Chapter 11

Adaptive Principal Component Analysis–Based Outliers Detection Through Neighborhood Voting in Wireless Sensor Networks

Ekaterina Aleksandrova
University of Glasgow, UK

Christos Anagnostopoulos
University of Glasgow, UK

ABSTRACT

This chapter introduces statistical learning methods and findings of a group decision-making algorithm in internet of things (IoT) and edge computing environments. The discussed methodology locally detects outliers in an on-line and adaptive mode. It is driven by three perspectives—opinion, confidence, and independence—and exploits the incremental principal component analysis using the power method for eigenvector and eigenvalue estimation and Knuth and Welford’s online algorithms for variance estimation. The methodology is implemented and evaluated over real contextual data in a wireless network of environmental sensors from where appropriate conclusions are drawn about the capabilities and limitations of the proposed solution in IoT environments.

INTRODUCTION

In the growing world of the ‘Internet of Things’, more and more computing and sensing devices are joining the wireless sensor networks (WSNs) at the network edge allowing on-line contextual data processing and localized knowledge exchange. However, keeping a WSN stabilised only a part of it is having a fully usable environment. In order to make the exchanged contextual data logically consistent to other computing and sensing devices, contextual knowledge has to be locally processed and only suf-
Adaptive Principal Component Analysis-Based Outliers Detection

Efficient statistics are sent to the relevant destination, e.g., sink nodes, micro-data centres, edge gateways, and IoT applications. This would require devices to have the computing capabilities and resources for making on-line calculations and adaptive transformations of the gathered contextual data, presenting a high energy cost for the whole WSN. Such concerns in the past have led to the idea of a centralised computing power in a dematerialised system with enormous capacities and simple architectures, also named as the ‘Cloud’ (Lopez, 2015).

Cloud computing has made the current state of IoT environments a possibility by completely transforming the data cycle but at certain costs. Having all contextual data transferred to a centralised data centre involves a significant communication overhead. Furthermore, the required computational power for processing the data varies from device to device, as well as the storage capacity for preserving the results (Sharma & Wang, 2017), nevertheless, that increases the need for more powerful data centre cores. Nowadays, such problems have brought the highlight back to the ‘edge’ of the network given that modern devices have gained considerable computational abilities. The new concept allows devices with smaller capabilities than data centres but powerful enough to process any incoming contextual information from any closely located other devices/sensors and if needed pass the data to the core of the system.

Maintaining consistency in the exchanged data requires not only lossless contextual data transfer but also identifying and correcting corrupted and/or intentionally altered pieces of data and/or dimensions. Such pieces of data deviate from the rest and are known in the literature as “outliers” (Deng, 2016). The existence of outliers in distributed environments and WSNs demands for the development of a powerful, local and adaptive machine learning technique for outlier detection within neighbours/cliques of a WSN, which does not introduce higher network bandwidth or energy consumption (Deng, 2016) (Anagnostopoulos & Hadjiefthymiades, 2014) (Harth & Anagnostopoulos, 2017).

Aim

The aim of this chapter is to propose and analyse an on-line adaptive statistical learning algorithm for outlier detection in wireless sensor networks based on the Principle Component Analysis technique via different voting policies among sensors belonging in the same WSN neighbourhood/clique. The objective is to incrementally and locally learn the minimum sufficient statistics over each sensing and computing device in a WSN such that, based on the opinion of the neighbouring nodes via voting, the outlier detection process is revised to further improve the true positive rate.

Problem

In a summarised form outliers (also called anomalies) are defined as “statistical deviations from the majority of existing data” (Deng, 2016). And when there is such a tremendous amount of contextual and time-series data collected in a present-day IoT environment, there is also a very high probability that a part of that data is corrupted by faulty devices, or it reflects an extreme change in the scanned environment or is alternated with malicious intentions. Having such anomalies would result in unexpected future decisions based on the processed data.

The existence of outliers in a distributed system calls for finding a powerful technique for outlier detection. But in order to be most beneficial from the added extra computations, the algorithm has to be efficient, without a substantial increase in the network bandwidth or power consumption (Deng, 2016), (Anagnostopoulos, Hadjiefthymiades, & Georgas, 2012) (Anagnostopoulos & Hadjiefthymiades, 2014).