Digital Signal Processing 54 (2016) 1–11

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A dynamic modeling approach for anomaly detection using stochastic differential equations

Yalda Rajabzadeh, Amir Hossein Rezaie *, Hamidreza Amindavar

Department of Electrical Engineering, Amirkabir University of Technology, Tehran, Iran

A R T I C L E I N F O

Article history:
Available online 6 April 2016

Keywords:
Anomaly detection
Stochastic differential equation
Diffusion process
Girsanov's theorem

A B S T R A C T

In this paper the stochastic differential equation (SDE) is utilized as a quantitative description of a natural phenomenon to distinguish normal and anomalous samples. In this framework, discrete samples are modeled as a continuous time-dependent diffusion process with time varying drift and diffusion coefficients. We employ a local non-parametric technique using kernel regression and polynomial fitting to learn coefficients of stochastic models. Next, a numerical discrete construction of likelihood over a sliding window is established using Girsanov's theorem to calculate an anomalous score for test observations. One of the benefits of the method is to estimate the ratio of probability density functions (PDFs) through the Girsanov's theorem instead of evaluating PDFs themselves. Another feature of employing SDE model is its generality, in the sense that it includes most of the well-known stochastic models. Performance of the new approach in comparison to other methods is demonstrated through simulated and real data. For real-world cases, we test our method on detecting anomalies in twitter user engagement data and discriminating speech samples from non-speech ones. In both simulated and real data, proposed algorithm outperforms other methods.

1. Introduction

Anomaly detection has received considerable attention in recent years and refers to the problem of extracting observations or patterns that do not behave according to the expected behavior of data. In other words, some alternative processes called anomalies cause deviation of system behavior from the normal situation. Many methods have been proposed for solving the anomaly detection problem which have employed different techniques such as statistical, classification and clustering, neural network, information and spectral theoretic based algorithms. Some applications of the anomaly detection are presented in medical and public health, industrial process monitoring, abrupt changes and irregular vision motion detection. Various authors have studied the anomaly detection problem quite extensively; complete surveys on these techniques and applications could be found in [1] and the papers cited therein.

Among different anomaly detection methods, statistical techniques have a vast importance in finding discordant observations in complex systems. Stochastic fluctuations in complex systems change the anomaly detection problem to a challenging task. Statistical anomaly detection methods in stochastic dynamics rely on fitting a parametric or non-parametric model or distribution to normal observations and extracting the nominal behavior of the system, then employing a statistical inference test to detect observations that do not belong to this model. Most studies have investigated a known underlying model, namely autoregressive (AR) [2], autoregressive conditional heteroscedasticity (AR-ARCH) [3], generalized autoregressive conditional heteroscedasticity (GARCH) [4].

The main drawback in these techniques is that the performance of such techniques is highly dependent upon the choice of the distribution or model that the data is generated from. The performance of the method will be poor if the model is under-specified, this means the true data generation model is more complex than the model used in the anomaly detection technique. This problem is even more challenging when non-linearity and non-Gaussianity are considered, so in practice a more flexible model is required.

To overcome these problems, we introduce a new application of stochastic differential equation in anomaly detection; which is modeling system dynamic uncertainty by a global and more flexible model that relies on milder assumptions in comparison with other parametric models. Our key idea is based on the fact that all statistical information of a time series could be achieved from its PDF while PDF estimation without a specific model assumption is known to be a hard problem [5]. Since the PDF of a process can be represented by a partial differential equation known as

* Corresponding author.
E-mail addresses: yalda_raj@aut.ac.ir (Y. Rajabzadeh), rezaie@aut.ac.ir (A.H. Rezaie), hamidami@aut.ac.ir (H. Amindavar).

http://dx.doi.org/10.1016/j.dsp.2016.03.006
1051-2004/

the Fokker Planck equation whose parameters are equivalent to its corresponding SDE [6–9], therefore, time series inference through SDE and PDF are analogous approaches. On the other hand, SDE is a continuous model which is measured at discrete instances of time so it is a continuous-discrete model, hence, it inherently assimilates the irregular space sampling in the formulation of the model which is not attained in other popular models. We note that continuous processes could encompass discretized processes as their subset, hence, it is more convenient to model the underlying processes in the continuous domain. In other words, this filtering model is a generalization of discrete time series models; many discrete time stochastic models can be interpreted as a SDE being sampled at discrete time; for example popular models such as AR, ARMA, ARCH and GARCH, are extensions of a particular type of stochastic differential equation [6,9–12].

In this paper, we propose a new algorithm employing SDE models to detect anomalies in stochastic processes. By the assumption that we have a training dataset that describes the normal operation of the system, we specify whether there exists an anomaly in the test data. This procedure consists of the following stages:

First we apply local polynomial kernel regression to estimate SDE coefficients. The conventional non-parametric parameter estimation method [8] is a batch algorithm where all samples are needed for estimating PDF and it also requires too many data samples. Beside this, for the non-stationary time series an ensemble of experimental realizations of process is necessary which is not generally available. Another method for short and non-stationary data is introduced in [13] based on some assumptions that may not be practically valid. Accordingly, we use an extension of the framework in [14] by utilizing polynomial fitting to obtain a closed form estimator for SDE coefficients in non-stationary case as well as stationary one.

Next, we construct the log-likelihood of the model parameters over a sliding window using Girsanov's formula instead of PDF. We use the log-likelihood value as an anomaly indicator for making a decision about the presence of anomaly in the test data. We would note that by using Girsanov's formula, the log-likelihood score is calculated without estimating the PDF of the process, a formidable task, since it is hard or impossible to find an explicit expression for PDF especially in non-linear time series.

The organization of the paper is as follows. In the next section, we briefly review stochastic differential equation modeling. Problem formulation from a statistical perspective, relation to SDE modeling as well as parameter estimation and log-likelihood calculation are discussed in section 3. Power and performance of this procedure is examined through Monte Carlo simulations in section 4; its application in twitter user engagement data and voice activity detection is also presented in this section. Finally, some conclusions are drawn in section 5.

2. Preliminaries

In this section, we represent a brief review of SDE and its parameter estimation approach.

Stochastic differential equations are equations that model dynamic characteristics of complex systems. They simplify quantifying uncertainty of the time series while preserve momentous information of them [8]. Stochastic differential equations are inherently extensions of ordinary differential equations (ODEs). They provide flexibility in choosing the stochastic component and it is not restricted to a constant-variance Gaussian process. General representation of a continuous Markovian process $\{X_t = X(t)\}_{t\geq 0}$ in the form of stochastic differential equation is as follows

$$dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t \tag{1}$$

here $\{W_t = W(t)\}_{t\geq 0}$ is the Wiener process which is the source of randomness, $\mu(\cdot,\cdot)$ and $\sigma^2(\cdot,\cdot)$ are referred to as the time and state dependent drift and diffusion coefficients, respectively. The Wiener process is a Gaussian delta-correlated noise with continuous path and vanishing mean. $\mu(\cdot,\cdot)$ presents the deterministic component which specifies the nominal dynamics of the system and $\sigma^2(\cdot,\cdot)$ determines the stochastic dynamics and represents how the noise affects the system. In fact, by SDE modeling, we separate and quantify deterministic and stochastic dynamics of the system.

Note that the diffusion coefficient is formulated by a function of time and process X_t and is not a constant while in most statistical models like autoregressive models, noise is introduced into the model by a constant-variance Gaussian process which is additively superimposed on the trajectory generated by a deterministic dynamic. The constant-variance Gaussian noise is not a realistic assumption in practice, especially in complex systems that are composed of many microscopic subsystems and exposed to numerous random influences.

(1) results in a solution that is a Markovian, continuous time stochastic process which is called a diffusion process. In [6,15] one can find sufficient conditions for existence of a solution to (1).

The same information in (1) is contained in the corresponding Fokker Planck equation expressing the temporal evolution of probability density function $P(X_t, t)$ of X_t:

$$\frac{\partial P(X_t,t)}{\partial t} = -\frac{\partial}{\partial X_t}\mu(X_t,t)P(X_t,t) + \frac{1}{2}\frac{\partial^2}{\partial X_t^2}\sigma^2(X_t,t)P(X_t,t) \tag{2}$$

As mentioned before, one important advantage of a SDE approach is based on (2) in which SDE coefficients characterize how the PDF evolves over time without getting involved in the high computational cost of PDF estimation. So, it is more convenient to work with the drift and diffusion terms rather than PDF.

There are two techniques for estimating drift and diffusion coefficients: parametric and non-parametric. Our discussion is confined to the systems which we do not have exact knowledge about the system dynamics in advance. Since the focus in this paper is on the non-parametric estimation, for information about parametric parameter estimation please refer to [7]. Two well known estimators in finite time step observations for $\mu(x,t)$ and $\sigma^2(x,t)$ are described as

$$\begin{aligned}\hat{\mu}(x,t) &= \lim_{\tau\to 0}\frac{1}{\tau}\int (X_\tau - x)P(X_\tau, t+\Delta t|x,t)dX_\tau \\ \hat{\sigma}^2(x,t) &= \lim_{\tau\to 0}\frac{1}{2\tau}\int (X_\tau - x)^2 P(X_\tau, t+\Delta t|x,t)dX_\tau,\end{aligned} \tag{3}$$

here $\hat{\mu}(x,t)$ and $\hat{\sigma}^2(x,t)$ are seen as instantaneous conditional mean and variance of the process, respectively [7]. In this approach the conditional PDF of observations, denoted by P in (3), is derived by histogram.

When τ is not small enough, first drift and diffusion coefficients are estimated by (3), and then extrapolate to $\tau = 0$. Details of implementation are presented in [16].

3. Problem formulation

In this section, we first describe a mathematical representation of the problem from statistical point of view, then the proposed method is presented.

3.1. Mathematical representation

Let $\{X_{1:N}\}$ be a sequence of observations from a continuous process X_t that are sampled at even or uneven time intervals $\Delta_i t$ at $\{t_1,\ldots,t_N\}$. The training dataset $\{X_{1:k}\}$ contains stationary observations sampled at normal operation of the system. Our goal is