We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900
Open access books available

186,000
International authors and editors

200M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Chapter 1

Introductory Chapter: Time Series Analysis (TSA) for Anomaly Detection in IoT

Nawaz Mohamudally

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.72669

1. Introduction

Observing data points over time with proper transform may reveal valuable information about systems behaviours and trends. This book entitled "Time Series Analysis (TSA) and Applications" comes at a very opportune period where business enterprises are overloaded with data and looking for swift analytics and on the other hand have not yet trusted the powerful algorithms such as deep learning and AI. Academics prefer simple tools like Matlab or Mathematica to run TSA. However, statistics and probabilistic instruments have gained wide acceptance for decades. Time Series Analysis had been often assimilated to finance and forecasting. The chapters presented here prove the contrary and show how far TSA is being applied across an array of disciplines and how efficient and effective this technique could be if it is fittingly utilised. In the same spirit, this chapter provides an overview of time series as applied to detect anomalies in Internet of Things (IoT) networks. Specific attention is paid to anomalies that occur in smart cities IoT use cases. The final aim of this research work partly described here is to mount plug n play anomaly detection engine (ADE).

The Internet has evolved from its original aim of providing access to web resources globally to what is commonly called today Internet of Things, where it is expected that objects will internetwork and have a presence on the Internet just with an IPv6 address for example. The objects market is estimated in billions and trillions, very far from the global human population. This has led to new business models with development of dedicated IoT networks such as SigFox, LoRa, Symphony Link, and NB-IoT, and production of IoT compliant devices from microcontrollers' manufacturers such as Microchip, Intel, and Raspberry PI. Software companies have come up with virtual machines and statistical tools for big data analytics whereas network devices constructors like Cisco and Juniper for instance have come up with network

INTECH
open science | open minds

© 2018 The Author(s). Licensee InTech. This chapter is distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. CC BY

Figure 1. IoT Value Chain.

gateways and routers to accommodate devices connection, routing, and IoT data transit. The myriad of technologies involved within the IoT ecosystem should empower smart environments as it happens likewise in smart cities. The next section introduces the IoT value chain and then lists some use cases of IoT in smart environments whereby anomalies arouse, followed by the classification of anomalies in the time domain, the time series models applicable and finally problematics in applying TSA to anomaly detection in IoT.

2. IoT value chain

The IoT value chain in **Figure 1** shows that the value added services to IoT & key differentiator is the data analytics part which comprises the anomaly detection component with the help of TSA. Data analytics in IoT could be a higher income generator than key technology enablers like SDN, IPv6, and 5G, even more than machine automation. We are talking about Analytics as a Service (AaaS). According to Cisco's annual Visual Networking Index, machine-to-machine (M2M) connections that support IoT applications will account for more than half of the world's 27.1 billion devices and connections by 2021.

2.1. Anomalies categories

Table 1 illustrates the anomalies descriptions in selected smart cities IoT use cases.

Let us now classify the anomalies in the time domain.

i. *Static vs. dynamic*: anomalies are defined as data points not following current patterns; static means in the same direction but with different characteristics whereas dynamic refers to opposite direction.

Smart	***Anomalies description***	***Benefits***
Water	Water leakages	To prevent water waste
Lighting	Broken bulbs	Save time and fuel for maintenance
Home	Gas leakage	Alert home users on the incident
Building	Electricity peak and pipe leakage	Energy monitoring
Farm	Anomalies in farm data and weather	Monitor growth
Goods	Traffic congestion spots	Optimize route and delivery

Table 1. Benefits of Anomaly Detection in Smart City Applications.

http://dx.doi.org/10.5772/intechopen.72669

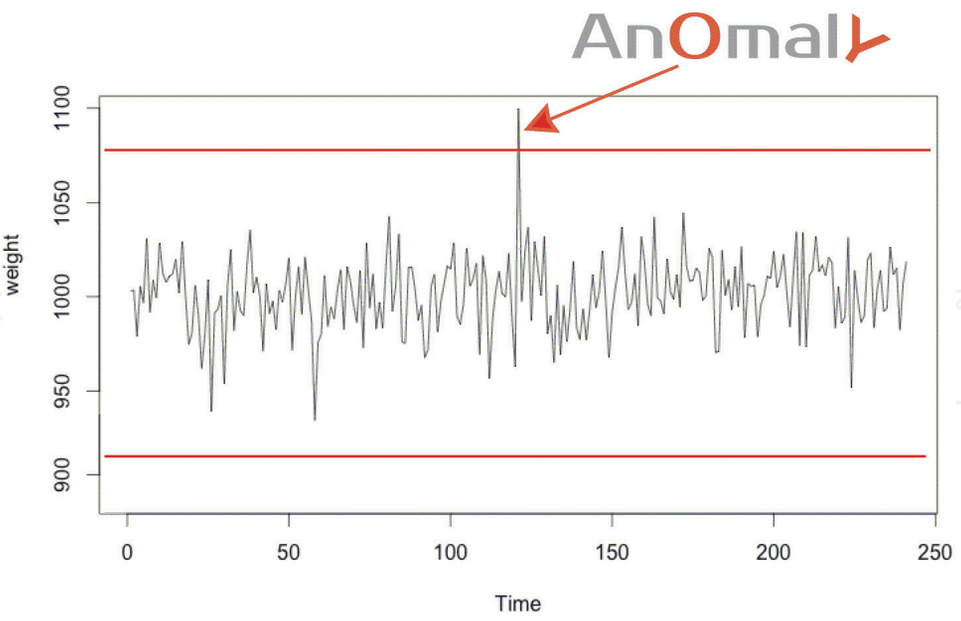

Figure 2. Outlier anomaly (https://anomaly.io/anomaly-detection-normal-distribution/).

ii. *Outlier*: an outlier is not necessarily an anomaly; it all depends on the defined threshold, for instance in the example in **Figure 2** showing sugar bags weight with respect to time, any bag <920 g or >1080 g is considered as an anomaly.

iii. *Contextual*: a data point could be an anomaly in one context but not in another. For example, a temperature of 35°C in January is an anomaly in a northern European country but normal in a southern hemisphere island for the same month.

iv. *Collective*: this happens when there is elongation in time of a particular anomaly like it happens in telecom transmission; there are accumulation of delays that result in jitters.

2.2. Time series models

There is actually no one size fit all solution for the development of an ADE as well as no de facto time series model that suits the ADE. Below are some of the popular time series models adopted for ADE in IoT.

i. *Autoregressive models*: an autoregressive model specifies that the output variable depends linearly on its own previous values. It is based on an approach that several points from the past generate a forecast of the next point with the addition of some random variable, which is usually white noise. The autoregressive integrated moving average (ARIMA) is applicable to stationary time series only.

ii. *Symbolic TSA*: data points are converted to bits and bytes 10100111001; then, Information Theory; Shannon, FFT, DFT, DWT is applied.

iii. *Seasonal-trend-Loess* (*STL*) *decomposition*: data points together with the noise or multiple data sets over a period are decomposed and analyzed to detect eventual anomalies.

iv. *Machine learning*: there are two main branches of machine learning namely supervised learning whereby the pattern for the anomaly is learnt and known, whereas in supervised mode, detection is done by inference or featuring. The latter is more challenging as the anomaly pattern is unknown and the algorithm learnt from the data points is to be analyzed. The supervised mode comprises the following methods: Decision Table, Random Forest, K-nearest Neighbor, SVMs, Deep Learning, Naive Bayes. The popular "*unsupervised*" algorithms are K-means clustering, DBSCAN, N-SVM, Stream Clustering, and LDA (Latent Dirichlet Allocation).

2.3. Problematics

Below listed are the 10 main issues, in which some are inherent to the IoT network and others to the time series properties.

i. Missing data points/holes: missing data can happen due to device malfunctioning, for instance, or issues related to device identification for example. "Potent, climate warming gases are being emitted into the atmosphere but are not being recorded in official inventories," a BBC (http://www.bbc.com/news/science-environment-40669449) investigation has found.

ii. Data corruption: for instance, data can be corrupted due to external factors or device malfunctioning; thus, it is important to ensure that the data points analyzed are accurate and come from the system under investigation.

iii. Encrypted data: in most IoT networks, data are encrypted during transmission and normally decrypted for customer usage. If detection is to be performed on encrypted data, anomaly detection might not be straightforward.

iv. Sensor fusion: data points from different sensors can be aggregated for a specific function. For example, different parameters like temperature, carbon footprint, wind speed can be captured from different sensors and merged for modelling on a server for environmental impact study. In such cases, the TSA needs to deal with multiple datasets. Sensor fusion is also assimilated to evolving sources.

v. Real-time detection: this is probably more inherent to the network itself, but the processing and programming aspects of the TSA are also determinants.

vi. Seasonality: also called as periodic time series, arrives when the time series is influenced by the seasonal factors such as day, night, month, and so on.

vii. Heteroscedasticity: it involves frequent changes in variances that can render the transformation of the time series more complex.

viii. Noisy data: data points with very low amplitude can be drowned into the intrinsic transmission electronic noise. Network equipment vendors are proposing edge computing routers that would actually clean the IoT device data in a closer location prior to run the complete analytics on the cloud.

ix. Traffic surge: at times, there could be excessive throughput like number SMS on the eve of New Year that could bring an overload on the ADE.

x. Non-linearity: date points that are not stationary and changing with time would require multivariate analysis.

3. Conclusion

This chapter highlights the challenges relevant to core elements involved in the development of an anomaly detection engine (ADE). It was found that an accurate and reliable ADE relies on three main selection factors namely, the quality of the data points, the time series transformation, and where analytics are executed. Moreover, due to the heterogeneous nature of networking environments, the convergence of communication and data protocols in IoT requires special attention when it comes to anomaly detection software development. For instance, raw data points from a smart water application are surely completely different from that from a health care IoT application; hence, the domain of application is another determinant factor in the construction of an efficient ADE. Machine learning in the unsupervised mode is indeed very efficient in situations where datasets are unpredictable. Moreover, cases where data points show nonlinear time series require multivariate analysis that makes the process more computing intensive. This property is not favorable to real-time anomaly detection as more computation at the ADE level will affect the accuracy of the ADE. From a software development perspective, the trend is similar to data mining tools embedded in popular database servers. Once the dataset is compiled, the user can choose the most appropriate statistical tool. In a near future, ERP solution providers will probably propose the ADE as a customizable module that would best fit the customers' requirements. Future work will investigate into the challenges from empirical experimentations and how anomaly detection can be translated as a service in cloud computing.

Author details

Nawaz Mohamudally

Address all correspondence to: alimohamudally@umail.utm.ac.mu

University of Technology, Mauritius, Port Louis, Mauritius

IntechOpen

IntechOpen