Time-series Novelty Detection Using One-class Support Vector Machines

Junshui Ma and Simon Perkins

{junshui,s.perkins}@lanl.gov
Space and Remote Sensing Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545

Abstract – **Time-series novelty detection, or anomaly detection, refers to the automatic identification of novel or abnormal events embedded in normal time-series points. Although it is a challenging topic in data mining, it has been acquiring increasing attention due to its huge potential for immediate applications. In this paper, a new algorithm for time-series novelty detection based on one-class support vector machines (SVMs) is proposed. The concepts of phase and projected phase spaces are first introduced, which allows us to convert a time-series into a set of vectors in the (projected) phase spaces. Then we interpret novel events in time-series as outliers of the "normal" distribution of the converted vectors in the (projected) phase spaces. One-class SVMs are employed as the outlier detectors. In order to obtain robust detection results, a technique to combine intermediate results at different phase spaces is also proposed. Experiments on both synthetic and measured data are presented to demonstrate the promising performance of the new algorithm.**

I. INTRODUCTION

Novelty detection, or anomaly detection, refers to automatic identification of unforeseen or abnormal phenomena embedded in a large amount of normal data [1, 10, 11]. One of its most attractive application scenarios is when time series are targeted [10, 11, 12, 13]. For example, in a safety-critical environment, it will be helpful to have an automatic supervising system to screen the time series generated by monitoring sensors, and to report abnormal observations. Meanwhile, novelty detection is a challenging topic, mainly because of the insufficient knowledge and inaccurate representative of the so-called "novelty" for a given system [1].

Despite its technical challenge, in the past over ten years novelty detection is a topic acquiring increasing attention, and a number of techniques have been proposed and investigated to address it. These techniques were experimentally proved to be effective in some cases, while they can fail in some other cases due to the assumptions and/or processes upon which they are based. For example, some were designed based on the assumption of possessing precise theoretical models of the underlying problem [2], or knowing the novelty conditions [3, 4, 5], which are unfortunately generally not true in real world. A wavelet-based signal trend shift detection method is proposed in [9]. Nevertheless, this method cannot detect short novel patterns embedded in normal signals. An interesting idea for novelty detection, inspired by the negative-selection mechanism of the immune system, was proposed in [10]. However, this method can potentially fail because the negative set will go to null with the increasing diversity of the normal set. The method, called TARZAN, proposed in [11] is based on converting the time series into a symbolic string. However, the procedure for discretizing and symbolizing real values in time series can potentially lose meaningful patterns in the original time series. The method presented in [15] is, strictly speaking, not novelty detection algorithm, because it requires knowing what kind of novelty is expecting.

In some other studies [6, 7, 8, 14], novel events were interpreted as outliers of the "normal" distribution function. The advantage of this direction is its theoretical tractability. Therefore, it leads to methods that can be well defined. This direction becomes especially appealing after an algorithm, called one-class SVMs, was proposed [6, 7], because the one-class SVMs can naturally detect outliers among a set of vectors. Following the direction of formulating novelty detection as outlier detection, this paper proposes a novelty detection algorithm for time series using one-class SVMs. As with other detection algorithms, it is impossible for this new algorithm to succeed in all scenarios. However, this algorithm can at least provide an alternative and complementary solution to some problems in which other available techniques may fail.

The main contributions of this paper are:

1) Introducing the (projected) phase space to allow one-class SVMs to be applied to time-series data;
2) Combining the one-class SVM outputs for different (projected) phase spaces to produce more robust novelty detection results.

The paper is organized as follows. Section II is devoted to a brief introduction to one-class SVMs. The conversion between a time series and a set of vectors in the (projected) phase space is proposed in Section III. The detection algorithm is presented in Section IV. Experiments on both synthetic and measured data are proposed in Section V to demonstrate the algorithmic performance.

0-7803-7898-9/03/$17.00 ©2003 IEEE

II. A BRIEF INTRODUCTION TO ONE-CLASS SUPPORT VECTOR MACHINES

We briefly introduce the basic concepts of one-class SVMs in this section. Its more detailed presentation can be found in [6] and [7].

Given a set of vectors $T = \{\mathbf{x}_i, i = 1 \dots l\}$, where $\mathbf{x}_i \in I \subseteq \mathbf{R}^E$, E is the dimension of I, and I is called the input space. A nonlinear function $\Phi(\mathbf{x})$ maps vector $\mathbf{x}$ from input space I into a huge, or even infinite, dimensional feature space F. We construct a hyper-plane in feature space F as

$$f(\mathbf{x}) = \mathbf{W}^T \Phi(\mathbf{x}) - \rho \tag{1}$$

to separate as many as possible of the mapped vectors $\{\Phi(\mathbf{x}_i), i = 1 \dots l\}$ from the origin in feature space F. The $\mathbf{W}$ and ρ in (1) are obtained by solving an optimization problem:

$$\begin{aligned} \min_{\mathbf{W},b} \quad & P = \frac{1}{2}\mathbf{W}^T\mathbf{W} + \frac{1}{\nu l}\sum_{i=1}^{l}\xi_i - \rho \\ s.t. \quad & (\mathbf{W}^T\Phi(\mathbf{x}) - \rho) \geq -\xi_i, \quad \xi_i \geq 0 \end{aligned} \tag{2}$$

where $\nu \in (0,1)$, and it is a parameter to trade-off the smoothness of $f(\mathbf{x})$ and fewer falling on the same side of the hyper-plane (1) as the origin in F.

After introducing Lagrange multipliers α_i for each vector $\mathbf{x}_i$, the dual problem of the optimization problem of (2) can be obtained. Solving the dual problem leads to

$$\mathbf{W} = \sum_{i=1}^{l} \alpha_i \Phi(\mathbf{x}_i),$$

where $0 \leq \alpha_i \leq \frac{1}{\nu l}$.

The famous *kernel trick* is the procedure of using a kernel function in input space I to replace the inner product of two vectors in feature space F. Accordingly, the hyper-plane (1) in feature space F becomes a nonlinear function in the input space I

$$f(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i K(\mathbf{x}_i, \mathbf{x}) - \rho, \tag{3}$$

where $K(\mathbf{x}_i, \mathbf{x}) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x})$, and it is a kernel function in the input space I. There are many admissible choices for kernel function $K(\mathbf{x}_i, \mathbf{x})$. The most widely used one in one-class SVMs is the RBF kernel function. That is,

$$K(\mathbf{x}_i, \mathbf{x}) = \exp\{-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2\}. \tag{4}$$

According to (2), any vector $\mathbf{x}$ with $f(\mathbf{x}) < 0$ is an outlier. Moreover, it can be proved that a vector $\mathbf{x}_i$ in the training set T is an outlier if and only if its α_i is $\frac{1}{\nu l}$.

It is proved that ν is the upper bound on the fraction of outliers over all training samples [6]. Therefore, the value of ν directly determines the sensitivity the outlier detection algorithm using one-class SVMs.

III. TIME SERIES VS. (PROJECTED) PHASE SPACE

According to Section II, one-class SVMs can only be applied to a set of vectors, and are not directly applicable to time-series type of data. Therefore, we have to figure out a method to convert the time series to a set of vectors. The most straightforward way is to unfold the time-series into a phase space using a time-delay embedding process [17].

More specifically, given a time series $x(t)$, $t = 1 \dots N$, it can be unfolded into its phase space Q, where $Q \subseteq \mathbf{R}^E$, and E is called the *embedding dimension* using a time-delay embedding process:

$$\mathbf{x}_E(t) = [x(t-E+1) \quad x(t-E+2) \quad \cdots \quad x(t)], \tag{5}$$

where $\mathbf{x}_t \in Q$. Thus, a time series $x(t)$ can be converted to a set of vectors $T_E(N)$,

$$T_E(N) = \{\mathbf{x}_E(t), t = E \cdots N\}. \tag{6}$$

Note that when we use the concept of phase space, we only use its mathematical form, and ignore its implied physical meaning. Also, it is obvious that vectors in $T_E(N)$ fail to meet the i.i.d. condition. This fact can cause the outlier detection results obtained using one-class SVMs to lose some nice properties, such as the PAC performance bound [6]. However, it does not damage the validity of using one-class SVMs for outlier detection, because the formulation of one-class SVMs were not derived based on the assumption that the data set follows the i.i.d. condition.

In some cases, when a time series is mostly composed of low frequency components, in phase space Q the set of vectors converted from this time series distribute along the diagonal vector **1**, where $\mathbf{1} = [1 \quad 1 \quad \cdots \quad 1]^T$. This point is demonstrated in Figure 1. If one-class SVMs are applied to a set of vectors like this, detection results will be heavily biased to the time-series points with either extremely large or extremely small values. Although novelty in many cases does appear in points with extreme values, this scenario potentially rules out the chance for detecting novel patterns formed by points with normal values. Therefore, the concept of *projected phase space* is introduced to cope with this bias.

The basic idea is to project all the vectors in a phase space to a subspace orthogonal to the diagonal vector **1**. That is, according to the projection theorem,

$$\mathbf{x}_E^{'}(t) = (\mathbf{I} - \frac{1}{E}\mathbf{1}\mathbf{1}^T)\mathbf{x}_E(t), \tag{7}$$

where **I** is the identity matrix, and $\mathbf{x}_E^{'}(t)$ is the projected vector in the projected phase space, denoted by $Q^{'}$.

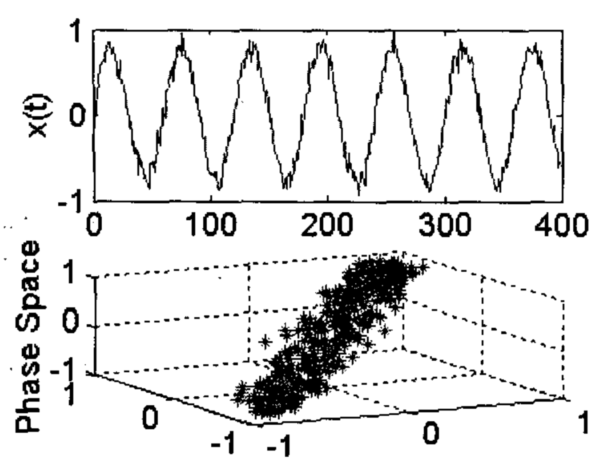

Fig. 1. A Time Series and Its Vector Set in a Phase Space (E=3)

Intuitively, the projection operation (7) is like applying a high-pass filter to the time series. The rationale of using this projection, instead of a high-pass filter, is that it can adaptively match the underlying data, and we need not specifically define any parameters, such as the cut-off frequency in high-pass filters.

Meanwhile, this projection operation is not guaranteed to be beneficial for every real world problem [19]. It is induced to enlarge our choices to handle situations where the one-class SVMs can be severely misled. In real world applications, whether the time series should be unfolded to phase spaces or to projected phase spaces is determined by both the property of the time series, and the kind of novelty we wish to detect. As suggested by other researchers, if we consider novelty detection as a problem of knowledge discovery and data mining, it should be an "iterative activity" [11], and "the discovery algorithm should be run several times with different parameter settings" [18].

IV. ONE-CLASS SVM-BASED NOVELTY DETECTION FOR TIME SERIES

After converting a time-series into a set of vectors in the (projected) phase space, a novelty detection algorithm for time series becomes readily available.

Given a time series $x(t)$, where $t = 1...N$, and its corresponding vector set $T_E(N) = \{\mathbf{x}_E^{(')}(t), t = E \cdots N\}$ in the (projected) phase space, we denote $i(E,t)$, where $t = 1...N$, and $i(E,t) \in \{0,1\}$, as the detection results when the embedding dimension is *E*. $i(E,t) = 1$ suggests that $x(t)$ is considered as a novel point in the (projected) phase space with an embedding dimension of *E*. The value of $i(E,t)$ is set to 1 if its corresponding time-series point $x(t)$ is the element of any outlier detected using one-class SVMs. For example, if $\mathbf{x}_E^{(')}(t)$ is detected as an outlier, all points from $i(E, t-E+1)$ to $i(E,t)$ are set to 1.

However, according to Section III, to unfold a time series into the (projected) phase space, one must first determine an embedding dimension *E*. Different embedding dimensions lead to different representations of the time series in phase spaces. Intuitively, if there is an intrinsic novel event happening, its novelty is supposed to manifest in its different representations.

In order to construct a robust novelty detection algorithm less dependent on a particular representation, we define $x(t)$ as a novel point only when the novelty indication function $I(t) = 1$, where $I(t) = \prod_{E \in S} i(E,t)$, and *S* is a set formed by a large range of embedding dimensions. The size of set *S*, along with the choice of its elements, plays an important role in trading-off between detection rate and false alarm.

V. EXPERIMENTS

Experiments based on both synthetic and measured data are presented to demonstrate the performance of our novelty detection algorithm.

A. Experiments Based on Synthetic Data

Two synthetic time series $x_1(t)$ and $x_2(t)$ are generated. $x_1(t)$ is a sinusoid signal with small additive noise, while $x_2(t)$ is the same as $x_1(t)$ except that it has a small segment of large additive noise. The algorithmic parameters are set as follows:

1) Embedding dimension range S, $S=[3 \quad 5 \quad 7 \quad \cdots \quad 19]$;
2) One-class SVM parameter $\nu = 0.02$;
3) Kernel function used by one-class SVM $K(\mathbf{x}_i,\mathbf{x}_j)=\exp\{-|\mathbf{x}_i-\mathbf{x}_j|^2/10\}$;

The original signals $x_1(t)$ and $x_2(t)$ are plots in Figure 2, along with the novelty indication function $I(t)$ obtained when both time series are unfolded into the projected phase spaces.

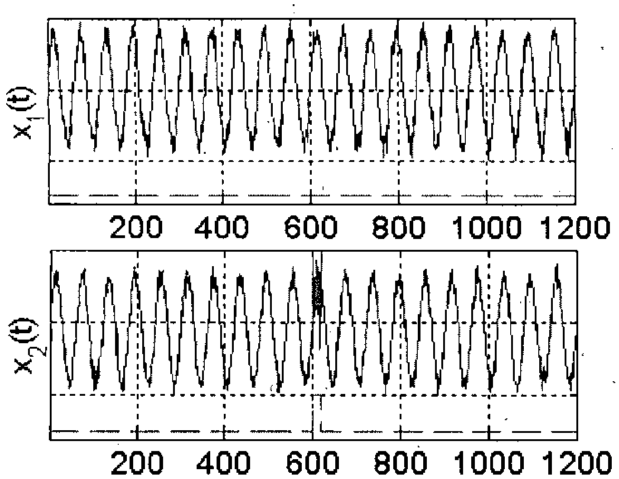

Fig. 2. Novelty Detection Results When Unfolded to Projected Phase Spaces

The solid curves in Figure 2 are the synthetic time series, and the dash-lines are the novelty indication function. Figure 2 shows that our detection algorithm successfully detects the novel points in $\mathbf{x}_2(t)$ without false alarms. Meanwhile, it also properly figures out that no part of $\mathbf{x}_1(t)$ can be considered as a novel point.

Because $x_1(t)$ and $x_2(t)$ are time series mainly composed of low frequency components, we predict that results obtained in projected phase spaces are more reliable. For comparison, the detection results obtained by unfolding both time series into the phase spaces are shown in Fig. 3. Compared with the result in Fig. 2, two false alarms are observed in this figure, which coincides with our prediction.

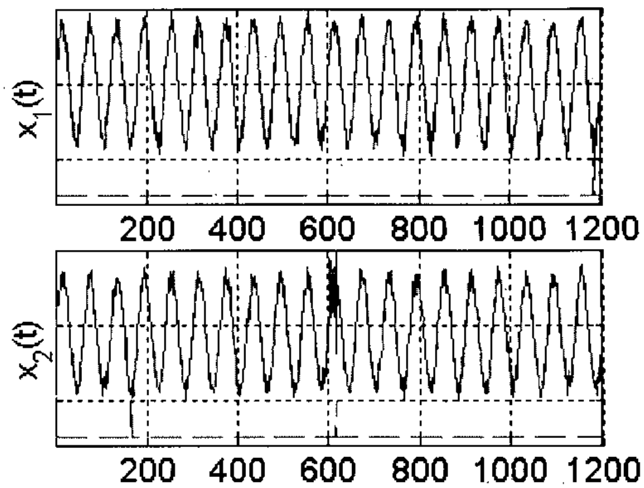

Fig. 3. Novelty Detection Results When Unfolded to Phase Spaces

B. An Experiment Based on Measured Data

The experiment is to apply the detection algorithm to the Santa Fe Institute Competition (SFIC) data [16], which is a 1000-point time series. The algorithmic parameters are set exactly the same as for the experiments in Subsection A, except that $\nu = 0.05$ is employed in this experiment. Because SFIC time series does not only have low frequent components, the novelty detection results obtained in phase spaces and project phased spaces are fairly similar. Therefore, we only plot out the result obtained in phase spaces in Figure 4. This result perfectly matches the human visual detection result.

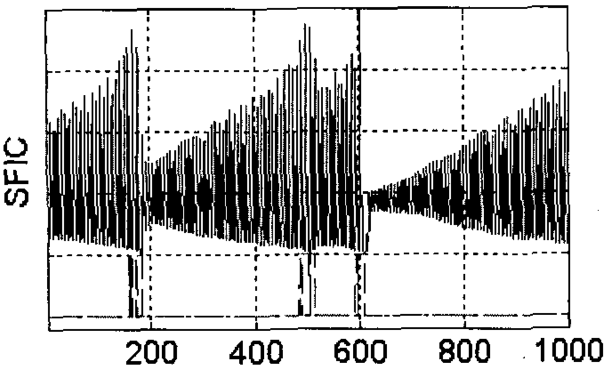

Fig. 4. Novelty Detection Results When Unfolded to Phase Spaces

VI. CONCLUSIONS

This paper proposes a new novelty detection algorithm for time series using one-class support vector machines. Experimental results demonstrate its promising performance. An interesting future direction related to this research is to find out the possible relationship between the value of ν and the confidence of the detected novelty.

ACKNOWLEDGEMENTS

This work is supported by the NASA project NRA-00-01-AISR-088 and by the Los Alamos Laboratory Directed Research and Development (LDRD) program.

REFERENCES

[1] Alexander Ypma and Rober P. Duin, Novelty Detection Using Self-Organizing Maps, in Progress in Connectionist-Based Information Systems, pp 1322-1325, London: Springer, 1997.

[2] R. Kozma, M. Kitamura, M. Sakuma, and Y. Yokoyama, Anomaly Detection by Neural Network Models and Statistical Time Series Analysis, in Proceedings of IEEE International Conference on Neural Networks, Orlando, Florida, June 27-29, 1994.

[3] Rolf Isermann, Process Fault Detection Based on Modeling and Estimation Method – A Survey, *Automatica*, vol. 20, pp.387-404, 1984.

[4] C. M. Bishop, Novelty Detection and Neural Network Validation, IEE Proceedings – Vision, Image and Signal Processing, vol. 141, no. 4, pp. 217-222, August, 1994.

[5] S.Roberts and L. Tarassenko. A Probabilistic Resource Allocating Network for Novelty Detection, Neural Computation, vol. 6, pp. 270-284, 1994.

[6] B. Schölkopf, R.C. Williamson, A.J. Smola, J. Shawe-Taylor, and J. Platt. Support vector method for novelty detection. In Neural Information Processing Systems, 2000.

[7] Colin Campbell, Kristin P. Bennett, A Linear Programming Approach to Novelty Detection, in Advances in Neural Information Processing Systems, vol 14, 2001.

[8] H. V. Jagadish, N. Kouda, and S. Muthukrishnan, Mining deviates in a time series database, in Proceedings of 25th International Conference on Very Large Data Bases, pp 102-113, 1999.

[9] C. Shahabi, X. Tian, and W. Zhao, Tsa-tree: A Wavelet-based Approach to Improve the Efficiency of Multi-level Surprise and Trend Queries. In Proceedings of 12th International Conference on Scientific and Statistical Database Management, 2000.

[10] Dipanker Dasgupta and Stephanie Forrest, Novelty Detection in Time Series Data Using Ideas from Immunology, In Proceedings of the 5th International Conference on Intelligent Systems, Reno, Nevada, June 19-21, 1996.

[11] E. Keogh, S Lonardi, and W Chiu, Finding Surprising Patterns in a Time Series Database In Linear Time and Space, In the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp 550-556, Edmonton, Alberta, Canada, July 23 - 26, 2002.

[12] Valery Guralnik, Jaideep Srivastava, Event Detection from Time Series Data. In Proceedings of the International Conference Knowledge Discovery and Data Mining, San Diego, California, 1999.

[13] Tom Brotherton, Tom Johnson, and George Chadderdon, Classification and Novelty Detection Using Linear Models and a Class Dependent-Elliptical Basis Function Neural Network, in Proceedings of the International Conference on Neural Networks, Anchorage, May 1998.

[14] Martin Lauer, A Mixture Approach to Novelty Detection Using Training Data With Outliers, Lecture Notes in Computer Science, vol. 2167, pp. 300-310, 2001.

[15] Richard J. Povinelli, Xin Feng. (in press) "A New Temporal Pattern Identification Method For Characterization And Prediction Of Complex Time Series Events," IEEE Transactions on Knowledge and Data Engineering.

[16] A. S. Weigend, N. A. Gershenfeld, Ed., Time-series Prediction: Forcasting the future and Understanding the Past, Addison-Wesley, 1994.

[17] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a Time Series", Physical Review Letters, vol. 45, pp. 712-716, 1980.

[18] G. Das, K.-I. Lin, H. Mannila, G. Renganathan, and P. Smyth. "Rule discovery from time series," in Proceedings of the 4th International Conference on Very Large Data Bases, pp. 606-617, 1998.

[19] J. Theiler, and S. Eubank, "Don't bleach chaotic data", Chaos, vol.3, pp. 771-782, 1993.