2012 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 23–26, 2012, SANTANDER, SPAIN

KERNELS FOR TIME SERIES OF EXPONENTIAL DECAY/GROWTH PROCESSES

Zineb Noumir, Paul Honeine

Institut Charles Delaunay (CNRS)
Université de technologie de Troyes
Troyes, France

Cédric Richard

Laboratoire Lagrange (CNRS)
Observatoire de la Côte d'Azur
Université de Nice Sophia-Antipolis
Nice, France

ABSTRACT

Many processes exhibit exponential behavior. When kernel-based machines are applied on this type of data, conventional kernels such as the Gaussian kernel are not appropriate. In this paper, we derive kernels adapted to time series of exponential decay or growth processes. We provide a theoretical study of these kernels, including the issue of universality. Experimental results are given on a case study: chlorine decay in water distribution systems.

Index Terms— Kernel function, one-class classification, normalization, kernel methods, support vector machines

1. INTRODUCTION

In a wide variety of situations, many processes exhibit an exponential behavior. This behavior follows from natural processes, when a quantity decreases/increases at a rate proportional to its value. Many quantities depend exponentially on time, either by increasing such as in virus spread, economic growth, nuclear chain reaction, or by decreasing such as given in heat transfer, radioactivity, pharmacology, and chemical reactions, only to name a few. Exponential decay[1] also arises in many time series, such as the measure of degradation based on vibration (bearings, dampers, etc), as well as any other characteristic whose evolution represents an evolution to failure [1, 2].

While all these problems are of interest, and without loss of generality, this paper deals with chemical reactions, and more precisely the decay of Chlorine concentration in a water supply network. Chlorine is the most used disinfectant to treat water for microbiological protection in a water supply network [3]. It is injected at treatment stations, and chlorine concentrations decay over time essentially as water disinfection takes place, while it must be present at the points of water consumption. Chlorine exhibits an exponential decay behavior [4, 5] while, at any given node (sensor), it largely fluctuates due to non-constant consumer demand, temperature, pH, etc. These fluctuations inhibit the use of an adapted model to characterize the domain of operation. See Figure 1.

By learning the system from available data, machine learning provides an elegant approach to overcome model-based techniques where a perfect understanding of the system is necessary. This is the idea behind kernel-based machines, pioneered by Vapnik's support vector machines (SVM) for classification and regression [6]. Domain description using the so-called one-class SVM is proposed in [7], by solving a quadratic programming problem. These machines share an interesting property: sparsity. In fact, only a small number of the training samples contributes to the final model. These samples, the so-called Support Vectors, describe the diversity within the training data.

The performance of a kernel-based learning machine is highly related to the choice of the kernel function. The kernel determines an implicit (nonlinear) map that transforms the data, from the input space to a feature space, and therefore determines its distribution in the latter space. Conventional kernels include the polynomial, the exponential and the Gaussian kernels. Because there is "no free lunch" in kernel selection [8], the choice of the kernel should reflect prior knowledge about the problem at hand.

In this paper, we design appropriate kernels for time series of exponential decay (or growth) processes. Conventional kernels, such as the Gaussian kernels, are not appropriate for this class of data. This is illustrated in practice by the large dispersion of these data in the resulting feature space. In consequence, one gets "less regularity" using kernel-based machines. This results in many support vectors, even when a single steady (constant-parameter) process is under investigation. In this paper, we propose a class of kernel functions to overcome these drawbacks. Inspired by the normalized kernel [9], we derive several kernels, and study the generalization ability in the sense of universality, as defined in [10].

The rest of the paper is organized as follows. Section 2 presents the time series for exponential decay processes. We derive kernels adapted for these data in Section 3. Section 4

This work was supported by ANR-08-SECU-013-02 VigiRes'Eau.

[1] Exponential decay includes exponential growth, when one measures the negative quantity.

978-1-4673-1026-0/12/$31.00

outlines the problem of chlorine concentration in water distribution systems, and provide experimental results on real time series. Conclusion and further directions are given in Section 5. But before, we present the classical one-class SVM for domain description.

One-class SVM for domain description

Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ be a set of samples, and let $\Phi(\cdot)$ be a non-linear transformation defined by some positive semi-definite kernel $\kappa(\cdot,\cdot)$, i.e., $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$ corresponds to the inner product between $\Phi(\boldsymbol{x}_i)$ and $\Phi(\boldsymbol{x}_j)$.

The one-class SVM finds a sphere, of minimum volume, containing all (or most of) the samples. Its center $\boldsymbol{c}$ and radius r are obtained by solving the following optimization problem:

$$\min_{r,\boldsymbol{c},\boldsymbol{\zeta}} \; r^2 + \frac{1}{\nu n}\sum_{i=1}^{n}\zeta_i$$
$$\text{subject to } \|\Phi(\boldsymbol{x}_i) - \boldsymbol{c}\|^2 \leq r^2 + \zeta_i \quad \text{for all } i = 1, 2, \ldots, n$$

where, $\zeta_1, \zeta_2, \ldots, \zeta_n$ are non-negative slack variables and ν is a positive parameter that specifies the tradeoff between the sphere volume and the number of outliers, i.e., samples lying outside the sphere. By introducing the Karush-Kuhn-Tucker optimality conditions, we get the optimal model of the center

$$\boldsymbol{c} = \sum_{i=1}^{n} \alpha_i \, \Phi(\boldsymbol{x}_i), \tag{1}$$

where the α_i's are the solution of the so-called dual problem:

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \alpha_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}_i) - \sum_{i,j=1}^{n} \alpha_i \alpha_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)$$
$$\text{subject to } \sum_{i=1}^{n} \alpha_i = 1 \text{ and } 0 \leq \alpha_i \leq \frac{1}{\nu n} \text{ for all } i = 1, 2, \ldots, n.$$

The one-class SVM for domain description can also be considered for novelty detection tasks. To this end, any new sample $\boldsymbol{x}$ can be regarded as novel to the domain described by the sphere, namely to the distribution of the training samples, when its distance to the center is beyond some threshold, where the distance is given as

$$\|\Phi(\boldsymbol{x}) - \boldsymbol{c}\|^2 = \sum_{i,j} \alpha_i \alpha_j \kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) - 2\sum_{i} \alpha_i \kappa(\boldsymbol{x}_i, \boldsymbol{x}) + \kappa(\boldsymbol{x}, \boldsymbol{x}).$$

The one-class SVM (derived above), as well as SVM machines for regression and classification [6], enjoy an important property: sparsity. It is well known that only a small fraction of the training samples contributes to the model (1). These samples, called Support Vectors (SVs), provide a sufficient representation of the samples, thus describing the diversity of the data. Other algorithms based on different SVs selection criteria have been recently applied with success for the one-class problem in [11, 12]. All these algorithms perform essentially identically, while the performance depends crucially on the choice of the kernel function. There is "no free lunch" in kernel selection [8]. The choice of the kernel should reflect prior knowledge about the problem at hand. The main contribution of this paper is to design appropriate kernels for time series of exponential decay processes.

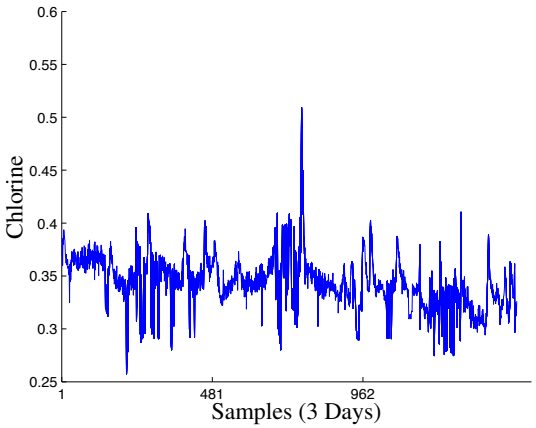

Fig. 1. The time series of chlorine variation for three days for a given sensor in the water distribution network of the city of Cannes, in France.

2. PROBLEM STATEMENT

A quantity follows an exponential decay process if it decreases at a rate proportional to its value. Mathematically, it satisfies the differential equation

$$\frac{\partial x(t)}{\partial t} = -\lambda \, x(t),$$

where $x(t)$ is the positive quantity at time t and λ is a positive number, often called the decay constant. Equivalently, the solution to this differential equation is

$$x(t) = x(0)\, e^{-\lambda\, t}, \tag{2}$$

where $x(0)$ is the initial quantity. When the parameter λ is constant, we can define other parameters, such as the (mean) lifetime $1/\lambda$ and the half-life $(\log 2)/\lambda$ of the process.

In real-life applications, the decay parameter λ may be time-varying, often by increasing at some instant. A more general model is

$$x(t) = x(0)\, e^{-\lambda(t)\, t}, \tag{3}$$

which allows to incorporate a process varying with time, e.g., having a decay parameter piecewise constant. As opposed to include further models to track the variations, e.g., the non-constant decay parameter $\lambda(t)$, we consider in this paper a machine learning approach.

We unfold the time-series into a phase space using time-delay embedding process [13], by converting the time-series $x(t)$ into a set of vectors

$$\boldsymbol{x}_i = [x(i-m+1) \ \cdots \ x(i-1) \ \ x(i)]^\top, \qquad (4)$$

for $i = m, m+1, \ldots$, and where m, the embedding dimension, is fixed in advance. [2]

With a set of (vector) samples $\{\boldsymbol{x}_m, \boldsymbol{x}_{m+1}, \boldsymbol{x}_{m+2}, \ldots\}$ at hand, one can consider classical machine learning techniques for pattern recognition or regression. Any kernel function can be applied, the most primitive one being the inner product,

$$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \boldsymbol{x}_i^\top \boldsymbol{x}_j, \qquad (5)$$

as well as the polynomial kernel, $\kappa_p(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i^\top \boldsymbol{x}_j + c)^p$, for $p > 0$ and $c \geq 0$. These kernels often perform poorly, as opposed to the Gaussian kernel

$$\kappa_G(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{1}{2\sigma^2}\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2},$$

where σ is the bandwidth parameter and $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|$ the Euclidean distance. The consistency of this kernel is "guaranteed" by the concept of universal kernels, as derived in [10].

The use of the Euclidean metric, with the conventional inner product or distance, within these classical kernels is not adapted for exponential decay time-series. To illustrate this issue, consider two samples, $\boldsymbol{x}_i, \boldsymbol{x}_j$ (e.g., two consecutive windows), drawn from the same constant-parameter process. It is easy to see that these measures of similarity and dissimilarity, i.e., $\boldsymbol{x}_i^\top \boldsymbol{x}_j$ and $\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2$, highly depend on the initial value, namely another sample $x(0)$. We propose to consider kernels that are independent of the multiplicative scaling x_0. The functions that satisfy the property of invariance under dilations are also called homogeneous of degree zero.

3. KERNELS FOR TIME SERIES OF EXPONENTIAL DECAY PROCESSES

A valid (reproducing) kernel function is a function that can be expressed as an inner product in some arbitrary space. By Mercer's theorem, this property is satisfied by positive semi-definite functions. Besides, one can engineer a valid kernel from other kernels by applying the simple rules given in Table 1. Moreover, one can provide valid kernels using a power series (polynomial or a Taylor series expansion) of other kernels, as long as the coefficients are non-negative.

3.1. Normalized kernels

As a first natural way to remove the dependence on the initial quantity, i.e., x_0, we consider the normalized kernel,

$$\kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j)}{\sqrt{\kappa(\boldsymbol{x}_i, \boldsymbol{x}_i)\,\kappa(\boldsymbol{x}_j, \boldsymbol{x}_j)}}, \qquad (6)$$

[2]The footnotes in this paper are given only for illustration:
We say $\boldsymbol{x}_i \sim \epsilon_\lambda$ if entries of $\boldsymbol{x}_i$, as defined in (4), satisfy (2) for some $x(0)$.

Rule	Expression
1. Linear combination	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \sum_k \gamma_k\, \kappa_k(\boldsymbol{x}_i, \boldsymbol{x}_j)$
2. Positive offset	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \kappa_0(\boldsymbol{x}_i, \boldsymbol{x}_j) + \gamma_0$
3. Product	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = \prod_k \kappa_k(\boldsymbol{x}_i, \boldsymbol{x}_j)$
4. Exponential	$\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{\kappa_0(\boldsymbol{x}_i, \boldsymbol{x}_j)}$
5. Normalization	$\kappa_0(\boldsymbol{x}_i, \boldsymbol{x}_j) / \sqrt{\kappa_0(\boldsymbol{x}_i, \boldsymbol{x}_i)\,\kappa_0(\boldsymbol{x}_j, \boldsymbol{x}_j)}$

Table 1. Some simple rules for engineering a valid kernel from available ones, with $\gamma_k \in R_+$.

where $\kappa(\cdot, \cdot)$ is the classical inner product defined in (5). This normalized kernel can be viewed as the (linear) inner product kernel applied after normalizing the data, namely by using $\boldsymbol{x}_i / \|\boldsymbol{x}_i\|$ for every $\boldsymbol{x}_i$. [3]

The normalization removes the scaling factor by projecting the data onto the unit-radius sphere, then applying the conventional inner-product kernel. This projection operation reduces the dimension of the data, while the classical inner-product kernel performs poorly in general. Next, we take advantage of the universality of the Gaussian kernel in order to provide more interesting kernels.

3.2. Kernels with logarithmic distance

We consider another way to derive kernels independent of the initial quantity, by modifying the metric and using the logarithmic distance instead, with

$$\|\log \boldsymbol{x}_i - \log \boldsymbol{x}_j\|,$$

where the logarithmic function $\log(\cdot)$ is applied to each entry of its (vector) argument. It is worth noting that this expression is independent of the base of the logarithm.

Therefore, one can easily replace the classical Euclidean distance with the above logarithmic distance, which allows us to refashion translation-invariant kernels into scale-invariant ones. This is the case of the Radial Basis Functions, with kernels of the form $f(\|\boldsymbol{x}_i - \boldsymbol{x}_j\|^2)$ such as the Gaussian kernel. The Gaussian kernel with logarithmic distance is defined by

$$\begin{aligned} \kappa_{G,\log}(\boldsymbol{x}_i, \boldsymbol{x}_j) &= \kappa_G(\log(\boldsymbol{x}_i), \log(\boldsymbol{x}_j)) \\ &= e^{-\frac{1}{2\sigma^2}\|\log \boldsymbol{x}_i - \log \boldsymbol{x}_j\|^2}. \end{aligned} \qquad (7)$$

This kernel can be equivalently obtained by considering the inner product (5) on the log-transformed data, namely $\kappa(\log(\boldsymbol{x}_i), \log(\boldsymbol{x}_j))$, and apply the Rule 4 (Exponential) and

[3]It is easy to see that, for $\boldsymbol{x}_i \sim \epsilon_{\lambda_i}$ and $\boldsymbol{x}_j \sim \epsilon_{\lambda_j}$, we have

$$\kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j) = \frac{\sum_\tau e^{-\lambda_i(i-\tau)-\lambda_j(j-\tau)}}{\sqrt{\sum_\tau e^{-2\lambda_i(i-\tau)} \sum_\tau e^{-2\lambda_j(j-\tau)}}}$$

where the summation is on $\tau = 0, 1, \ldots, m-1$, which is independent of the initial value $x(0)$.

the Rule 5 (Normalization) successively. This leads to a scale-invariant Gaussian kernel. [4]

The use of a logarithmic scale may sometimes be preferred on the linear scale. This is due to the property that this nonlinear transformation stretches low values, and shorten greater ones. Considering a distribution of samples $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$, then in the logarithmic scale, samples closer to the origin get further away from the "center" of the distribution[5]. Therefore, these samples are more likely to be outliers when applying the one-class SVM algorithm. This yields an interesting single-side detection property which, for the exponential decay data, corresponds to distinguishing large values of the decay parameter λ, namely when decay goes faster.

Finally, the logarithm can be viewed as a power transformation. By using a positive parameter μ, one can consider the family of power transformations with the so-called Box-Cox transformation, where each $x(i)$ is substituted with:

$$\begin{cases} \big(x(i)^\mu - 1\big)/\mu & \text{if } \mu \neq 0 \\ \log x(i) & \text{if } \mu = 0 \end{cases}$$

This transformation has shown its suitability in many statistical problems. This is the case when one needs to transform not normal-like distributed data into data that does follow approximately a normal distribution, namely turning skew unimodal distributions into nearly symmetric normal-like distributions. The use of this transformation allows us to provide a tunable transformation, with the tunable parameter μ, and therefore give a new class of kernels. However, this is beyond the scope of this paper.

3.3. More kernels

The normalized kernel (6) is a linear kernel applied on the normalized data. Based on this kernel, we propose two different kernels and study some of their properties.

Definition 1. *The logarithmic (normalized) kernel is defined as follows:*

$$\kappa_{\log,n}(\boldsymbol{x}_i, \boldsymbol{x}_j) = -\log(1 - \kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j))$$

Definition 2. *The inverse hyperbolic tangent (*artanh*) kernel is defined as follows:*

$$\kappa_{\operatorname{artanh},n}(\boldsymbol{x}_i, \boldsymbol{x}_j) = \operatorname{artanh}(\kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j))$$

[4] For $\boldsymbol{x}_i \sim \epsilon_{\lambda_i}$ and $\boldsymbol{x}_j \sim \epsilon_{\lambda_j}$, i.e., λ_i (resp. λ_j) is constant for all the entries of $\boldsymbol{x}_i$ (resp. λ_j), we have

$$\kappa_{G,\log}(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{1}{2\sigma^2}\sum_\tau |\lambda_i(i-\tau) - \lambda_j(j-\tau)|^2}.$$

For the same process, i.e., $\boldsymbol{x}_i, \boldsymbol{x}_j \sim \epsilon_\lambda$, we get

$$\kappa_{G,\log}(\boldsymbol{x}_i, \boldsymbol{x}_j) = e^{-\frac{1}{2\sigma^2}\lambda m|i-j|^2}.$$

[5] To show this, consider three values ($m = 1$ for illustration) with values $x(i) < x(j) < x(k)$, and $\|x(i) - x(j)\| = \|x(j) - x(k)\|$. Then, we get in the logarithmic scale: $\|\log(x(i)) - \log(x(j))\| > \|\log(x(j)) - \log(x(k))\|$.

Proposition 3. *The logarithmic (normalized) kernel and the* artanh *kernel are valid reproducing kernels.*

Proof. To prove this, we consider the expansion of each function into an infinite series:

$$-\log(1 - \kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j)) = \sum_{k=0}^{\infty} \frac{\kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j)^k}{k}, \qquad (8)$$

and

$$\operatorname{artanh}(\kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j)) = \sum_{k=0}^{\infty} \frac{\kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j)^{2k+1}}{2k+1}, \qquad (9)$$

where we have $-1 < \kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j) < 1$. The validity of these kernels results from applying rules of Table 1, namely Rule 1 and Rule 3. It is worth noting that these rules were also considered to show that the exponential kernel is a valid kernel, by writing it as an infinite series. □

These kernels are illustrated in Figure 2, as a function of the normalized kernel $\kappa_n(\cdot, \cdot)$. The artanh kernel is related to the logarithmic (normalized) kernel, since

$$\operatorname{artanh}(z) = \tfrac{1}{2}(\log(1 + z) - \log(1 - z))$$

for any z, and more specifically for $z = \kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j)$. Beyond these simple relations, the logarithmic (normalized) kernel defines a feature space that allows to approximate all continuous functions. This property is studied next, with the concept of "universal kernels".

Universality

The concept of universal kernels is proposed in order to study the generalization ability of machine learning classifiers, independent of the learning scheme. In [10], the authors establish the existence of the class of "universal" kernels, as the set of kernels that are consistent for a large variety of classification problems, provided a suitably chosen regularization. Examples of universal kernels include the exponential and Gaussian kernels. This is formalized here for the proposed kernel.

Proposition 4. *The logarithmic (normalized) kernel is a universal kernel.*

Proof. From the expansion of the logarithmic (normalized) kernel into an infinite series, as given in (8), this kernel takes the form

$$\sum_{k=0}^{\infty} a_k \left(\kappa_n(\boldsymbol{x}_i, \boldsymbol{x}_j)\right)^k,$$

with $a_k > 0$, for all $k \geq 0$. Due to Corollary 10 in [10] (given here[6] for completeness of this paper), we get the universality of this kernel. □

[6] *Corollary 10 in [10]*: Let $0 < r \leq \infty$ and $f\colon (-r, r) \to \mathbb{R}$ be a C^∞-function that can be expanded into its Taylor series in 0, namely $f(z) = \sum_{n=0}^{\infty} a_n z^n$ for all $-r \leq z \leq r$. If we have $a_n > 0$ for all n, then $\kappa(\boldsymbol{x}_i, \boldsymbol{x}_j) = f(\langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle)$ defines a universal kernel on every compact subset of $\{\boldsymbol{x} \in \mathbb{R}^d\colon \|\boldsymbol{x}\|_2^2 \leq r\}$.

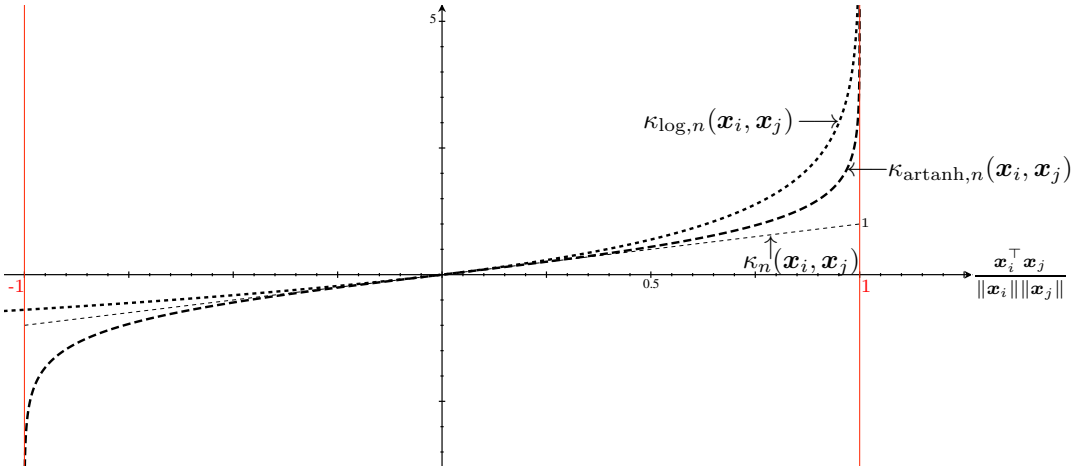

Fig. 2. Illustrations of some kernels, as a function of $\frac{\boldsymbol{x}_i^\top \boldsymbol{x}_j}{\|\boldsymbol{x}_i\| \|\boldsymbol{x}_j\|}$.

4. EXPERIMENTATIONS

Chlorine in water distribution systems

A drinking water system is a collection of pipes connecting different nodes, reservoirs, pumps and valves used to transport drinking water to consumers. To regulate the quality of drinking water, all pathogenic microorganisms must be destroyed using a chemical disinfectant. The best widely used and inexpensive disinfectant is chlorine. For this reason, the Environmental Protection Agency (USA) and many countries such as France adopt the injection of chlorine into water supply networks, and enforces regulations regarding the minimum amount of chlorine within a drinking water distribution system. Chlorine is injected at treatment stations, and chlorine concentrations decay over time essentially as water disinfection takes place (and reaction occurring at the pipe walls), while it must be present at the points of water consumption.

Chlorine sensors are expensive, costly to implement, and unfeasible to equip all the nodes for observability issues. Thus, control algorithms remain inefficient, making automated closed-loop control not yet widely accepted. Currently, open-loop manual control schemes are often investigated. Treatment stations inject just enough chlorine to maintain a minimum predicted chlorine residual everywhere in the network (typically the allowable minimum is 0.2 mg/l). Occasional offline measurements of chlorine concentration taken within some given nodes are sometimes used to adjust the chlorine input.

It is often assumed that the chlorine distribution within a single pipe is described by

$$\frac{\partial x(y,t)}{\partial t} + \nu \frac{\partial x(y,t)}{\partial y} = -\lambda\, x(y,t) \qquad (10)$$

where $x(y,t)$ is the chlorine concentration at position y within the pipe and time t, ν is the water velocity, and λ is the kinetics of water, also called first-order reaction constant. One also assumes that chlorine completely mixes at pipe junctions. For a single pipe, the exact solution of the above differential equation is given by

$$x(y,t) = x(y - \nu t, 0)\, e^{-\lambda t},$$

when $y > \nu\, t$, otherwise $x(y,t) = x(0, t - y/\nu)\, e^{-\lambda\, y/\nu}$.

In practice, the problem is more complicated. On the one hand, the first-order reaction parameter λ is a function of several parameters such as the wall reaction rate (which depends on the pipe material, roughness, and age, and cannot be directly measured), and the bulk reaction rate (itself a function of temperature, pH, ...). On the other hand, the water velocity ν is never constant, due to the ever fluctuating consumer demand (varies instantly, within the week, within the seasons, ...), while it is most of the time unknown (nominal values of aggregate water demand are often assumed). Combining all these uncertainties and variabilities, a model-based approach is unrealistic.

Experimental results

In this paper, we study chlorine variation at a single location (sensor), and propose to use domain description using one-class SVM. The measurements of chlorine were taken from the public water supplies of the city of Cannes, in France. We considered 3 days of chlorine concentration measures, with sampling at the rate of a sample every 3 minutes. This time series exhibits large fluctuations due to the variations in water consumption and an inefficient control system (no closed-loop control, only one sensor for chlorine observability). See Figure 1.

To capture the structure of the time series, a sliding window of length $m = 10$ was used, with

$$\boldsymbol{x}_i = [x(i-9) \ \cdots \ x(i-1) \ \ x(i)]^\top.$$

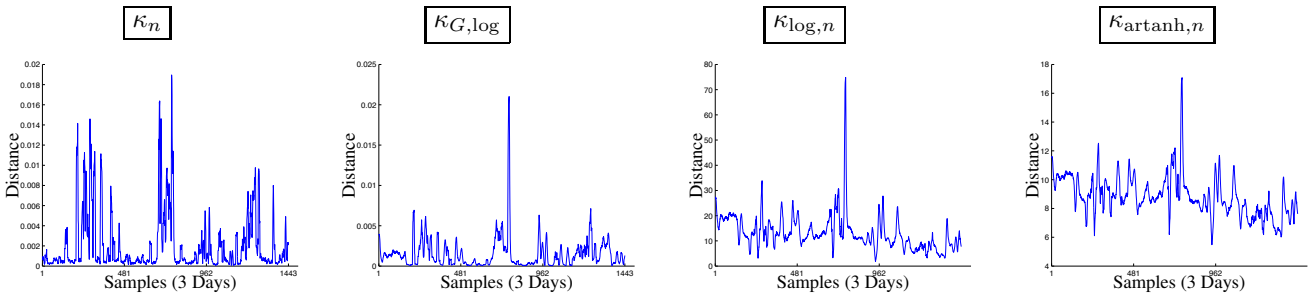

Fig. 3. The distance profile of all entries to the center of the one-class. The center is determined by classical one-class SVM algorithm, using (from left to right) the normalized linear kernel, the gaussian kernel with logarithmic distance, the logarithmic (normalized) kernel, and the inverse hyperbolic tangent (artanh) kernel.

As applications for the one-class SVM method, we considered the four types of kernel as studied in this paper:

- the normalized kernel, in (6);
- the Gaussian kernel with logarithmic distance, in (7);
- the logarithmic (normalized) kernel, in Definition 1;
- and the inverse hyperbolic tangent (artanh) kernel, in Definition 2.

Figure 3 shows the profile of the distance of data to the center of the one-class, as estimated by the one-class SVM algorithm, where different kernels were used. To get comparable results for all these kernels, the number of support vectors was set to 145, which corresponds to 10% of the samples. It is easy to see that the normalized kernel (left figure) provides the largest fluctuations, as opposed to the other proposed kernels. This illustration can be viewed as a better representation of the time series with the latter kernels. One may also consider a threshold (no illustrated) on the value of the distance, in order to provide a decision rule for one-class detection.

5. CONCLUSION

In this paper, we studied time series from exponential decay/growth processes, such as the chlorine concentration in a water supply network. We derived several kernels appropriate for these data, and studied some theoretical properties. Experiments conducted on real chlorine time series showed the performance of these kernels using a one-class SVM algorithm. As for future work, we are studying other theoretical properties, as well as the use of the proposed approach for classification and regression tasks. We are also investigating the use of these kernels when measurements from several nodes of the network are available.

6. REFERENCES

[1] N. Gebraeel and J. Pan, "Prognostic degradation models for computing and updating residual life distributions in a time-varying environment," *IEEE Transactions On Reliability*, vol. 57, no. 4, pp. 539–550, 2008.

[2] K. Kaiser and N. Gebraeel, "Predictive maintenance management using sensor-based degradation models," *IEEE Transactions on Systems, Man, and Cybernetics, Part A : Systems and Humans*, vol. 39, pp. 840–849, 2009.

[3] C. Gavrila and G. Tarlea, "Modeling water quality in drinking water distribution systems," *Revista De Chimie*, vol. 57, pp. 101–103, 2006.

[4] M.R. Islam, M.H. Chaudry, and R.M. Clark, "Inverse modeling of chlorine concentration in pipe networks under dynamic condition," *Journal of Environmental Engineering*, vol. 123, pp. 1033–1040, 1997.

[5] R.M. Clark and M. Sivaganesan, "Predicting chlorine residuals in drinking water: second order model," *Journal of water resources planning and management*, vol. 128, pp. 152–161, 2002.

[6] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer-Verlag, New York, 1995.

[7] D.M.J. Tax, "One-class classification," Phd thesis, Delft University of Technology, Delft, June 2001.

[8] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*, Springer, 1996.

[9] A. Graf, A. Smola, and S. Borer, "Classification in a normalized feature space using support vector machines," *IEEE Transactions on Neural Networks*, vol. 14, no. 3, pp. 597–605, 2003.

[10] I. Steinwart, "On the influence of the kernel on the consistency of support vector machines," *J. Mach. Learn. Res.*, vol. 2, pp. 67–93, mar 2002.

[11] Z. Noumir, P. Honeine, and C. Richard, "One-class machines based on the coherence criterion," *Proc. IEEE workshop on Statistical Signal Processing*, 5–8 August 2012.

[12] Z. Noumir, P. Honeine, and C. Richard, "One simple one-class classification methods," *Proc. IEEE International Symposium on Information Theory*, 1–6 July 2012.

[13] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, "Geometry from a Time Series," *Physical Review Letters*, vol. 45, no. 9, pp. 712–716, 1980.