

A SVM-based behavior monitoring algorithm towards detection of un-desired events in critical infrastructures

Y. Jiang¹, J. Jiang¹, and P. Capodiecì²

¹Digital Media & Systems Research Institute, University of Bradford, UK
j.jiang1@bradford.ac.uk, paolo.capodiecì@Selex-Comms.com

²Selex Communications S.p.A, Italy

Abstract. In this paper, we report our recent research activities under MICIE, a European project funded under Framework-7 Programme, in which a SVM-based behaviour modeling and learning algorithm is described. The proposed algorithm further exploits the adapted learning capability in SVM by using statistics analysis and K-S test verification to introduce an automated parameter control mechanism, and hence the SVM learning and detection can be made adaptive to the statistics of the input data. Experiments on telecommunication network data sets support that the proposed algorithm is able to detect undesired events effectively, presenting a good potential for development of computer-aided monitoring software tools for protection of critical infrastructures.

1 Introduction

The term “Critical Infrastructures (CIs)” relates to “those assets or parts thereof which are essential for the maintenance of critical societal functions, including the supply chain, health, safety, security, economic or social well-being of people”. According to COM(2006)787 Critical Infrastructure encompass the following sectors and related sub sectors: (i) Energy to include Oil and gas production, refining, treatment, storage and distribution by pipelines, as well as electricity generation and transmission; (ii) Information communication technologies (ICT) to include Internet, satellite, broadcasting, and instrumentation automation and control systems etc. (iii) water, food and health facilities and supply chains; (iv) transport and finance systems. As CIs can be damaged, destroyed or disrupted by deliberate acts of terrorism, natural disasters, negligence, accidents or computer hacking, criminal activities and malicious behaviours, it becomes extremely important to save the lives and property of people at risk in the EU from terrorism, natural disasters and accidents, and any disruptions or manipulations of CIs should be brief, infrequent, manageable, geographically isolated and minimally detrimental to the welfare of the Member States, their citizens and the European Union. To this end, the MICIE project (funded under FP7) is to support the improvement of Critical Infrastructure Protection capability in Europe through the design and implementation of an on-line “MICIE alerting system”, which is able to predict, in real time, the cascade effects (expressed in risk levels of being no more able to provide the services with the targeted QoS) on a given CI

of some undesired events happened in the reference CI and/or in other independent CIs. It is expected that a range of software tools will be developed under MICIE via intelligent computing approaches to support the activities and decisions to be taken by the CI operators.

The challenging issue here is that critical infrastructures are dependent with each other in terms of their securities, and thus, to develop a useful MICIE alerting system, two categories of software tools need to be developed, which means that not only the behavior of each individual CI needs to be modelled, monitored and controlled, but also the inter-dependency among different CIs.

In this paper, we focus on the issue of behaviour modelling of individual CIs, and hence a monitoring software tool can be developed first to detect unusual patterns and undesired events towards robust risk analysis, prediction and hence a range of meta-data can be produced to communicate with other CIs and pave the way for inter-dependency modelling as well as integrated approaches for the final MICIE alerting system.

Essentially, behavior modelling of individual CI is about analyzing the raw information generated to indicate the operational status of the CI. Examples of such raw information include the network traffic data for telecommunication CIs or sets of control parameters often referred to as KPIs (key performance indicators). Existing approach adopted by most of the CI industry is rule-based, where a set of thresholds is set up according to operational experiences, and anything beyond one or more thresholds could prompt investigation or other actions by engineers. While such approaches are effective to maintain a reasonable level of security, it is essentially labour intensive and thus the running cost is very high. Recently, artificial intelligence approaches are introduced to examine the possibility of computer-aided security of facilities to protect the CIs, especially the anomaly detection around the telecommunication networks [1-5]. These approaches are represented by statistics-based and neural network based. While statistics analysis [1,2], such as Bayesian and hidden Markov etc. is used to determine the most appropriate threshold values to complete the control and detection of those abnormal behaviours, the neural network based approaches [4,5] represents machine learning of training data set supported by ground truth to analyze the variation of input data and hence capture the behaviour of the CI. In this paper, we combine these two approaches together to exploit the statistics analysis for adaptive estimation of controlling parameters to drive the machine learning approach, and hence an adaptive to input statistics machine learning approach can be designed for the purpose of modelling the behavior of CIs. In comparison with existing techniques, our proposed algorithm achieves a range of advantages, including: (i) the behavior modelling and unusual pattern detection are made adaptive to the variation of input data via its statistics analysis; (ii) automated determination of controlling parameters for SVM to achieve the best possible performances for pattern detection and modelling.

The rest of the paper is organized in two sections, where section 2 describes the proposed algorithm, section 3 reports experimental results and concluding remarks.

2. The Proposed Algorithm Design

As one of the most popular machine learning approaches, SVM has received tremendous attention in a number of areas to deal with learning, training and optimising problems. General SVM uses a kernel-mapping technique to separate linearly non-separable cases in high dimensional space [7]. SVM not only separates data from different classes, but also separates data to its maximum margin. In other words, SVM not only divide mixed data sets into classes, but also optimize such a classification. However, the weakness of general SVM lies in the fact that it requires labeled data sets to get it trained, yet many practical applications, especially the information associated with many CIs, do not have such labelled data to provide a so-called ground truth. This is because the operators do not have clear ideas about which patterns are regarded as abnormal and which are regarded as normal until investigation of individual cases is completed, and the outcome of such investigations is often case sensitive. To this end, we propose a statistics-based one-class SVM algorithm to complete the automated machine learning of CI's input information and hence leading to successful behavior modelling of CIs.

One-class-SVM is essentially an extension of support vector machines [6,8] used for detecting the outliers [7, 9]. The idea is that it first maps the data into high dimensional space, and then maximizes the margin between the mapped data and the origin. The main difference from general SVMs is the fact that a constraint parameter ν is introduced to control the maximum percentage of outliers in the dataset, which can also be used to indicate the priori. As a result, the one-class-SVM is capable of being adaptive to input changes or different players when used to model the behavior of CIs. This is because that the priori specifies the maximum likelihood that outlier detections can be made, and hence such detection is less sensitive to changes of inputs generated by different CIs or different elements within a CI. In comparison with neural network based approaches [4], SVM presents a range of advantages, which can be summarized as: (i) while artificial neural networks (ANNs) can suffer from multiple local minima, the solution to an SVM is often global and unique; (ii) while ANNs use empirical risk minimization, SVMs use structural risk minimization. As a result, the computational complexity of SVM is not dependent on the dimensionality of the input space, and SVMs can also have a simple geometric interpretation and generate a sparse solution; (iii) from wide range of reports on evaluation of both SVMs and ANNs [9,10], it is generally concluded that SVMs often outperform ANNs in many classification-relevant applications due to the fact that they are less prone to over-fittings.

Given a data set describing the operational status of the target CI: $T = \{x_1, x_2, \dots, x_l\}$, where $x \in R^N$ is an input vector with N elements, i.e. each item inside the status data set is regarded as an N-dimensional vector, a learning system such as SVM-based can be established to process the N-dimensional vectors to model, analyze, and monitor the CI's operation. The essential task is to find a function f that generates the value "+1" for most of the vectors in the data set, and "-1" for the other very small part. The Strategy for such a classification and detection is to

use a one-class-SVM [9] and map the input data into a Hilbert space H according to a mapping function $X = \phi(x)$, and separate the data from the origin to its maximum margin.

As a result, to separate the mapped data from the origin to its maximum margin is equivalent to solving the following quadratic optimisation problem:

$$\min_{w \in F} : \frac{1}{2} \|w\|^2 + \frac{1}{\nu l} \sum_i \xi_i - \rho \quad (1)$$

Subject to:

$$f(x) = w\phi(x) - \rho \geq -\xi_i, \xi_i > 0, i = 1, \dots, l \quad (2)$$

Where $\nu \in (0,1)$ is a constraint parameter to limit the maximum proportion of the high performance responses among all the ordinary responses as such that a maximum of $\nu \times 100\%$ are expected to return negative values according to $f(x) = w\phi(x) - \rho$. ξ_i are slack variables acting as penalties in the objective function.

It is proved [9] that $\nu \times 100$ is the upper bound percentage of the data that are expected to be outliers in the training data, and a vector x_i is detected to be outlier in the training set, if and only if $\alpha_i = 1/(\nu l)$. α_i is the parameter directly determines the sensitivity of outlier detection using one-class-SVM. Its selection is dependent on the specifications and requirements for protection of individual CIs and specific expectations by the engineers who operate the CI protection. Inevitably, such parameter setting would be extremely complicated as it is connected to many other factors, such as investigation of suspicious events or patterns, their related human labour costs, and understanding of all the operational status information sets etc. As a matter of fact, such information data sets are normally provided by user partners within the MICIE consortium and hence making it difficult and time consuming for technology or research partners to understand such operational data sets before any artificial intelligence and machine learning algorithms could be developed. Further, CI operators are often sensitive in handing out critical information for confidential purposes, which make it additionally difficult to get the collaboration going smoothly.

Under this circumstance, we propose to adopt a light-touch approach, where focus of research is to analyse the operational data sets by estimating their statistics to provide a context simulation for machine learning without too much regards to their specific meaning to those CI operators. Otherwise, we could be trapped into the situation that research partners need to learn the whole CI operation process before any software tools could be developed, yet such learning process is often made extremely difficult and almost impossible for security reasons. In this way, research under MICIE becomes two important steps: (i) pre-processing the operational data sets and convert them into input vectors; (ii) estimate and analyze their statistics to activate the

machine learning such as the one-class SVM as described above. In this paper, we report our initial results on the second step to solve the essential problem that how these parameters could be adaptively determined to drive the SVM.

Given the input operational data sets, $\psi = \{y_1, y_2, \dots, y_M\}$, we estimate their statistics features, such as the mean and variance, as follows:

$$\mu = \frac{1}{M} \sum_{i=1}^M y_i \quad (3)$$

$$\sigma^2 = \frac{1}{M} \sum_i (y_i - \mu)^2 \quad (4)$$

We then apply K-S test to all the samples to verify a probability function distribution (PDF) to characterize the input information source, for which it is most likely expected to use Gaussian distribution. Consequently, we propose the following technique to estimate the SVM parameter γ via exploiting the above statistics analysis of the input operational data sources:

$$\gamma = P(y_i > \mu + \sigma) \quad (5)$$

In other words, the constraint parameter is estimated to be the probability of those data samples that are larger than $\mu + \sigma$. As μ indicates the mean value of all samples and σ the standard deviation, their combination would provide an adaptive constraint to characterize the input data source, which is equivalent to a threshold that is adaptive to the variation of input statistics and such adaptability is fully automatic. Figure-1 illustrates such an adaptability, where the shaded area indicates $P(y_i > \mu + \sigma)$. As seen, when the statistics of input changes, indicated by the change of μ or σ , the shaded area will also change correspondingly.

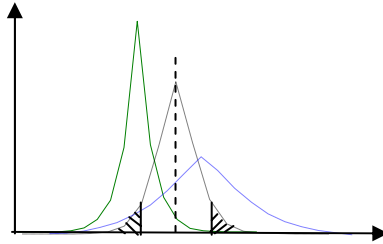


Figure-1: Adaptability illustration to the variation of input statistics

3. Experimental Results and Concluding Remarks

To evaluate the proposed algorithm, we carried out two phases of experiments, where the first phase is to run the K-S test to verify the Gaussian distribution and the second

phase is to detect unusual patterns out of telecommunication network traffic data sets, which is regarded as one of the most important CIs under the section of ICT.

Given the input samples, the K-S test is characterized by the following operation:

$$H = \begin{cases} 0 & \text{if } \gamma_{ks} = \max_{y \in \zeta_j, j \in [1, N]} \left| \sum_{i \leq y} C_i - \int f(y) dy \right| < D_\alpha \\ 1 & \text{else} \end{cases} \quad (6)$$

The corresponding hypothesis is given as:

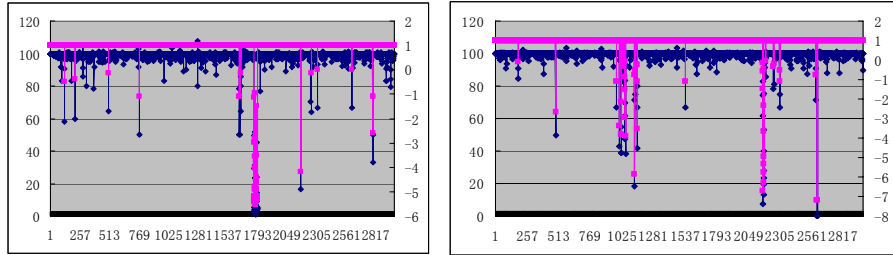
- H=0: It can be established that the tested samples come from a population with probability distribution $f(y)$ at the given significance level D_α ;
- H=1: It can be established that the tested samples do not come from a population with probability distribution $f(y)$ at the given significance level.

Table-I: K-S test results

Sample sets	$f(y)$	γ_{ks}	$D_{0.05}$	H
Set-1	Gaussian	0.0026	0.0059	1
	Laplace	0.0265	0.0059	0
Set-2	Gaussian	0.0023	0.0057	1
	Laplace	0.0254	0.0057	0
Set-3	Gaussian	0.0047	0.0060	1
	Laplace	0.0234	0.0060	0

Table-I illustrates all the K-S test results, where we divided the input samples into three sub-sets for efficiency of analysis purposes, and tested both Gaussian distribution and Laplace distribution. As seen, all the test results indicate that the samples we tested conform to the Gaussian distribution for a number of significance levels. In case that the K-S test indicates a strong fit-in with Laplace distribution, the constraint parameter should be determined via Laplace PDF in equation (5). As seen, the advantage of the proposed algorithm lies in the fact that the SVM learning is not only made adaptive to the statistics features of the input data, but also to the varying nature of its probability distributions. For complicated behaviour of CIs, such approach will prove useful that a number of PDFs are required to present a piece-wise simulation and characterization of the input information source.

Figure-2 illustrates the experimental results of the outlier detection via the one-class SVM learning mechanism for four different sets of traffic data captured from four network nodes.



(a) Outlier detection for data set-1 (b) Outlier detection for data set-2

Figure 2: Illustration of experimental results for outlier (undesired events) detection

While the blue points represent the normal behavior of the telecommunication network nodes, the pink highlighted points indicate suspicious patterns or events, which may need attention or even investigations by the CI operators. Via our industrial partner's investigations, it is verified that the proposed algorithm achieved around 80% accuracy in detecting those true undesired events in comparison with the ground truth. The associated false positive rate is around 17%, which is a reasonable price to be paid for the outlier detection. There exist certain relationship between the false positive rate and true positive rate, and the general trend is that the closer the true positive rate is to 100%, the larger the false positive rate, which could reach 100% in certain cases.

In this paper, we have described our latest research activities under the MICIE project, where a SVM-based machine learning algorithm is proposed to automatically monitor the behaviour of CIs and detect unusual patterns or events out of the operational data sets. The essential idea is to carry out statistics analysis to pre-process the input information and hence provide a set of context information to drive the machine learning module and make it adaptive to the statistics of input. In this way, the machine learning algorithm can be made sort of universal, in which the meaning of input data could be less regarded via concentrating on capturing their statistics rather than their functionalities. Experimental results support that such an approach is effective and also efficient in terms of running costs, due to the fact that the entire process is computer-aided, self-adapted, and the exposed parameters are tractable by human users. In practical cases, the statistics estimation and analysis need to be regularly run to ensure that the hypothesis of a certain probability distribution is correct, and the specific values of mean and standard deviation are updated.

Finally, the authors wish to acknowledge the financial support for the research work supported by the MICIE project under the European Framework-7 Programme (Contract No: 225353).

References:

1. A. Patcha, J-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends", *Computer Networks*, Vol. 51, pp.3448-3470, 2007.
2. S. Rajasegarar, C. Leckie, M. Palaniswami, "Anomaly detection in wireless sensor networks", *IEEE Wireless Communications*, August, 2008.
3. S.-J. Han, S.-B. Cho, "Evolutionary Neural Networks for Anomaly Detection Based on the Behavior of a Program", *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Vol. 36, No. 3, 2006.
4. S. Muthuraman, J. Jiang, "Anomaly detection in telecommunication network performance data," *Proceedings of the 2007 International Conference on Artificial Intelligence*, Monte Carlo Resort, Las Vegas, Nevada, USA, June, 2007.
5. T. Shon, J. Moon, "A hybrid machine learning approach to network anomaly detection", *Information Sciences*, Vol. 177, pp. 3799-3821, 2007.
6. L.M. Manevitz, M. Yousef, "One-Class SVMs for document classification", *Journal of Machine Learning Research*, Vol. 2, pp. 139-154, 2001.
7. S.S. Keerthi and C.J. Lin, "Asymptotic behaviors of support vector machines with Gaussian Kernel", *Neural Computation*, vol. 15, no. 7, pp. 1667-1689, 2003.
8. B. Schölkopf, R. Williamson et al. "Support vector method for novelty detection", *Neural Information processing Systems*, MIT Press, pp 582-588, 2000;
9. Y. Li and J. Jiang "Combination of SVM knowledge for microcalcification detection in digital mammograms" *Lecture Notes in Computer Science*, Springer Verlag, Vol 317, pp359-365, 2004;
10. Kalatzis I, Piliouras N. et al 'Comparative evaluation of probabilistic neural network versus support vector machines classifiers in discriminating EPR signals of depressive patients from healthy control', *Image and Signal Processing and Analysis*, ISPA2003, Vol 2, 18-20 Sept. 2003, pp981-985.