Adaptive Cognitive Management and Knowledge Discovery Framework Based on Internet of Things Big Data

Runxuan Bai¹ *, JianYu Lv², Chunnan Shang³

¹Faculty of Science and Engineering, University of Nottingham, Ningbo, Zhejiang, 315100
²School of information, University of international business and Economics
³Southwestern university of finance and economics

Abstract. In the future of large-scale industrial automation applications in the Internet of Things big data management and knowledge discovery, the importance of the industrial Internet is increasing day by day. The Internet of Things (IoT), computing intelligence, machine type communication, big data and sensor technology and other diversification Technology can be combined to improve the efficiency of data management and knowledge discovery for large-scale automated applications. To this end, we need to propose a cognitive-oriented IoT big data framework (COIB-framework) and its implementation architecture, IoT big data hierarchical architecture, data organization and knowledge exploration subsystems to achieve effective data management and knowledge discovery suitable for large-scale industrial automation applications. In this work, we need to propose a cognitive-oriented IoT big data framework (COIB-framework), as well as implementation architecture, IoT big data layered architecture, data organization and knowledge mining subsystems, to meet the needs of large-scale industrial automation applications. Discussion and analysis show that the proposed framework and architecture are based on the intelligent industrial application of IoT big data provides a reasonable solution.

1. Introduction

IoT objects can be defined as any tangible things associated with real-world entities (such as people, machines, and animals), with unique identification and self-directed data transmission capabilities on the network [1]. To technically construct IoT objects, you need embed three tiny components, such as sensor chips, solenoids, and adjusting capacitors, into a very small container that can be easily placed into any real-world entity related to IoT applications; however, the technical configuration actually depends on the type of entity and application [2]. More specifically, we can define IoT objects as sensors or RFID devices or any smart objects that have an Internet connection through a physical IP and are able to autonomously without any human intervention to transfer data to the network [3].

In almost all IoT applications, massive data will be dumped into storage, which is very necessary for data analysis, modeling, information conversion, knowledge production, and decision generation [4]. In the upcoming large-scale industrial automation, thousands of exabytes of data need to be stored [5]. For
example, if data is extracted every 5-15 minutes, IoT objects used in large-scale industrial power grid infrastructure will generate huge log data at a rate of 1tb per day [6].

Therefore, such IoT big data needs a suitable analysis framework to generate knowledge to measure the operating efficiency, load distribution, machine maintenance, etc. in industrial automation applications [7]. For the storage of large-scale IoT big data, digital technology has been greatly developed; however, the speed of data management and knowledge discovery may not be greatly improved and cannot be performed on the basis of the timeline [8]. The big data management and knowledge discovery of the Internet of Things are designed to provide a COIB framework for real-time data management, so knowledge mining researchers [9]. It is still a challenging problem. In order to achieve effective data management and knowledge discovery, the COIB framework always needs suitable data management and analysis tools to convert large-scale heterogeneous agile data streams into actionable knowledge [10]. The corresponding architecture is given in Figure1.

2. Coib-architecture organization

To implement the COIB-framework, we must ensure that the automation environment has a high degree of industrial Internet connection between the Internet of Things objects manufactured on different machines to regulate the function and operating efficiency of the machine. The functions of the COIB-framework organization are shown in Figure2. For large-scale parallel and distributed environments, each IoT network segment has independent application monitoring and control. Therefore, the entire industrial automation environment can be logically divided into several IoT segments to be compatible with standard network configuration management requirements. The IoT data segment is responsible for generating the IoT raw data stream and serving as the original data source for its corresponding data segment. These IoT big data streams are highly unstructured in terms of name, scale, abstraction level, etc., ensuring a high degree of data inconsistency performance, redundancy, incompleteness, and other anomalies. Therefore, big data aggregators perform data fusion operations on these large IoT big data streams. Data fusion operations can use some standard data semantics to eliminate these anomalies, resulting in Clean data for total quality management.

![Figure 1. Knowledge warehouse system architecture.](image-url)
The IoT big data classifier divides the cleaned data into multiple clusters based on data behavior, characteristics, data types, and data types, and is easy to use. In industrial automation applications, the data domain includes operational data, production data, status data, and maintenance data, etc. HBase storage is responsible for expanding these data clusters and storing them in its multiple storage nodes. In the HBase system, you can set up tables like a relational database so that each table contains rows and columns, and each table must have an element defined as the primary key. For better supervision and access, the primary node controls the storage node.

3. Functional analysis of COIB framework

3.1. IoT Big Data Fusion
It includes the fusion of large-scale data streams according to some standard data semantics. At each periodic time interval, the data aggregator checks the availability of data in the original data source of a single IoT, and if some availability is found, it continues to perform data integration process to achieve data integrity, consistency, integrity, and total quality management. Data quality assessment is an important aspect of total quality management and can further achieve the expected quality of big data analysis results. For real-time monitoring applications, once in if the process is executed at periodic intervals, the subsequent functional process must be performed to enable effective data management and knowledge discovery within the critical timeline.

3.2. IoT Big Data Classification
Real-time data classification and clustering for agile data streams is a challenging problem in the IoT big data environment. The data classifier initiates the classification process to formulate the fused data
into multiple data groups so that they exist in each data group homogeneous data. For automated applications, fused data can be classified into multiple groups with multiple event types, such as machine status data, functional data, inventory data, production data, product quality data, etc. Some of the non-linear data classification processes may be suitable for large-scale IoT big data classification problems. For time-critical big data classification and clustering, only simple "If Then else" classification statements can be used to aggregate data into multiple data clusters. To better illustrate the problem, let us design a data classifier based on five clusters as described in Algorithm1.

3.3. IoT Big Data Storage
At present, the technology has been installed in the large-scale data storage in the IoT big data environment. The HBase database management system is responsible for IoT big data storage across storage nodes led by control nodes. Based on the application, the entire IoT space can be divided into disjoint segments in order to implement the HBase system in each segment, thereby ensuring distributed control of large-scale IoT large databases. The control node of one HBase system must communicate with the control node of another HBase system to achieve effective data distribution and load balancing. Algorithm1 ensures that data clusters are formulated based on event types, and these data clusters can be effectively managed by the HBase system for monitoring automation applications. For future big data transmission for large-scale automation applications based on the Internet of Things. In the following steps, a standard IoT big data storage engineering method is described.

- Step 1. Identify the total space of IoT.
- Step 2. Identify disjoint logical IoT network segments.
- Step 3. Observe the heterogeneous distributed data sources in each network segment.
- Step 4. Treat the data source of each market segment as a Bigraw data source for IoT.
- Step 5. Set up a physical sensing and actual storage layer between the physical sensing and actual storage layer.
- Step 6. Set up and configure the storage management tool (HBase).
- Step 7. Check storage incompatibility.
- Step 8. Normalize storage incompatibilities.
- Step 9. Implement Algorithm 1 and store the stored content to the corresponding storage server.

```
(1) Input agile heterogeneous data streams.
(2) Process executes an If...then...else classifier system to store the data streams based on event types.
(3) Output storage of data streams into the respective data clusters such that each cluster stores the identical data of a specific event type.
(4) C_1, C_2, C_3, C_4, C_5: data clusters;
(5) // Check periodic aggregated data streams (DS_i; i = 1, ..., n);
(6) While (availability of DS_i = true)
    (7) {
        (8) If (DS_i = C_1, event type)
            (9) Then move.data.streams (DS_i, C_1);
        (10) Else if (DS_i = C_2, event type)
            (11) Then move.data.streams (DS_i, C_2);
        (12) Else if (DS_i = C_3, event type)
            (13) Then move.data.streams (DS_i, C_3);
        (14) Else if (DS_i = C_4, event type)
            (15) Then move.data.streams (DS_i, C_4);
        (16) Else
            (17) move.data.streams (DS_i, C_5);
        (18) }
    (19) Set (availability of DS_i = false)
    (20) // Check the move operations
    (21) If (completed.move = true)
        (22) Then goto Step (2); // for next periodic DS_i;
    (23) Else continue the move operations;
```

**Figure 3.** Algorithm 1. Data classifier system based on five clusters
3.4. Big Data Analysis
Data analysis is based on standard IoT big data management and knowledge discovery strategies, taking into account data modeling, visualization, and the presentation of published data and knowledge. The main goal of analysis is to transform IoT big data into knowledge, actions, and decisions in real time. Cognitive computing intelligence tools are used as catalysts for data management and knowledge discovery to generate cognitive decisions, plans, and drives for large-scale industrial automation applications. A real-time data analysis method based on fuzzy neural genetic algorithm is proposed, which can be applied IoT big data management and knowledge discovery for future large-scale industrial automation applications. The detailed method is described in Algorithm 2.

![Figure 4. Algorithm 2. Real-time data analysis based on fuzzy neural genetic algorithm](image)

4. Results analysis and discussion
This article considers the theoretical analysis of our proposed COIB framework, which contains heterogeneous IoT big data to further analyze the industry's strategy, tactics, and operational decisions. Operational decisions play a vital role in industrial automation applications, while strategy and tactical decision-making involves planning automation for industrial applications. The COIB framework mainly focuses on the prospects of IoT big data management and knowledge discovery, so a detailed review is made in Table 1. The researchers analyzed the overall big data in the IoT environment numerous systems, models, and frameworks for management and knowledge discovery activities. The review shows that the main real-time data mining and knowledge discovery activities can also collect Meta data, information, and explicit knowledge used to standardize automated applications. And explicit knowledge. All the big data management and knowledge discovery activities of Table 1 are integrated into a unified knowledge discovery strategy in which three different areas of data mining, data warehouse, and machine intelligence are integrated to achieve Automation applications. Almost all automation applications are business-critical, safety-critical, and mission-critical. Their failure can lead to significant business, social and community property damage. Therefore, timely data management and knowledge discovery applications expected by addressing potential threats and challenges, to minimize these losses have a key role.

4.1. Implementation Architecture
The implementation architecture shows how the COIB framework is implemented in large-scale industrial automation applications. The architecture is shown in Figure 5. The data center is responsible for performing data aggregation, classification, and storage operations. A dedicated and powerful server can be used in each center Perform their own operations. Knowledge production centers acquire accurate data and use cognitive computing intelligence (CI) tools to continue to generate explicit knowledge.
The action center receives this knowledge, identifies important actions, uses the action queue to prioritize these actions, and sends them to the supervisory center for immediate action. Based on the actions, the supervisory center automatically generates supervisory instructions and transmits them to the IoT-based microcontroller objects to take action on operations. Since IoT big data management mainly includes data management and knowledge discovery processes, we will focus on the four subsystems to be managed under the COIB framework (see Figure 6). In order to improve data management and the efficiency of knowledge discovery and the management of these four subsystems play a vital role.

4.2. Knowledge Exploration
Knowledge organizations from IoT big data always need databases and knowledge bases to infer cognitive output, such as decisions, plans, and incentives, to monitor time-critical automation.
applications. IoT big databases consist of an ordered collection of facts, which are organized in a standard NoSQL framework, as shown in Box1. The problem now is to design a framework for the Internet of Things knowledge base, which consists of a predefined set of rules used to impart cognitive judgments and decisions to achieve the desired intelligence. The knowledge base framework of the Internet of Things inherits the characteristics of natural language processing model, concept dependency model, logical reasoning model, and predicate model, and builds a knowledge base information system (KBIS). KBIS consists of rules-based, math-based, statistics-based, and the text, case-based reasoning and structure-based framework are composed to integrate human historical experience, analytical skills, logical reasoning and innovative ideas, and to ensure the establishment of an effective knowledge management system. Figure 7. The framework is composed of four knowledge management activities: knowledge acquisition, knowledge storage, knowledge dissemination, and knowledge application. Knowledge acquisition implements knowledge workstations for knowledge discovery within expert knowledge networks.

Figure 7. IoT Knowledge Discovery Subsystem.

```
Rowkey M/c-id {
    Column family (IoT object-1) {
        Column T1: value, event data
        Column T2: value, event data
        ...
        Column Tn: value, event data
    }
    Column family (IoT object-2) {
        Column T1: value, event data
        Column T2: value, event data
        ...
        Column Tn: value, event data
    }
    ...
    Column family (IoT object-n) {
        Column T1: value, event data
        Column T2: value, event data
        ...
        Column Tn: value, event data
    }
}
```

Figure 8. Square1. HBase architecture of data organization subsystem
4.3. Statistical Analysis

In the future of IoT big data management and knowledge discovery, smart technology platforms such as sensors, RFID devices, and wearable smart devices can evolve into large-scale IoT big data environments. Therefore, for analysis purposes, we have used some real-time data sets of specific event types and further convert them into fuzzy event data sets (eds) in the range of [0, 1]. Table 1 gives detailed statistical analysis. The analysis shows that in industrial automation applications, if the deviation from the event instance set is small, the load distribution is more balanced, so mc-1 is more abnormal in load balancing than other machines. Here, for analysis, more than a thousand event instances are studied.

| Parameter       | Min  | Max  | Mean | Deviation |
|-----------------|------|------|------|-----------|
| MC-1 load (real)| 0.043| 1.000| 0.447| 0.348     |
| MC-2 load (real)| 0.292| 0.967| 0.430| 0.070     |
| MC-3 load (real)| 0.352| 0.929| 0.170| 0.086     |
| MC-4 load (real)| 0.128| 0.987| 0.345| 0.012     |
| MC-5 load (real)| 0.002| 0.945| 0.383| 0.113     |

**Table 1.** MCs distributed load learning

Figure 9 depicts the load distribution analysis of event instances for industrial automation applications. In Figure 10, MC load data is obtained on the x-axis to analyze the MC’s set of event instances with 95% confidence through the survivor function reliability. According to the reliability value of a single MC, MC’s MTTF (mean time to failure) can be produced without any manual intervention.

![Figure 9](image-url)
In a large IoT big data environment, a large number of event instances are regularly generated for analysis and real-time decision making. These huge sets of event instances are highly unstructured, ambiguous, and even insufficient to discover knowledge of strategic value. Therefore, in this case, machine learning tools can largely support the conversion of such a huge event instance into some explicit knowledge for effective business use, such as production planning, scheduling, decision-making, strategic construction, and more use.

Event instances are randomly assigned to five data samples, 70% of the event instances are used for training, 15% of the event instances are used for verification, and the remaining 15% are used for testing. In this analysis, we are in the largest number of event instances. Absolute zero error or minimum error probability=0.04859, as shown in Figure11. The zero error or minimal error of the learning system is zero error or very rare. Once the error calculation is controlled, the calculation output and the network generated output can be the average squared difference between them is mined to analyze the overall performance of the pre-configured machine learning system, and the system can be set up for the operational aspects of the automation application.

5. Conclusion
In this article, we propose a COIB framework for effective data management and knowledge discovery on IoT big data. We also propose an implementation architecture and a layered architecture for the entire IoT big data to improve the COIB framework in large-scale industries feasibility of implementation in an automated environment. The COIB framework follows the principles of a data-centric architecture and integrates large data streams by effectively considering data access paths, which may be suitable for large-scale space-time query management. A pyramid-shaped Internet of Things is proposed big data management system highlights important subsystems from IoT object management to real-world application management. Finally, data organization and knowledge mining subsystems that implement machine-level big data flow management are given. In addition, the entire system structure and framework the integration of the provides a real-time platform for big data management and knowledge discovery of the Internet of Things for large-scale automation applications. We have added a new example for functional analysis. In this example, the event data set is normalized and converted into a standard fuzzy data set with more than 1000 active instances, whose range is in the range of [0,1]. In statistical analysis, we estimated two cognitive parameters, such as load distribution and reliability analysis for industrial automation applications. In the computational analysis, we use the conjugate
gradient back propagation algorithm as a machine learning algorithm to calculate the calculation output and network of the fuzzy event data set the error between the generated output instances, thereby measuring the overall performance of the machine learning system.

References

[1] M. Ma, P. Wang, and C.-H. Chu, “Data management for internet of things: challenges, approaches and opportunities” in Proceedings of the Green Computing and Communications (Green-Com’13), IEEE and Internet of Things (iThings/CPSCom), IEEE International Conference on and IEEE Cyber, Physical and Social Computing, IEEE, 2013.

[2] A. Ahrary and R. D. A. Ludena, “Big data application to the vegetable production and distribution system,” in Proceedings of the IEEE 10th International Colloquium on Signal Processing and Its Applications (CSPA’14), pp. 20–24, Kuala Lumpur, Malaysia, March 2014.

[3] P. Barnaghi, A. Sheth, and C. Henson, “From data to actionable knowledge: big data challenges in the web of things” IEEE Intelligent Systems, vol. 28, no. 6, pp. 6–11, 2013.

[4] Z. Peng, Z. Jingling, and L. Qing, “Message oriented middleware data processing model in Internet of things,” in Proceedings of the 2nd International Conference on Computer Science and Network Technology (ICCSNT ‘12), pp. 94–97, Changchun, China, December 2012.

[5] Munila Turhong. Association and Application of Big Data and Internet of Things [J]. Integrated Circuit Applications, 2020, 37(03): 102-103.

[6] Zhao Dawei, Li Fan. Application of Big Data and Cloud Computing in Internet of Things [J]. China New Communications, 2020, 22(04): 126.

[7] Niu Xiaoli. Research on Big Data Storage and Management Technology of Internet of Things [J]. Computer Programming Skills and Maintenance, 2020(02): 67-68+71.

[8] Xuandan. Internet of Things Information Construction Based on Cloud Computing Network Environment and Big Data [J]. China New Communications, 2020, 22(03): 54.

[9] Yang Longpin. Design of encrypted storage system for effective information of big data in Internet of Things [J]. Automation and Instrumentation, 2019(12): 53-56+60.

[10] Zhang Baoyan. Research on the Application of Big Data in Internet of Things [J]. Shanxi Electronic Technology, 2019(06): 94-96.