$$

Translation invariance still holds, but the scaling factors out of the difference pattern, so it disappears when we normalize.

$$\begin{aligned}
\Delta\hat{\mathbf{y}}^{(n)} &= \frac{\Delta\mathbf{y}^{(n)}}{||\Delta\mathbf{y}^{(n)}||} \\
&= \frac{a\Delta\mathbf{x}^{(n)}}{a||\Delta\mathbf{x}^{(n)}||} \\
&= \Delta\hat{\mathbf{x}}^{(n)}
\end{aligned}$$

Effectively, this operation projects the trajectories onto the unit sphere and destroys some of the distance information that was present in the difference pattern attractor. In particular, all points on a line extending from the origin will map to the same point on the sphere. The modified embedding is, however, a useful depiction of the time evolution of the system. The regime detection proposed here will take advantage of the time structure of the trajectories rather than static distance information.

5. DETECTION WITH CROSS-CORRENTROPY

Consider the discrete random processes $\{\mathbf{X}_{n_1} : n_1 \in N_1\}$ and $\{\mathbf{Y}_{n_2} : n_2 \in N_2\}$, where N_1 and N_2 are time index sets. Then the cross-correntropy function is

$$\nu_{x,y}(n_1, n_2) = E[\kappa(x_{n_1}, y_{n_2})], \tag{1}$$

where $E[\cdot]$ is the expectation operator over the random processes and κ is a continuous positive definite kernel function. The correntropy function is a similarity measure between time series, that induces a metric, the correntropy-induced metric (CIM) [2]. It is common to use the Gaussian kernel, in which case (1) takes the form

$$\nu_{x,y;\sigma}(n_1, n_2) = E[G_\sigma(||x_{n_1} - y_{n_2}||)], \tag{2}$$

where σ is the Gaussian bandwidth. Selection of the the kernel bandwidth is an active area of research, but in our application, $0.5 \leq \sigma \leq 1.5$ worked well, because we are on a unit sphere with maximum geodesic distance of π.

For the problem of regime detection, we consider trajectory segments of length N, which are finite realizations of the random process. The expected value in (2) is replaced by the sample mean. The cross-correntropy between trajectory segments has the range

$$0 < \nu_{x,y;\sigma}(n_1, n_2) \leq 1.$$

Consider the streaming time series at time n_1. In the modified embedding space, the trajectory segment formed from the previous N points is

$$\Delta\hat{\mathbf{x}} = [\Delta\hat{\mathbf{x}}^{(n_1)}, \Delta\hat{\mathbf{x}}^{(n_1-1)}, ..., \Delta\hat{\mathbf{x}}^{(n_1-N+1)}].$$

Similarly, the length N trajectory segment ending at time n_2 in the training set is

$$\Delta\hat{\mathbf{y}} = [\Delta\hat{\mathbf{y}}^{(n_2)}, \Delta\hat{\mathbf{y}}^{(n_2-1)}, ..., \Delta\hat{\mathbf{y}}^{(n_2-N+1)}].$$

The sample correntropy between these two trajectories is

$$\hat{\nu}(\Delta\hat{\mathbf{x}}, \Delta\hat{\mathbf{y}}) = \frac{1}{N} \sum_{i=0}^{N-1} G_\sigma(||\Delta\hat{\mathbf{x}}^{(n_1-i)} - \Delta\hat{\mathbf{y}}^{(n_2-i)}||). \tag{3}$$

In on-line streaming, we will calculate (3) for all length N trajectory segments in the training set. If there exists a training segment $\Delta\hat{\mathbf{y}}$, such that $\hat{\nu}(\Delta\hat{\mathbf{x}}, \Delta\hat{\mathbf{y}}) \approx 1$, then the current sample x_{n_1} is assigned to the desired regime.

The trajectories in the modified embedding space are on the surface of a sphere. We replace the Euclidean distance in (3) with the geodesic length on the sphere, to obtain a better depiction of the distance between points. All modified embedding space points are unit vectors, so the dot product between two points is the cosine of the angle between them. The geodesic length is therefore the arc cosine of the dot product.

$$\hat{\nu}(\Delta\hat{\mathbf{x}}, \Delta\hat{\mathbf{y}}) = \frac{1}{N} \sum_{i=0}^{N-1} G_\sigma(\text{acos}(< \Delta\hat{\mathbf{x}}^{(n_1-i)}, \Delta\hat{\mathbf{y}}^{(n_2-i)} >))$$

The method of regime detection based on cross-correntropy in the proposed scale and translation invariant modified embedding space is described in Algorithm 1.

Algorithm 1 On-line regime detection in the modified embedding space with cross-correntropy

---

I. *Embed training set y*
Given a discrete time series $\{y_{n_2} : 1 \leq n_2 \leq L\}$
Select a time delay τ, and embedding dimension m
for $i = (m-1)\tau + 1$ **to** L **do**
   $\mathbf{y}^{(i)} = [y_i, y_{i-\tau}, ..., y_{i-(m-1)\tau}]$
   $\Delta\mathbf{y}^{(i)} = [y_i - y_{i-(m-1)\tau}, ..., y_{i-(m-2)\tau} - y_{i-(m-1)\tau}]$
   $\Delta\hat{\mathbf{y}}^{(i)} = \frac{\Delta\mathbf{y}^{(i)}}{||\Delta\mathbf{y}^{(i)}||}$
end for

II. *Regime detection of streaming data x at time n_1*
Set desired trajectory length to N
Set Gaussian kernel bandwidth σ
Set threshold ϵ on cross-correntropy
$\mathbf{x}^{(n_1)} = [x_{n_1}, x_{n_1-\tau}, ..., x_{n_1-(m-1)\tau}]$
$\Delta\mathbf{x}^{(n_1)} = [x_{n_1} - x_{n_1-(m-1)\tau}, ..., x_{n_1-(m-2)\tau} - x_{n_1-(m-1)\tau}]$
$\Delta\hat{\mathbf{x}}^{(n_1)} = \frac{\Delta\mathbf{x}^{(n_1)}}{||\Delta\mathbf{x}^{(n_1)}||}$
for $j = (m-1)\tau + N$ **to** L **do**
   $\nu(j) = \frac{1}{N} \sum_{k=0}^{N-1} G_\sigma(\text{acos}(< \Delta\hat{\mathbf{x}}^{(n_1-k)}, \Delta\hat{\mathbf{y}}^{(j-k)} >))$
end for
if $\max \nu > \epsilon$ **then**
   Sample x_{n_1} is in the desired regime
end if

For simplicity, only the case of single regime detection is shown. In the multi-regime case, one needs only to add further training sets to the embedding space. Then incoming trajectory segments are compared with the training sets of all regimes. The regime that produces the highest correntropy is selected.

6. RESULTS

We test the algorithm with a single desired regime of quasi-periodic oscillations produced from a chaotic regime commonly observed in the gas and oil industry. From the training set data (Fig. 1), mutual information was used to select a time delay of 3 and the Grassberger-Procaccia algorithm [9] revealed that the correlation dimension is 1.8. An embedding dimension of 4 sucessfully unfolded the attractor in this case. The training set is real data that has been mean-centered.

To demonstrate the effectiveness of our algorithm, we compare its performance with the standard time delay embedding. The embedding parameters and detection procedure remain the same. We
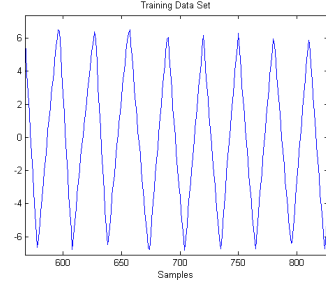

Fig. 1. Example of the training set oscillation data. The total training set length is 2500 samples.

set the segment length $N = 10$, and the kernel bandwidth $\sigma = 0.9$. The first test case is seen in Fig. 2. This is also real data that has been mean-centered. There are two oscillatory segments and a non-oscillatory middle segment. As seen in Table 1, the standard embedding slightly outperforms the modified embedding, which is expected because the use of distance as a discriminating characteristic is impaired in the modified space.
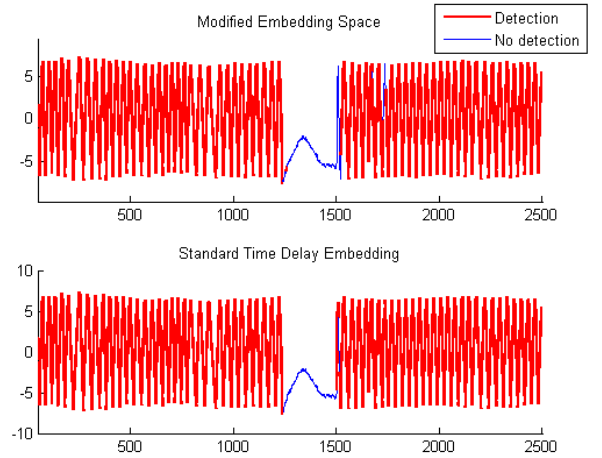

Fig. 2. First data set. The test data is similar in mean and amplitude to the training set. Algorithm 1 (top) performs slightly worse than the standard embedding (bottom).

In the previous test, the standard embedding worked because the training and testing data were similar in mean and scale. We artificially scale and translate sections of the testing set, as seen in Fig. 3, and perform a second test. We manually skew the test set so that regime boundaries are precisely known.

In the second test, our algorithm greatly outperforms the standard embedding. In the modified space, there is no difference between the testing and training data aside from the sudden jumps between regimes.

Our method can detect dynamics before periodicity is evident. We define the *reaction time* for a detector as the number of missed detections following the onset of a regime (Fig. 4). Time-frequency methods can also detect oscillations by identifying a strong spectral component at the fundamental frequency. In an on-line setting, at least one period of the oscillations must elapse to have meaning in the frequency domain, which is too much delay in some applications.
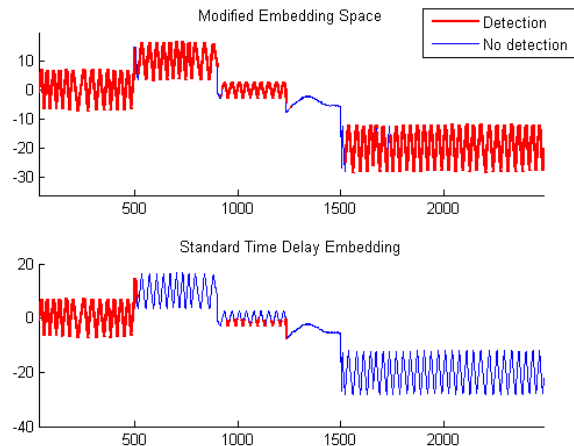
Fig. 3. Second data set. The testing data contains scaled and translated versions of the oscillations. In this case, Algorithm 1 (top) greatly outperforms the standard embedding (bottom).

Table 1. Error rate, true positive rate (TPR), and false positive rate (FPR) for the two tests.

Embedding	Error Rate	TPR	FPR
Test 1 (Fig. 2)			
Modified	1.9%	98.4%	4.1%
Standard	0.8%	99.6%	3.7%
Test 2 (Fig. 3)			
Modified	3.4%	96.5%	2.2%
Standard	65.7%	26.7%	3.0%

The segment length, N, offers the user control over the reaction time of the detector. In Table 2, we represent the tradeoff between reaction time and error rate. The test set is the same as in Fig. 2 and Algorithm 1 is again employed. The reaction time is determined from the transition to the second oscillatory regime. For the N values tested, all reaction times were well less than the average oscillation period of about 30 samples.

Table 2. Tradeoff between reaction time and error rate as a function of trajectory length.

Length N	Reaction Time	Error Rate
2	5	3.7%
4	8	2.5%
6	17	2.1%
8	17	1.7%
10	21	1.9%

7. CONCLUSIONS

The method presented in this paper is a simple way to achieve on-line, multi-regime detection in a nonstationary environment. Through two sequential modifications on the standard time delay embedding, we achieved a representation of dynamics that is invariant to constant scaling and translation. Cross-correntropy allows us
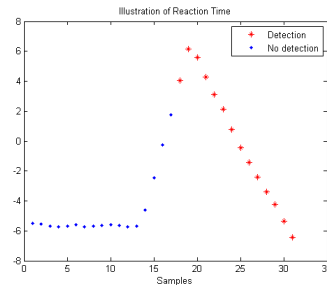

Fig. 4. Following the onset of oscillation, there are 4 missed detections, and therefore the detection delay is 4 samples.

to fully exploit the time structure of the trajectories, which sit on a sphere in the modifed embedding space.

In the case of quasi-periodic oscillations, when there is no scaling and translation between the training and testing sets, the modified embedding was shown to perform similarly to the standard embedding. When these two distortions are present, the modified embedding is unaffected, while the standard embedding fails. Our method has the benefit of short detector reaction time, relative to frequency domain methods, which in general will fail for the first period of oscillation.

8. REFERENCES

[1] Charu C. Aggarwal, *Data Streams: Models and Algorithms*, Springer, 2006.

[2] José C. Príncipe, *Information Theoretic Learning*, Springer, 2010.

[3] Donald B. Percival and Andrew T. Walden, *Spectral analysis for physical applications*, Cambridge University Press, 1993.

[4] Eamonn Keogh and Shruti Kasetty, "On the need for time series data mining benchmarks: A survey and empirical demonstration," in *SIGKDD*, 2002, pp. 102–111.

[5] Holger Kantz and Thomas Schreiber, *Nonlinear Time Series Analysis*, Cambridge University Press, New York, NY, USA, 2003.

[6] Ignacio Santamaría, Puskal P. Pokharel, and José C. Príncipe, "Generalized correlation function: definition, properties, and application to blind equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6-1, pp. 2187–2197, 2006.

[7] Floris Takens, "Detecting strange attractors in turbulence," *Lecture Notes in Mathematics*, vol. 898, no. 1, pp. 366–381, 1981.

[8] Henry D. I. Abarbanel, *Analysis of Observed Chaotic Data*, Springer, 1996.

[9] Peter Grassberger and Itamar Procaccia, "Characterization of strange attractors," *Phys. Rev. Lett.*, vol. 50, no. 5, pp. 346–349, Jan 1983.

[10] J.P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *EPL (Europhysics Letters)*, vol. 4, no. 9, pp. 973, 1987.