Title
Sensor Network Data Fault Types

Authors
Ni, Kevin
Ramanathan, Nithya
Hajj Chehade, Nabil
et al.

Publication Date
2007-10-10

Peer reviewed

Sensor Network Data Fault Types

Kevin Ni, Nithya Ramanathan, Nabil Hajj Chehade, Laura Balzano, Sheela Nair, Sadaf Zahedi, Greg Pottie, Mark Hansen, and Mani Srivastava

Introduction: Sensor Network Data Faults

Data faults are common in deployments

- **Fault detection is a difficult and complex task**
  There are many factors that influence data and could cause faults.
  Faults are application and sensor type dependent.
- **Fault detection systems must be based on models of sensor behavior, the environment, and models of faulty data**

Lack of a detailed study of sensor faults

- **There is no significant list of features to consider modeling**
  - A detailed list of features provides for systematic description of faults
- **A taxonomy of faults aids in fault detection by providing known fault models with suspect behavior**
- **A list of actual faults will also aid in testing fault detection systems**

Description: Fault detection design considerations

Modeling data and faults

- **The set of models to be used for fault detection contain models for both good and faulty data**
- **Models are based on many features of a sensor network.**
- **Data is then classified by the model that most closely matches a selected feature vector.**
- **Human input is required to update and create models and validate other unusual measurements**
- **Data is available to the fusion center where decisions on sensor quality are made**

Sensor network features

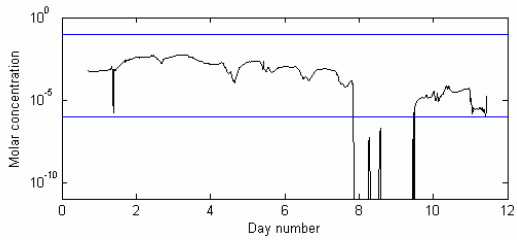
- **Much of the features and modeling must be based on prior knowledge determined by human input.**
- **The features that help define expected behavior have three major categories**
  - Environment features – This is the context in which a sensor is deployed. It also includes the sensing application and location.
  - System features – This includes sensor behavior and limitations. Also may include different modalities of a sensor network
  - Data features – These are features of the data such as mean, variance, and correlation. They can be considered in both time and space domains, although they are not limited to these two.

Taxonomy of Faults and Application

Faults

- **Faults are defined relative to a model of expected behavior**
  A fault or anomalous behavior can only be deemed as such if it is out of bounds of the expected behavior.
- **Two distinct approaches to defining faults**
  - The data-centric view examines the data produced by a sensor and describes fault models based on data features
  - The system view defines a physical malfunction with a sensor and how this feature may manifest itself in the data
- **These two views are related to one another and faults can be mapped between the two.**
  It may be easier or more convenient to define a fault in one domain
  One fault in one domain may map to multiple faults in the other domain
- **Data-centric faults**
  - Outliers – A single isolated event that is outside of the expected range of values to be returned.
  - Stuck-at faults – A series of data values with little or no variation for a period of time longer than expected
  - Spikes – A change in gradient over a period of time much greater than expected
  - Excessive Noise – Data exhibits much higher noise than expected, but may still track the phenomenon.
- **System faults**
  - Calibration fault – Calibration errors can cause sensor data to exhibit unexpected offsets, gains, or drift
  - Hardware fault – A very general heading for any part of the sensor to malfunction. This can manifest itself many ways
  - Low Battery – Low battery affects the sensor in different ways as well.
  - Environment out of range – the actual phenomenon exceeds the detection capabilities of the sensor
  - Clipping – related to environment out of range, the data tends to exhibit "rooftops" and "floors".
- **Faulty data can be classified in three ways**
  - The faulty data still provides some useful information about the particular phenomenon of interest. Any conclusion will have greater uncertainty
  - The faulty data is totally useless and can be discarded
  - The data is irrelevant to the sensing application and can be ignored.

Application



- **We give an example of an approach to analyzing a specific fault.**
- **Ammonium concentration reported from a sensor in soil over the course of several days.**
- **Determining whether or not a fault exists requires looking at several features to determine expected behavior**
  - Data exceeds the detection capability of the sensor.
  - Data is not consistent with a reasonable environmental model based upon the a model of the phenomenon. This includes sharp changes in concentration, and concentration being reported outside of a reasonable range.
- **From a system perspective**
  - We can also examine environmental changes, such as irrigation
  - Perhaps water caused a short in the hardware.
  - The environment is not expected to be out of the range of the sensor.
- **The behavior after "recovery" is also questionable**
  - One can exploit other characteristics such as spatial correlation to confirm the behavior