4.5 Industry: Supply Chain Management, Smart Farming, Chemical Process

Vargheese *et al.* [196] propose a system that improves shoppers' experience by enhancing the On the Shelf Availability (OSA) of products. Furthermore, the system also looks to forecast demand and provide insights on buyers' behaviour. A multi-tiered approach is employed, where sensors like video cameras, process video streams locally and analyse the products on the shelf, this data is then verified by other sensors like light, infra-red and RFID sensors and the metadata produced is sent to the the cloud to be further processed. In the cloud, this real time data is combined with models from learning systems, data from enterprise Point of Sale (POS) systems and inventory systems to recommend action plans to maintain the OSA of products. The staff of the store are informed and action is taken to restock products. Weather data, local events and promotion details are then analysed with the current OSA to provide demand forecasting and to model buyers behaviour which is fed back into the system.

Nechifor *et al.* [140] describe the use of real time data in analytics in a cold chain monitoring [2] process. Trucks are used for transporting perishable goods and drugs that require particular thermal and humidity conditions, sensors measure the position and conditions in the truck and of each package, while actuators - air conditioning and ventilation can be controlled automatically. On a larger scale, predictions can be made on delays in routes and when necessary to satisfy the product condition needs, longer but faster routes (less congestion) might be selected.

Similarly, Verdouw *et al.* [198] and Robak *et al.* [172] examine supply chains - the integrated, physical flow from raw material to end products with a shared objective, and formulate a framework based on their virtualisation in the IoT. At its highest level, a virtual supply chain supports intelligent analysis and reporting. This is applied to floricultural and a Fourth Party Logistics (4PL) integrator respectively, where business intelligence, data mining and predictive analytics can provide early warning in case of disruptions or unexpected deviations and advanced forecasting about consequences of the detected changes when the product reaches destination.

In the above examples on product and supply chain management, we see a common theme of predictive analytics being employed to business processes. This predictive analytics is often powered by learning from data to discover models or through data mining for patterns in data. The effectiveness of these algorithms benefits from the big data of the IoT in providing a large observation space for discovering patterns and trends. Real time data from sensors then provide the information required to immediately control actuators to rectify problems like products being out of stock on the shelf or conditions in trucks being unsuitable for perishable food.

Kamilaris *et al.* [104] describe the use of a Complex Event Processing (CEP) engine to discover significant events on semantically-enriched data streams from sensors within two smart farming scenarios. One scenario included detecting the fertility of cows from temperature readings and other information on a dairy farm to suggest the best insemination timings. The other was to adaptively control the soil conditions for crop cultivation.

The chemical process industry deploys inferential industrial IoT sensors to process monitoring chains [42]. Some techniques applied by sensors include linear regression, artificial neural networks (ANN) and Gaussian process regression which predict variables using available process data. These predictions enable quality monitoring and advance control systems in plants to automatically react and prescribe process modifications "to prevent off-grade products".

5 TYPES OF ANALYTICS AND THEIR IMPORTANCE

Following the study of the current work in analytics in the IoT, we explore a classification of analytics that is applicable to these domains. We derive a categorisation of analytical capabilities from business analytics literature, which the term analytics comes from. Bertolucci *et al.* [26]

Table 3. Summary of Application References by their Domains and Analytical Capabilities

Domain \ Capability	Descriptive	Diagnostic	Discovery	Predictive	Prescriptive
Health	[37, 136]	[37, 117]	[20, 86, 136]	[87]	[37]
Transport	[124]		[47, 99, 155]	[82]	[115]
Living	[41, 167]	[157]	[41, 66, 75, 138]		
Environment	[210]		[64, 178]	[4, 9, 135]	
Industry			[172, 198]	[42, 140, 198]	[42, 104, 196]

propose descriptive, predictive and prescriptive categories while Gartner [105] [33] proposes the extra category of diagnostic analytics. Finally, Corcoran *et al.* [45] introduce the additional category of discovery analytics. We build upon these to form a comprehensive classification of analytic capabilities consisting of five categories: descriptive, diagnostic, discovery, predictive and prescriptive analytics. Each category is described in detail in Section 5.1 and we also summarise how each IoT application surveyed in the previous section is categorised in Table 3. Each application domain has applications which support multiple analytical capabilities. We also note that all the categories of capabilities are well-represented in the literature survey, while mature domains like the industrial IoT focus on high value analytics.

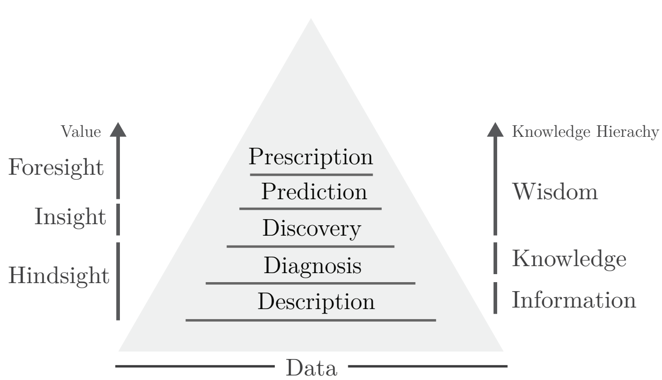

Fig. 4. Analytics and the Knowledge and Value Hierachies

Fig. 4 looks at how each analytical capability fits within the Knowledge Hierarchy [24] which is a common framework used in the Knowledge Management domain. This categorisation of analytic capabilities enables us to establish what the aim of analysis is and allows us to relate to the vision of IoT deployment as often expressed in research roadmaps. The value of each capability, is also highlighted in the figure. The knowledge hierarchy starts with data at the base, examples of which are facts, figures and observations (e.g. the raw data produced by IoT 'things'). Information is interpreted data with context, for example, temperature as represented by descriptive analytics: an average over a month or a categorical description of the day being sunny and warm. Knowledge is information within a context with added understanding and meaning, perhaps possible reasons for the high average temperature this month. Finally, wisdom is knowledge with insight, for example, discovering a particular trend in temperature and projecting it across future months while providing cost saving energy management solutions for heating a smart home based on these predictions. Each component of the knowledge hierarchy builds on the previous tier and we can see something similar with analytical capabilities. To add a practical view from business management literature to our discussion, a review of organisations adopting analytics [112] categorised them as Aspirational, Experienced and Transformed. Aspirational organisations were seen to use analytics in hindsight as a justification for actions, utilising the data, information and knowledge tiers in the process. Experienced organisations utilised insights to guide decisions and transformed organisations were characterised by their ability to use analytics to prescribe their actions, effectively applying foresight in their decision making process.

5.1 Five Categories of Analytics Capabilities

5.1.1 Descriptive Analytics. It helps us to answer the question, "what happened?". It can take the form of describing, summarising or presenting raw IoT data that has been gathered. Data are decoded, interpreted in context, fused and then presented so that it can be understood and might take the form of a chart, a report, statistics or some aggregation of information.

5.1.2 Diagnostic Analytics. It is the process of understanding why something has happened. This goes one step deeper then descriptive analytics in that we try to find out the root cause and explanations for the IoT data. Both descriptive and diagnostic analytics give us hindsight on what and why things have happened.

5.1.3 Discovery in Analytics. Through the application of inference, reasoning or detecting non trivial information from raw IoT data, we have the capability of Discovery in Analytics. Given the acute problem of volume that big data presents, Discovery in Analytics is also very valuable in narrowing down the search space of analytics applications. Discovery in Analytics on data tries to answer the question of what happened that we don't know about and the outcome is insight into what happened. What differentiates this from the previous types of analytics is using the data to detect something new, novel or different (e.g. trends, exceptions or clusters) rather than describing or explaining it.

5.1.4 Predictive Analytics. For the final two categories of analytics, we move from hindsight and insight to foresight. Predictive Analytics tries to answer the question: "what is likely to happen?". It uses past data and knowledge to predict future outcomes [76] and provides methods to assess the quality of these predictions [184].

5.1.5 Prescriptive Analytics. It looks at the question of what should I do about what has happened or is likely to happen. It enables decision-makers to not only look into the future about opportunities (and issues) that are potentially out there, but it also presents the best course of action to act on foresight in a timely manner [22] with the consideration of uncertainty. This form of analytical capability is closely coupled with optimisation, answering 'what if' questions so as to evaluate and present the best solution.

5.2 Specific Types of Analytics

Having looked at analytical capabilities which help to define the aims of analytics, we look at specific analytics that can guide stakeholders involved in the deployment of analytics on IoT applications. A summary of the specific types of analytics and their corresponding analytical capabilities can be found in Fig. 5.

5.2.1 Visual Analytics. Visual analytics combines interactive visualisations with data analytics techniques "for an effective understanding, reasoning and decision making on the basis of very

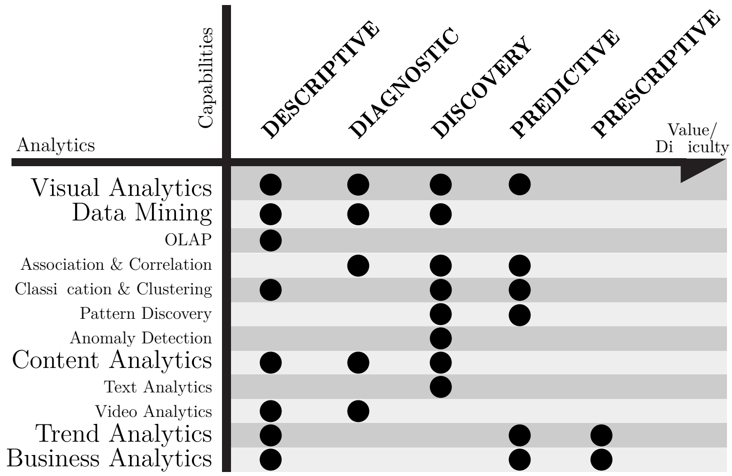

Fig. 5. Classification of Types of Analytics

large and complex data sets" [106]. Hence, visual analytics can contribute to not only describing and diagnosing what happened but also help users to discover new insights. In the work by Zhang et. al [?], we see visual analytics being applied to health care data and describing, through answering of questions like "What is the distribution of pregnancy age?", diagnosing, through hypothesising two disease patterns due to "diarrhoea" and "fever" not being correlated and discovery, through detecting the delayed outbreak of two diseases.

5.2.2 Data Mining. Data Mining is part of the Knowledge Discovery from Data (KDD) process in which interesting patterns and knowledge are discovered from large amounts of data [79]. The IoT is a source for a large amount of data in which the techniques of data mining can be applied. These include:

Multi-dimensional data summary is often associated with Online analytical processing (OLAP) operations that make use of background knowledge of the domain to allow presentation of data at different levels of abstraction. For example, you could drill-down and roll-up data to present it at different degrees of summarisation.

Association & correlation is the process of finding the relationship between two variables which vary according to some pattern. This could allow us to find out whether buying product A, led to buying product B with a degree of confidence and support.

Classification is the process of finding some model or function that has the ability to distinguish between data classes or concepts.

Clustering is the process of grouping data objects into classes without labels. The clustered data objects have maximum similarity to in-class objects and minimum similarity between objects from other classes.

Pattern discovery is the process of detecting and extracting interesting patterns from data, an example of which are frequent item sets, a set of items that often appear together in a transactional data set. *Anomaly detection* refers to the problem of "finding patterns in data that do not conform to expected behaviour" [34].

5.2.3 Content and Text Analytics. Content Analytics is the broad area of which analytical techniques are applied to digital content. Text analytics is the derivation of high quality information from unstructured text, for example, extracting named entities and relations, analyse sentiment, extract events and time series information, etc.

5.2.4 Video Analytics. Video Analytics (VA) is about the use of specialised software and hardware "to analyse captured video and automatically identify specific objects, events, behaviour or attitudes in video footage in real-time" [66].

5.2.5 Trend Analytics. Trend analytics is concerned with looking at data and events across time, understanding it and making predictions to future trends and providing early warning systems. Trend analytics is also closely related to the analysis of time-series information [35], where looking at a time-series we try to find a 'long-term change in the mean level'.

5.2.6 Business Analytics. Business Analytics is the practice of using an organisations data to gain insights through analytical techniques that can better inform business decisions and automate and optimise business processes.

5.3 A Layered Taxonomy of Data, Analytics and Applications for the IoT

Fig. 6 shows a layered taxonomy of analytics for the IoT that summarises our survey with respect to analytics capabilities and specific analytics. There are three layers in the taxonomy: data, analytics and applications. Within each layer are various concepts, classes and techniques which are well-defined in background literature and gathered from reviews in each area.

In the analytics layer, visual analytics processes are defined by Keim *et al.* [107] while techniques for each data type are summarised in surveys [5, 134, 189]. Data mining [67, 114, 183], text analytics [3] and video analytics [117] each are well-described in the referenced authoritative texts. Time-series forecasting [123], analysis and control [29] have also been reviewed in detail. Literature also covers business analytics processes [111], prescriptive analytics [22] and techniques [193].

In the application layer, themes and domains are from Section 3.2 while the IoT applications from each domain surveyed in Section 4 are shown connected to their various analytics capabilities. Analytics techniques can then be referenced under each capability.

In the data layer, big data as defined in Section 1 is summarised along with terms used throughout the survey including currency, types of data and their sources. Two other terms for big data, veracity and variability are introduced for completeness. Veracity is concerned with the noise within data and how accurate the data is for whatever purpose it is to serve. Variability is concerned with data whose meaning changes due to differences in interpretation of data within a specific context. Finally, processes, distribution levels and distributed technologies for storage and compute are covered in Section 6 that follows this.

6 ENABLING INFRASTRUCTURE FOR IOT ANALYTICS

In the previous section we looked at classifying analytics and building a taxonomy for understanding analytics. In this section, we will review work that enables analytics to be applied on IoT data.

Enabling infrastructure for analytics on the IoT are components, techniques and technology that contribute to the process whereby data is utilised in analytics applications. Fig. 7 shows the process of how data goes through the steps of generation and collection, aggregation and integration and finally is applied in analytics applications [38]. Storage and compute are abstract processes involved with each step of this data flow. In practice, data could be pipelined from one step to another, hence, need not necessarily be stored, physically, in a separate location. Compute could also be done on the device or in transit and need not imply a separate compute component.

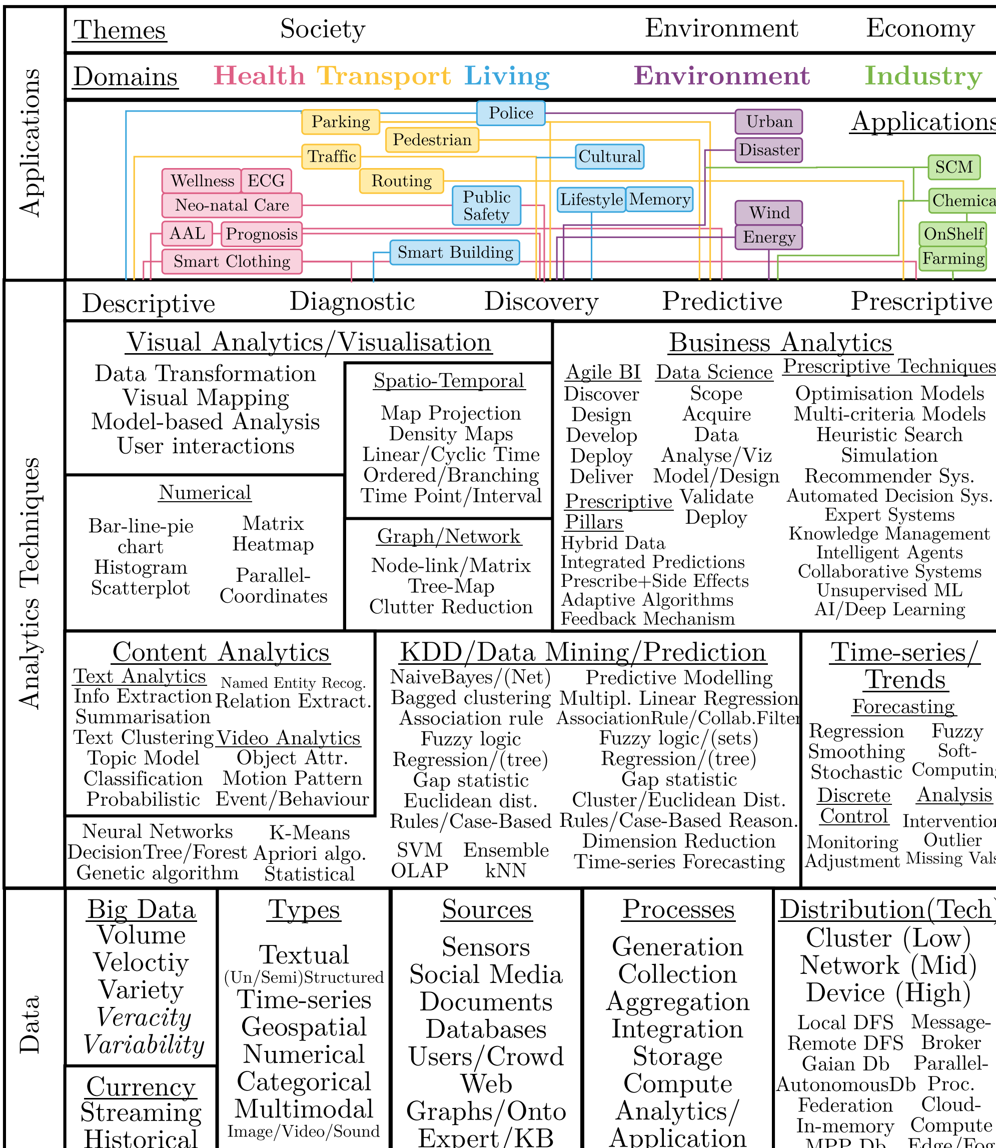

Fig. 6. Layered Taxonomy of Analytics From Data to Application

The following sections elaborate on each step of the data flow in IoT analytics from data generation and collection to aggregation and integration with storage and compute alongside. Fig. 8 summarises the technologies covered.

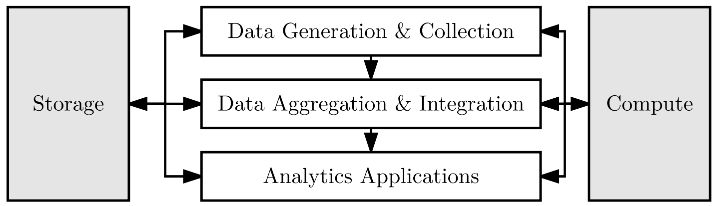

Fig. 7. Data Flow Process for Analytics Applications

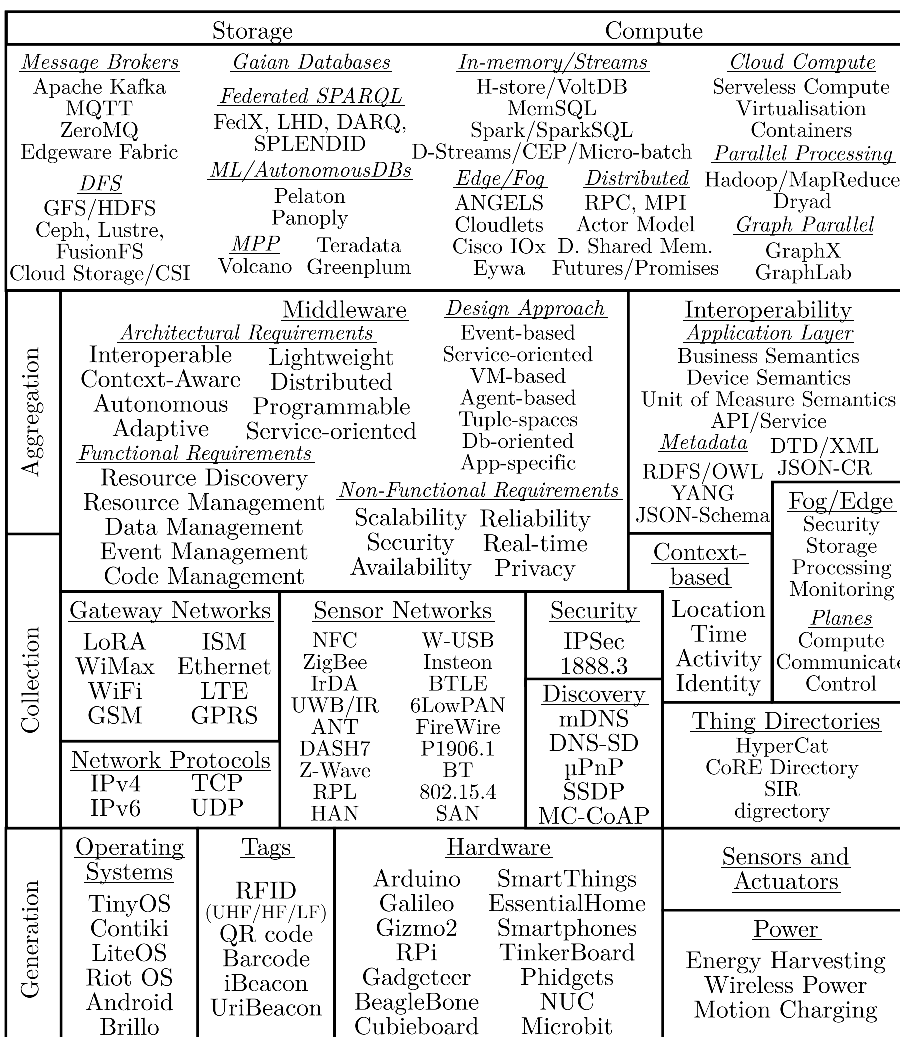

Fig. 8. IoT Enabling Infrastructure for Analytics: Generation, Collection, Aggregation, Storage, Compute