April 2022

DATA QUALITY FOR IOT

Mohsin Alam

Juan Cazila

Fabio Salvemini

DATA QUALITY FOR IOT

AUTHORS:
Mohsin Alam
Juan Cazila
Fabio Salvemini

ABSTRACT

Techniques are described herein to monitor the data quality of sensors, edge computing and other devices. This may provide users with information regarding the quality of data provided.

DETAILED DESCRIPTION

A critical challenge today is understanding the quality level of "new" data before making critical decisions in everyday operations (e.g., Network, Applications, or Industrial). The Internet of Things (IoT) environment should be kept updated with unified control and reporting, because of the impact of data quality depreciations over time. Users need additional input regarding critical operational decisions. More efficient decisions reduce industrial risk, save lives, reduce operational costs, and improve user productivity.

For example, stopping an oil and gas operation may be necessary if lives are at risk, but can result in hundreds of thousands of dollars lost every minute; thus, data quality is critical to ensure that the operation is stopped when necessary, but not needlessly so.

IoT environments usually prioritize areas such as industrial safety, productivity, operating expense reduction, and sustainability (green). Users also face diversity and integration challenges. Too many systems are in silos with poor integration, and are therefore hard to follow and control: much data comes from different systems without effective correlation.

In accordance with techniques described herein, IoT sensors and devices (edge computing) may be categorized in different areas covering Central Processing Units (CPUs), memory, battery status, physical security tampering, cyber security aspects, etc. A mechanism is provided to generate data “metering,” location, environment aspects (e.g., pollution, gas influence, etc.), software version, bug considerations, etc. The mechanism

may be active in real time (e.g., as of the day and time of the sample, last booting time, last administration connection, etc.).

Every category may have a respective assigned value and may be considered in a unified dashboard to obtain the correct Error Probability Sample (EPS) "accuracy" value that may be associated with the sample value in the central console. An IoT manager or central console may work with the "accuracy" levels in an independent mode or may group several values that will be considered as data to inform a single operational decision. For example, ten sensors may send an alert to the central console, and a single "accuracy" value may consolidate hundreds of samples received from these ten sensors. Thus, the platform operator may have a single value to inform the correct decision in a critical moment.

As illustrated in Figure 1 below, a Sensor/Application Logical Unit (SLU) engine may receive information from the server and qualify the internal performance of the sensors. An EPS algorithm may consider information from multiple Autonomous Systems (ASs) (e.g., sub-systems within a network). An analytics engine (real-time and/or batch) may use the EPS value before making important decisions regarding the systems. The analytics engine may also pass information to the SLU engine and EPS algorithm when data information is suspected.

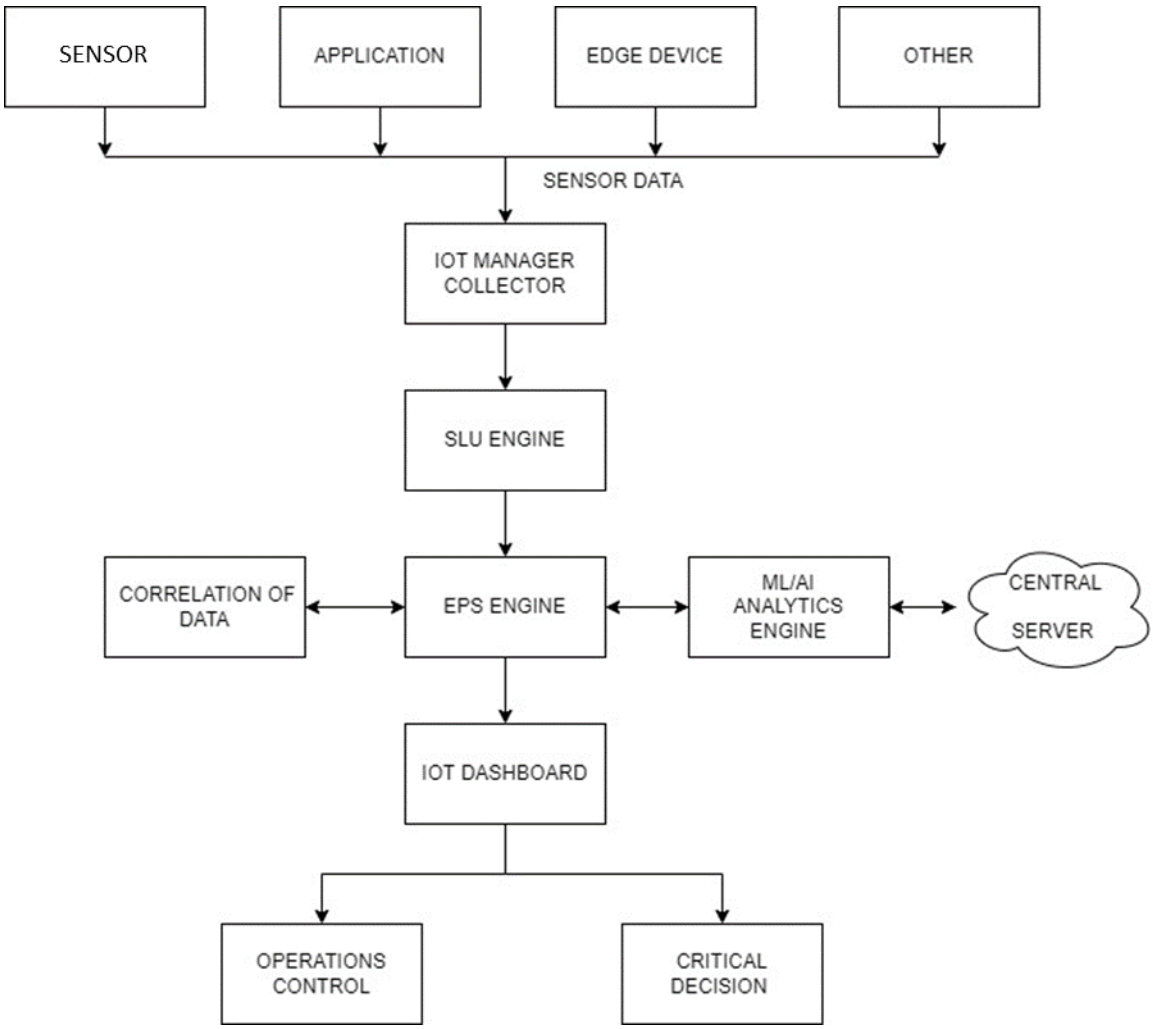

Figure 1

An SLU may be assigned for each "data generator"/sensor. The SLU engine may assign the SLU value considering the model, Operating System (OS), battery level, calibration state, security, etc. If needed, the SLU engine may re-assign value in real-time operation when suspicious values come from battery, security gaps, processors, and others, and may penalize the SLU value accordingly. The SLU engine may be connected to servers, and daily updates may improve the performance and decisions for the analytics.

An analytics engine may use the EPS algorithm in real time to perform the percentage level of error of the total received samples/data/information before making decisions or presenting a display in the operation dashboard (e.g., for human-based decisions).

The EPS algorithm may use the SLU value to calculate the Optimum Possible Total Sample Value (OPTSV) versus the Resultant Total Value (RTV), and may inform the analytic engine of this value.

The analytics engine may also use the RTV to understand how the Quality of Value (QoV) is changing over the time, if at all. The analytics engine may store the RTV or EPS in a database and generate a graph to correlate the same.

Figure 2 below illustrates a graph of RTV over time.

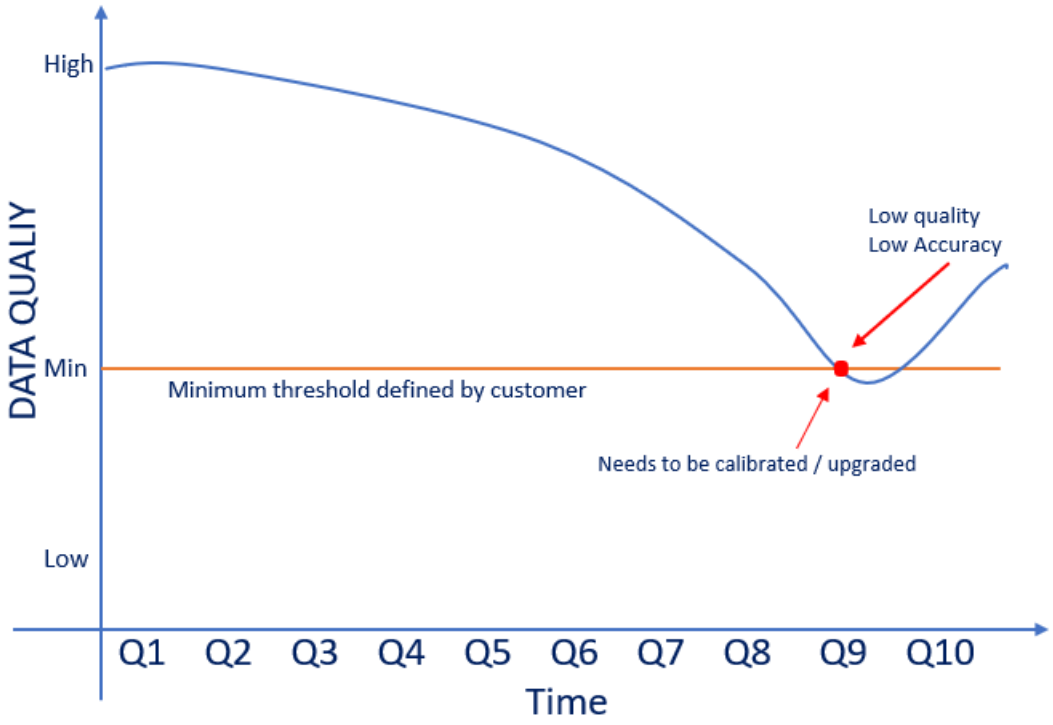

Figure 2

The analytics engine may enable the customer to set a reliable alert to be sent if the RTV falls below threshold. The analytics engine may receive information from multiple subsystems and sensors. This data information may match an SLU parameter (including the applications). Thus, the analytics engine may take actions considering not only the "data" but also the SLU value or the RTV.

Sensors and "edge" computing routers that are processing data may also be added in the process to provide data to the centralized dashboard.

AI/ML supervised algorithms may help produce an enhanced data quality model. The algorithm(s) may be trained using a plethora of data from one or multiple users, and the categorization may be entirely automated within few years.

Also, non-supervised AI/ML algorithms may help categorize third party samples that come from sensors of various vendors and may serve as a foundation for a platform for multiple environments.

In summary, techniques are described herein to monitor the data quality of sensors, edge computing and other devices. This may provide users with information regarding the quality of the data provided.