Journal of Computer Science and Applications.
ISSN 2231-1270 Volume 8, Number 1 (2016), pp. 31-39

http://www.irphouse.com

Semantic Filtering of IoT Data using Symbolic Aggregate Approximation (SAX)

R. Mahalakshmi
MCA, M.Phil, Lecturer, Thyagarajar College, Madurai, India.

Dr. S. Kannan
Associate Professor, Madurai Kamaraj University, Madurai, India.

Abstract

Semantic Filtering is a process of annotating raw data in order to derive meaningful information from it. Devices or Things in Internet of Things (IoT) generate a huge amount of sensed raw data. The huge sensed raw data has to be interpreted meaningfully with the context, to facilitate Human / Machine perception. Meaningful interpretation of raw data can be achieved through Semantic Filtering. And Semantic Filtering provides Semantic Annotations and Abnormal Event detections from the raw sensor data. Data dimensionality is also reduced in the process which results in reducing the storage space needed for this data.

INTRODUCTION

Internet of Things (IoT) is an evolving technology that connects everyday devices to internet in order to achieve smart solutions. All these devices and objects, from home appliances to bicycles, sport watches to environmental sensors, traffic lights and tools used in factories have their unique identifiers and communicate with other peers and servers through Internet. So IoT is essentially a network of Things or devices that can sense its surroundings, communicate the sensed information with other connected devices or Applications using Internet in order to achieve smart solutions even without human intervention. Devices in IoT could be simple, sensor embedded or connected objects like Air Conditioners, Environmental Monitors, and Security systems or complex inventory control devices deployed in `

Factory Automation Systems. These devices generate huge amount of data continuously. And IoT Device generated data is heterogenic, plain and raw in nature. Clearly, the sensed data from the IoT devices is characterized by 3V's of Big Data (Volume, Velocity, and Variety), which gets discussed by Arkady Zaslavsky et al in [4].

In order to address this complex issue and process IoT data effectively, enhanced data processing techniques are needed. How to identify and address the need for meaningful inference from raw data is the pertaining objective of the current work Semantic Filtering of IoT Data.

Semantics is nothing but a meaningful inference made from a stream of raw data. Semantic technologies have been used in the recent years as one of the key solutions to provide formalized representations of the real world data (Sensor Data). The advantage of applying semantic technologies to sensor data is actually the conceptualization and abstract representation of the raw data and making them machine interpretable. And interlinking this semantic data with existing resources on the Web will further enhance the usability of this raw data.

Adding semantics to different levels of IoT ensures that data originating from different sources is unambiguously accessible and process-able across different domains and users. Observation or measurement data collected from the real world can be semantically described to facilitate automated processing and integration in relation to domain knowledge and other existing resources in the cyber world.

Applying semantic technologies to IoT promotes interoperability among various resources, data providers, consumers and facilitates effective data access and integration, resource discovery, semantic reasoning, and knowledge extraction. According to P. Barnaghi, W. Wang, C. Henson, K. Taylor [5], Data annotations and semantic descriptions can be used at different levels and semantic annotations can be applied to various resources in IoT.

Xiang Su, Jukka Riekki, Jukka K. Nurminen, Johanna Nieminen and Markus Koskimies [6] illustrate RDF as best suited to represent physical world data in their work. By adding Semantics to Internet of Things, They have analyzed different semantic web technologies like RDF/XML, SenML/XML, and EN. Their objective was also to study the energy consumption by IoT Nodes (That had to be minimal) by adding Semantics to the IoT Data formats.

As per Barnaghi et al [7], IoT Data has big data challenges to tackle with. So issues related to interoperability, automation, and data analytics will require common description and data representation frameworks and machine-readable and machine-interpretable data descriptions.

Several technologies are developed in the Semantic Web, such as ontologies, semantic annotation, Linked Data and semantic Web services. Topical ontologies represent the knowledge from a certain domain providing a basic understanding of the concepts that serve as building blocks for further processing. Ganz, Barnaghi and Carrez [8] provide a solution for constructing the structure and relations of ontologies based on Real World Data.

In the Current work titled **Semantic Filtering of IoT Data**, the main objective is to derive meaningful annotations from large amounts of sensory data (real world data) in a human and or machine interpretable format. And abnormal events can be easily identified from the data. Essentially, Semantic Filtering converts raw data into useful information by Abstracting only the needed data and adding appropriate annotations/descriptions to it. And also it detects abnormal events based on the anomalies from data pattern / abstraction derived.

SEMANTIC FILTERING OF IOT DATA

Semantic Filtering of IoT Data has Semantic Annotation and Event Detection as the methodical outcomes.

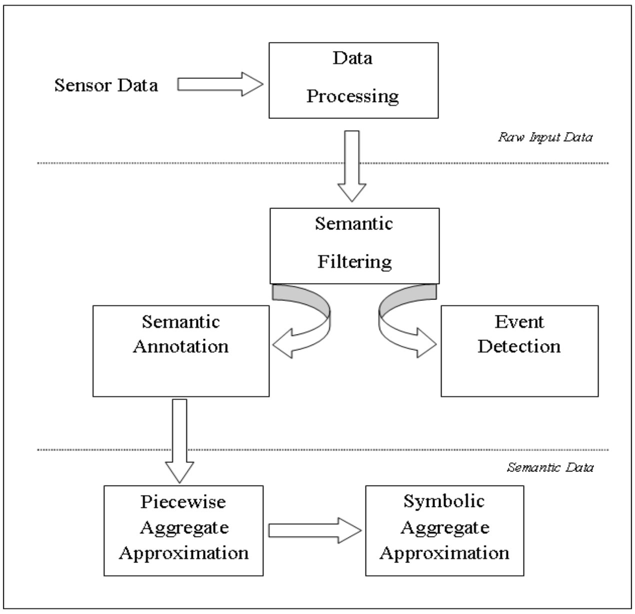

Figure 1: Semantic Filtering – Functional Parts.

And the functional structure (as also depicted in Figure 1 Semantic Filtering – Functional Parts.) of Semantic Filtering is:

1. Data Processing normalizes the raw data.
2. Semantic Filtering converts the normalized data into symbolic data
3. Semantic Annotation is given to the symbolic data based on range using
   a. Piecewise Aggregate Approximation (PAA).
   b. Symbolic Aggregate Approximation (SAX).
4. Event Detection detects any abnormal data pattern from the sensor data.

A. Semantic Annotation

In Semantic Annotation, first the raw heterogenic, time-tied IoT Data from Temperature, Pressure, and Humidity Sensors were normalized. As the second step, the normalized data is dimensionally reduced using PAA and SAX [9] by Jessica Lin, Eamonn Keogh, Stefano Lonardi and Bill Chiu. Here the raw data was divided into N sub-sequences using Piecewise Aggregate Approximation. Symbolic Aggregation was performed on the sub-sequenced data. Both PAA and SAX is further explained in the following sub sections.

1. Piecewise Aggregate Approximation (PAA)

PAA is Approximating a data curve by dividing it into equal length segments or subsequences and recording the mean value of the data points that fall within the segment. This representation reduces the data from n dimensions to N dimensions by dividing the data curve into N equal-sized 'frames'. The mean value of the data falling within a frame is calculated, and a vector of these values becomes the data reduced representation as per Figure 2 IoT – Piecewise Aggregate Approximation. (When N = n, the transformed representation is identical to the original representation. When N = 1, the transformed representation is simply the mean of the original sequence.)

So PAA can be summarized in the following steps:
Divide the IoT Data curve represented into w equal subspaces (where actual Data size is n).
While Data Size <= n
Find the mean for each space.
Plot the linear representation of each subspace.
So the entire data space (dimensionality) is reduced.

Discretization of these subspaces is done using the predetermined breakpoints (Breakpoints are sorted list of numbers that fall under the Gaussian curve).
For the Sensor data curve, a piecewise aggregate approximation will appear in the following way:

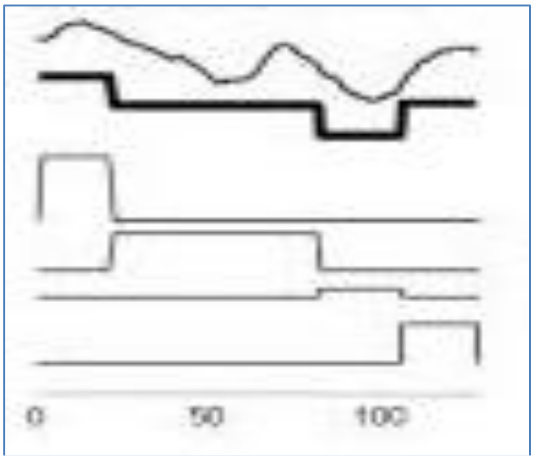

Figure 2 : IoT – Piecewise Aggregate Approximation

Finally the mean values are mapped to a single coefficient for each of the frames and aggregated. This forms the Piecewise Aggregated Approximation of the original data. Now the dimensionality of the original data is reduced. So this is the lower bounded and averaged representation of the (huge) actual data.

2. Symbolic Aggregate Approximation (SAX)
Basic idea of SAX is to convert the continuous data into discrete format. To convert the PAA coefficients obtained to symbols, breakpoints that divide the distribution

space into α equiprobable regions are determined (where α is the alphabet size specified by the user). Probabilistic approach is also applied in determining the breakpoints. Once the breakpoints are determined, each region is assigned a symbol (For example a, b, c are used as symbols here). The PAA coefficients can then be easily mapped to the symbols corresponding to the regions in which they reside. The symbols are assigned in a bottom-up fashion, i.e. the PAA coefficient that falls in the lowest region is converted to "a", in the one above to "b", and so forth. Figure SAX symbolization depicts the above theory.

To abstract from numerical values and to create higher-level concepts from the large amount of data produced by sensor devices, the symbolic aggregate approximation (SAX) dimensionality reduction mechanism [10] is used. SAX discretises the data and generates symbolic words representing patterns from the sensor data. Data discretization serves as building block for many pattern and event detection algorithms.

In the current work, the symbolic approximate aggregation is used to transform time tied IoT data into symbolized words as shown in Figure 3. IoT - SAX

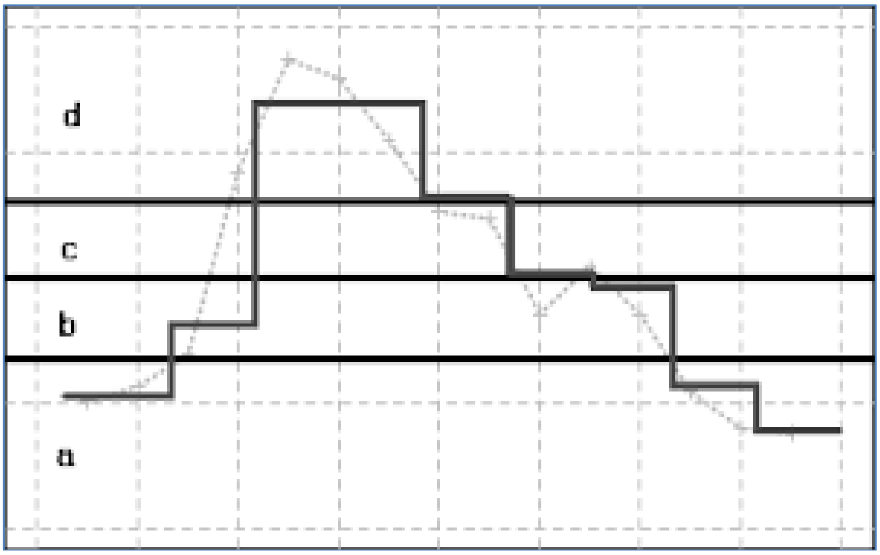

Figure 3: IoT – SAX

Finally with the Symbolized Data ("***abdccba***" from the Figure 3 IoT - SAX), Semantic annotations are provided based on the type of the sensor used.

B. Event Detection

IoT sensor data is continuous and is basically like a periodic update of the real world information sensed by IoT devices. It would also be highly beneficial to look for a pattern or an abnormal event in this data to make sense of this data in a real world scenario. In IoT Smart Home environment, data from the home sensors can be analyzed for regular patterns based on the occupant's daily usage. Based on this, when an abnormal event / data pattern is discovered, the appliances can be controlled accordingly. For example, if there was no motion in the room, but the A/C was left on, the temperature and humidity would drop abnormally. This can be detected and automated alert can be raised for controlling the appliances even from remote locations. This is essentially a sample anomaly detection that can be very useful in IoT.

Abnormal events in a data pattern can be identified by using anomaly detection. Anomaly is an abnormal/improbable data, given a data collection or dataset (Belonging to a specific domain). For example for Temperature value during night to have a data value as 60 (When the average would be normally 25 to 30) is an anomalous data. The following Figure 4 IoT Sensor Data – Anomaly Detection is a sample anomaly detected data plot for a Temperature Sensor data.

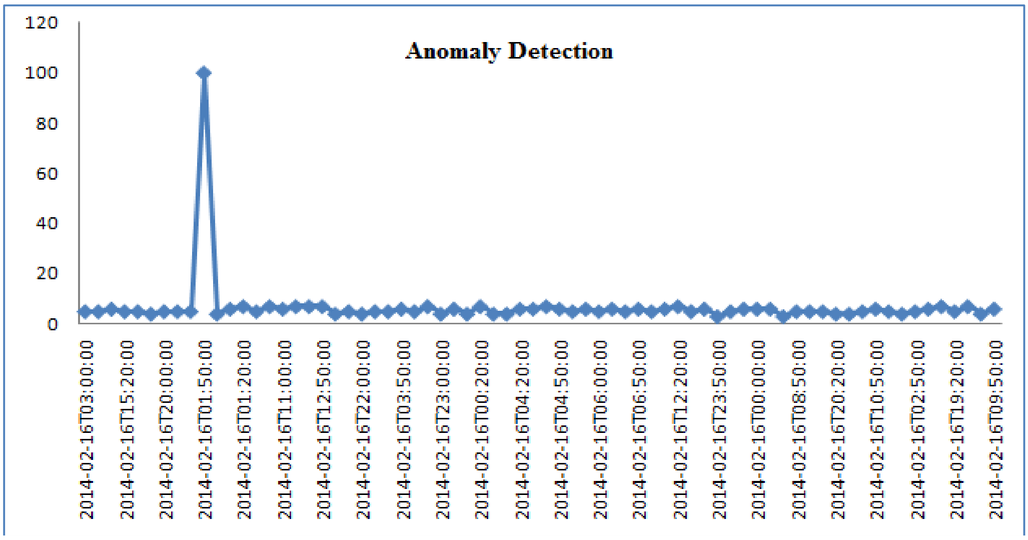

Figure 4: IoT Sensor Data – Anomaly Detection

To detect an anomaly / event from the IoT Temperature Sensor data, the subsequences generated by PAA are taken. With their mean value at hand, the individual data points within the frame can be compared against a threshold to detect any anomalous data. This was repeated for all the subsequences / Frames generated by PAA.

RESULTS AND DISCUSSION

Semantic Filtering of IoT Data was verified for the Smart City Domain of IoT. The real world data collected from Temperature, Humidity and atmospheric Pressure sensors were used for evaluation of the current work. These sensors deployed were for environmental monitoring in the city of Aarhus, Denmark during the month of February 2014. Semantic Annotations generated were based on the Threshold that was set based on the city’s normal weather conditions.

Using SAX, the Temperature Sensor Data was converted to the symbolic string "**bcdbbccbbc**". Thus by converting the original raw data, (Refer to Figure 5. Temperature Sensor Data)

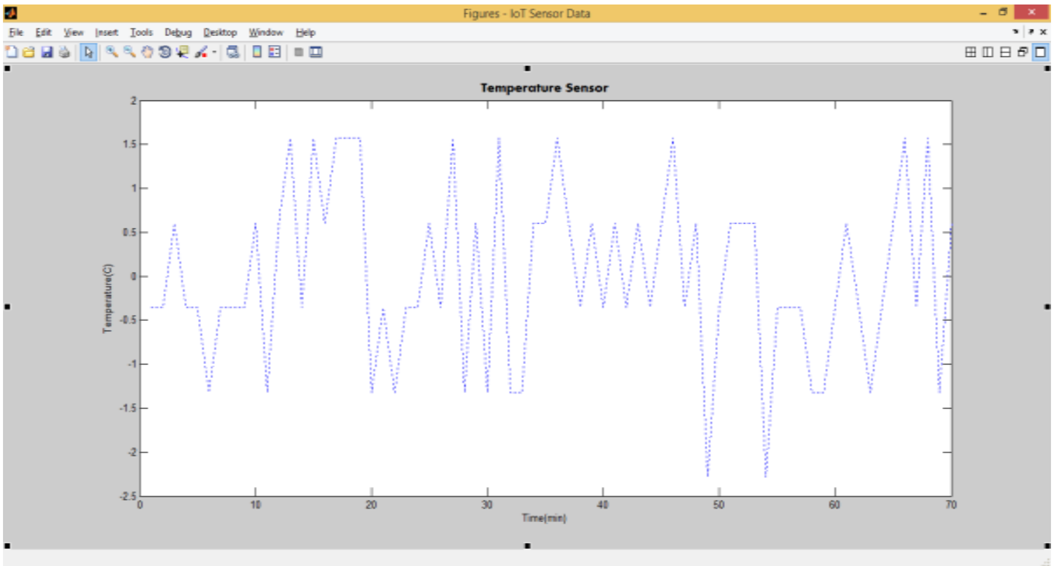

Figure 5 : IoT – Temperature Sensor Data

To symbolic data (Refer to Figure 6 Symbolic Temperature Data), obviously the Data volume is reduced.

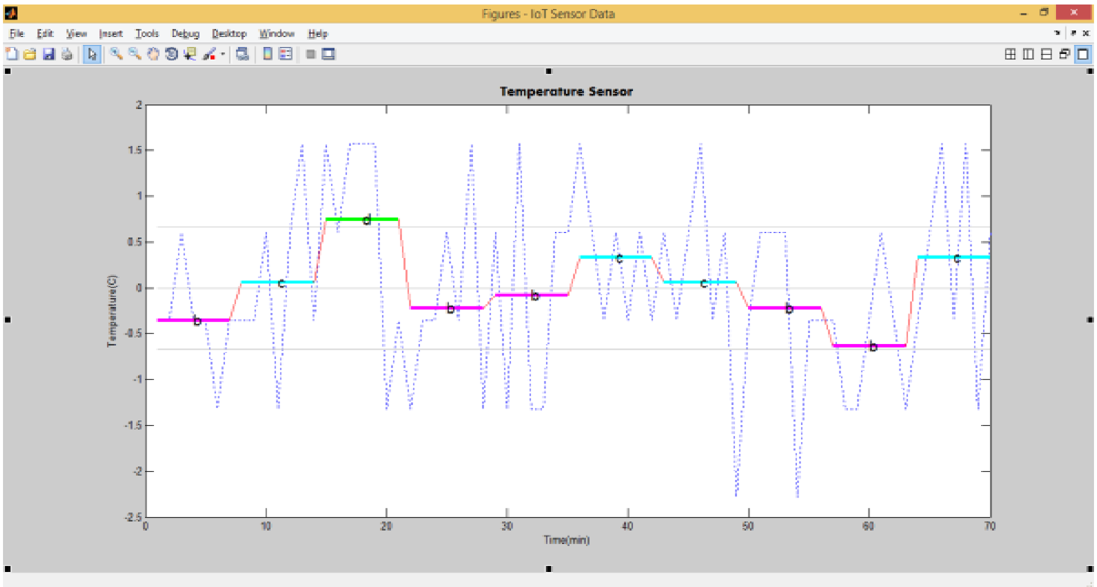

Figure 6: Symbolic Temperature Data

Now finally semantically annotating it (Refer to Table 1 Semantic Annotation for Temperature) makes the data more interpretable by human / machines easily.

Table 1: Semantic Annotation for Temperature

Symbol	Inference / Annotation	Range
a	Below Normal	< 5
b	Normal	>5 to <10
c	High	>10
d	Very High	>15

Event detection for abnormal data instances was also captured for Temperature Sensor by introducing anomalistic data value. Applying this for IoT HealthCare domain, anomaly / event detection from the sensor data (immediately) could be life saving. Current work was carried out with Environment sensors like Temperature, Air Pressure and Humidity is taken. This event detection could be expanded easily for other critical areas of IoT.

CONCLUSION

In this work semantic filtering of IoT data, semantic annotations and abnormal event detection using symbolic representation from raw sensor data were arrived at. In the process, Data dimensionality was reduced. The proposed semantic annotation would be a machine-readable and machine-interpretable representation of the raw data. The annotations could be used in control and monitoring applications that use the sensory data to observe the status of an environment or a physical entity or it could be used to provide an overall view of the changes and related occurrences over a period. Evaluation of Semantic Filtering is conducted with the help of a prototypical implementation and real world data gathered from sensors deployed.

However, the current work can be extended to ontological construction and also to explore the possibility of linking to the existing ontology. Execution time does not reflect the requirements for real-time processing; therefore more investigation has to be made, to make this approach more suitable for learning and ontology construction from real-time streaming data. Also further investigations on finding the best parameters for the whole process, in terms of window lengths (for the dimensionality reduction process), maximum number of symbols (that can be used in symbolic representation) and the threshold can be set (for semantic annotations) would give more precision to the results obtained.

REFERENCES

[1] McEwen. A and Casimally. H, "Designing the Internet of Things", ISBN - 13:978-1118430620, Wiley 2013.

[2] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, Marimuthu Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions", September 2013.

[3] William Stallings, "The Internet of Things: Network and Security Architecture", December 2015.

[4] Arkady Zaslavsky, Charith Perera and Dimitrios Georgakopoulos, "Sensing as a Service and Big Data", 2013.

[5] P. Barnaghi, W. Wang, C. Henson, K. Taylor, "Semantics for the Internet of Things: early progress and back to the future", International Journal on Semantic Web and Information Systems, IGI Global, September 2012.

[6] Xiang Su, Jukka Riekki, Jukka K. Nurminen, Johanna Nieminen and Markus Koskimies, "Adding semantics to internet of things", 2010.

[7] Payam Barnaghi, Senior Member, IEEE, and Francois Carrez, Wei Wang, Suparna De, Gilbert Cassar, Klaus Moessner, "Information Abstraction for Heterogeneous Real World Internet Data" , IEEE Sensors Journal, vol. 13, no. 10, pp. 3793-3805, 2013.

[8] Frieder Ganz, Member, IEEE, Payam Barnaghi, Senior Member, IEEE, and Francois Carrez, "Automated Semantic Knowledge Acquisition from Sensor Data", IEEE Systems Journal 2014.

[9] Jessica Lin, Eamonn Keogh, Stefano Lonardi, Bill Chiu, "A Symbolic Representation of Time Series, with Implications for Streaming Algorithms", June 2003.

[10] XiChen, Huajun Chen, Ningyu Zhang, Jue Huang and Wen Zhang, "Large-Scale Real-Time Semantic Processing Framework for Internet of Things", September 2015.