Abstract— This paper describes a framework for perception creation from sensor data. We propose using data abstraction techniques, in particular Symbolic Aggregate Approximation (SAX), to analyse and create patterns from sensor data. The created patterns are then linked to semantic descriptions that define thematic, spatial and temporal features, providing highly granular abstract representation of the raw sensor data. This helps to reduce the size of the data that needs to be communicated from the sensor nodes to the gateways or high-level processing components. We then discuss a method that uses abstract patterns created by SAX method and occurrences of different observations in a knowledge-based model to create perceptions from sensor data.

I. INTRODUCTION

It is estimated that the number of internet connected devices will grow to 50 billion devices by 2020 [1]. Many of these devices will include sensors that can send observation and measurement data about the physical world. The amount of data generated by billions of sensors (machine sensing) and humans carrying devices (citizen/social sensing) has grown from 281 exabytes in 2007 to 1,800 exabytes in 2011 [2]. This combination of cyber-physical data can help us to better understand events and changes in our surrounding environments, leading to applications with the ability to monitor and control buildings, homes and city infrastructures, and provide better healthcare and elderly care services. Efficient use of the sensor data involves making sense of massive amounts of data in order to convert it into information and knowledge from which humans can gain insights and base decisions. In this work we discuss how raw sensor observations can be processed and analysed to create perceptions.

Perception is the primary basis of human intelligence and experience. Creating higher-level abstractions and generating machine perception from the machine and citizen sensing data is a key enabler for developing situation-aware applications that can intelligently respond to changes in the real world [3]. In sensor data, perceptual hypotheses represent the semantics, or meaning, of observations and measurements.

The deluge of data generated by machine and human sensors needs to be processed to create meaningful higher-level abstractions that are machine-interpretable or human-understandable and to make this raw data useful for decision making. Our work represents an efficient solution for multi-granularity representation of continuous observations made by sensors, creating abstractions from the raw sensor data by representing them as patterns, and semantically annotating these patterns based on thematic and spatiotemporal features to create higher-level perceptions. The sensor connectivity and annotation of the data is enabled by our previous work on developing an intelligent gateway for semantic sensor networks [4]. We discuss the process of creating machine-interpretable and human-understandable perceptions from the data. Our approach to systematically generating perceptions from observation data is grounded in an abductive logic framework called Parsimonious Covering Theory (PCT) [5]. PCT uses domain-specific background knowledge to determine the best explanation for a set of observations. We describe how to approximate PCT and derive abstractions from observations on the Web using background knowledge and the inference capabilities of an off-the-shelf reasoner. In summary, our approach translates low-level sensor data to high-level abstractions through three steps:

1) Collect and semantically annotate sensor data
2) Find patterns in the sensor data with SAX (Section 3)
3) Translate the SAX patterns into abstractions with PCT (Section 4).

The rest of this paper is organised as follows: Section II describes the concept of the Internet of Things and integration of the physical world data into the digital world. Section III discusses the pattern generation technique for sensor observation and measurement data. Section IV describes the proposed framework for perception creation using the Parsimonious Covering Theory. Section V concludes the paper and describes future work.

II. INTERNET OF THINGS AND REAL WORLD DATA

Extending the current Internet and providing connection and communication between devices and physical objects, or “Things”, is a growing trend that is often referred to as the Internet of Things (IoT) [1]. IoT enables sensors, devices and human users to report their observations and measurements.
from the physical environment to other devices and/or parties via the Internet. In recent years, the primary focus has been on developing infrastructure and enabling technologies for IoT. However, intelligent data processing mechanisms that exploit the IoT data, process it and integrate it into the existing business processes and/or create situation-awareness and actionable knowledge are equally important in utilising IoT data and applications. The observation and measurement data is the lower-level data that is captured by sensor devices or human users. This data needs to be made available to potential users via common interfaces or to be stored in repositories for future access. However, IoT data consumers, users, and applications are often interested in the higher-level concepts, such as events. To make sense of the sensor data, extraction processes are required to transform raw sensor data into higher-level knowledge that refers to an event, a pattern or concepts that are comprehensible to machines and human users. This can be accomplished by using reasoning mechanisms. The creation of high-level abstractions from real-world data can be used to create situation awareness and intelligent decision making. To collect real world data from heterogeneous sensor resources, we use a gateway component [4] that runs on the University of Surrey’s smart campus testbed [6]. The gateway allows the accessing and collecting of sensor data using the Web service interfaces and hides the complexity of the underlying devices and sensor nodes from application level users. In our gateway component, we also annotate the sensor data according to the W3C Semantic Sensor Network (SSN) ontology [7][8]. The semantic annotations define the type of data, as well as the spatiotemporal attributes of the data and observation resources. These annotations are used in the perception computation process when different types of observations are fed into the PCT model to create meaningful abstractions from the data (described in section IV).

III. RAW SENSOR DATA AND OBSERVATIONS

In order to create meaningful abstractions from sensor data we first transform the raw sensor data into a set of observations. For example, if a sensor device measures the temperature of a room, the sensor reading can be represented as a digital value with a unit of measurement e.g. 36 degree Celsius. The first step is to transform this raw data into an observation; e.g. “hot” or “high temperature.” The raw sensor data, in combination with other information, can be then used to extract meaningful knowledge: e.g. high temperature in a room in the winter could indicate that the heater is set to high, or in the summer it could indicate that the air conditioning does not work; or high temperature in combination with smoke detection via another sensor could indicate a fire.

In this paper we create lower-level abstractions (here we call them observations) by collecting and processing streaming sensor data. The streaming data is divided into different segments and a pattern is created for each segment. These patterns represent an aggregation of a set of raw sensor data during a period of time. The patterns are then compared with previously identified and tagged patterns to create observations. The pattern construction is performed using the Symbolic Aggregate Approximation (SAX) technique [9]. SAX is used in data mining and time series data for dimensionality reduction and creating symbolic patterns. SAX divides a time series data into equal segments and then creates a string representation for each segment. For example, Figure 1 shows the data captured via an ambient noise sensor and illustrates the SAX patterns created from this data. The green line (the upper line) represents the digitised noise sensor data collected by the sensor and the blue line (the lower line) shows the patterns created using the SAX technique. The SAX symbolic patterns are represented as string “words.” For example, in Figure 1 the first symbolic pattern is represented as “fggfffhffffyjghbff” in SAX, consisting of 20 characters from a vocabulary of 10 letters (a-j) (the string size and letters are adjustable in SAX; for more information refer to [10], and [11]). The SAX patterns create the lower-level abstractions that are used to create the initial observations for computing the perceptions. The next section describes how the lower-level abstractions are represented as observations and discusses how the observations are then used in the perception creation process.

IV. PERCEPTION COMPUTATION

The initial step in analysing the sensor data to compute perceptions is to transform the data into a set of observations (as described in Section III). For this purpose we use symbolic patterns created using SAX. The next step is to use knowledge extraction processes to transform the observations into higher-level abstractions that can be directly, or by using reasoning mechanisms, related to an event or a phenomenon. This involves identifying the patterns, annotating them and then using the reasoning method to relate the lower-level observations to different events or situations [16]. Using the observations created in the previous step, along with domain specific background knowledge, we then infer perceptions from the sensor data using an OWL (Web Ontology Language) reasoner. Figure 2 illustrates how such background knowledge is represented in the W3C Semantic...
Sensor Network (SSN) ontology. In particular, it shows the relation between an observable property (e.g., cold temperature) to a real-world event (e.g., open window). The sso and dul namespaces in Figure 2 refer to the W3C SSN and DOLCE [17] ontologies, respectively. We propose using the Parsimonious Covering Theory (PCT) [13] to create perceptions from the observations. PCT uses domain-specific background knowledge, in the form of a bipartite graph, and determines the best explanation for a set of observations that are created using the SAX technique and pattern identification. It is becoming more common to find the background knowledge required for PCT openly available on the Web (as Linked Data [12]). To take advantage of this trend, Henson et al. have shown in an earlier work how to approximate PCT via the Web Ontology Language (OWL) [14] and they also described how to derive abstractions from observations on the Web using background knowledge and inference techniques [15].

![Figure 2. Representation of background domain knowledge in the W3C Semantic Sensor Network Ontology, utilised for creating perceptions of sensor data](image)

In this work, we use the annotated SAX patterns as observations for the PCT method. The annotated patterns are created by comparing the created symbolic patterns with similar existing annotated patterns. The string representation of the SAX mechanism enables us to compare the patterns using a string similarity function. However, this limits the pattern identification and annotation process to a set of patterns that are annotated and predefined to the system. The PCT method then uses background knowledge to define the reasoning models in order to process the observations (i.e. SAX patterns) and draw conclusions for possible perceptions that can be computed from the data.

The observation creation mechanism using the SAX mechanism enables the automated construction of lower-level abstractions (i.e. annotated SAX patterns), and higher-level abstractions (i.e. perceptions) that can then be computed using the PCT method. Figure 3 shows the different components of the perception creation process in our proposed framework. The patterns in SAX are created by analysing raw sensor data as time series data; so that it does not require any prior threshold definition or heuristics to set the parameters for pattern creation. However, the current model is still restricted to the patterns that can be associated to a set of predefined observations and we still require predefined and existing models that can be used in the PCT method.

As shown in Figure 3, first the data is collected via the sensors and then the raw sensor data streams are represented using SAX patterns. The pattern representation in SAX is the form of string. The symbolic patterns are then matched to a set of predefined observations according to the semantic annotation of the raw sensor data and the existing annotated patterns for different types of data. The annotated patterns form the lower-level abstractions (or what is referred to as the observations). The observations are then used as input for the PCT method. The PCT method uses the observations and background knowledge to infer possible perceptions from the data. An abductive reasoning mechanism, as discussed in [15], is used to infer the perceptions. The framework allows the sensor data to be represented at different granularity levels (i.e. raw data stream, symbolic patterns, and observations) and enables computing perceptions from the data.

![Figure 3. Perception creation using lower-level abstractions (annotated SAX patterns) and the PCT method](image)

V. CONCLUSIONS AND FUTURE WORK

This paper describes a conceptual framework for computing perception from sensor data. This involves creating lower-level abstractions from streaming raw sensor data using Symbolic Aggregate Approximation (SAX) and applying a method based on Parsimonious Covering Theory (PCT) to create higher-level abstractions that can represent machine-interpretable and/or human-understandable knowledge. We describe our framework using streaming sensor data and explain how different perceptions can be computed using the proposed method. The paper reports ongoing work and the results are still at an early stage. The proposed solution is still limited to already known observations and pre-defined inference models in PCT. We also need to use a set of predefined concepts to annotate the lower-level SAX
abstractions. The latter limits the extensibility of the method and hinders its application to a broader range of heterogeneous sensor data. Future work will focus on automating the annotation process for the SAX patterns and creating mechanisms to compute a broader range of higher-level abstractions using the SAX patterns and the PCT method.

REFERENCES