

Research Article

A Cognitive Adopted Framework for IoT Big-Data Management and Knowledge Discovery Prospective

Nilamadhab Mishra, Chung-Chih Lin, and Hsien-Tsung Chang

Department of Computer Science and Information Engineering, Chang Gung University, 259 Wen-Hwa 1st Road, Kwei-Shan, Tao-Yuan 333, Taiwan

Correspondence should be addressed to Hsien-Tsung Chang; smallpig@widelab.org

Received 21 July 2014; Revised 23 October 2014; Accepted 23 October 2014

Academic Editor: Sheng Xiao

Copyright © 2015 Nilamadhab Mishra et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In future IoT big-data management and knowledge discovery for large scale industrial automation application, the importance of industrial internet is increasing day by day. Several diversified technologies such as IoT (Internet of Things), computational intelligence, machine type communication, big-data, and sensor technology can be incorporated together to improve the data management and knowledge discovery efficiency of large scale automation applications. So in this work, we need to propose a Cognitive Oriented IoT Big-data Framework (COIB-framework) along with implementation architecture, IoT big-data layering architecture, and data organization and knowledge exploration subsystem for effective data management and knowledge discovery that is well-suited with the large scale industrial automation applications. The discussion and analysis show that the proposed framework and architectures create a reasonable solution in implementing IoT big-data based smart industrial applications.

1. Introduction

The IoT objects may be defined as any tangible thing associated with the real world entities such as man, machine, and animals, having unique identification and self-directed data transfer capability over the network. To technically construct an IoT object, three tiny components such as sensing chip, electromagnetic coil, and regulating capacitor are to be embedded into a very small sized container that can easily place into any real world entity associated with the IoT applications; however, the technical configuration actually depends on types of entities and applications. More specifically we can define IoT object as a sensor or RFID device or any smart object having internet connectivity over physical IP and capable to transmit the data to the network autonomously without any human interference [1]. The IoT technology integrates with the big-data approach with the purpose of regulating smart industrial automation application [2]. In a large scale industrial automation applications, thousands of automated machines are fabricated with such trillions of IoT chips to technically construct the IoT objects and the networks of such trillions of IoT objects may constitute a large scale industrial IoT environment, from where huge

structured, semistructured, and unstructured IoT big-data are produced in a real timescale [3]. Basically, the structured data are having precise layouts whereas the semistructured and unstructured data are in log format that ensure high data redundancy, inconsistency, anomalies, incompleteness, and so forth and also create a hazard to data management and knowledge discovery. The IoT big-data management system may be defined as a distributed computing system that especially deals with such semistructured and unstructured IoT big-data [4]. Due to high elasticity, flexibility, and dynamicity, the IoT objects penetrate into each and every real-time monitoring application such as business, scientific, industrial, health care, agriculture, animal farming, transportation, and many more [5]. In almost all IoT applications, a huge amount of data is dumped into the storage that are highly required for the purpose of data analysis, modelling, information transformation, knowledge production, and the decision generation. In upcoming large scale industrial automation, the quantity of IoT objects may be some trillions that can generate some thousand exabytes of data. For example, the IoT objects used in a large scale industrial electrical power grid infrastructure generates huge log data at the rate of one terabyte per day if data are extracted in every 5–15 minutes [6].

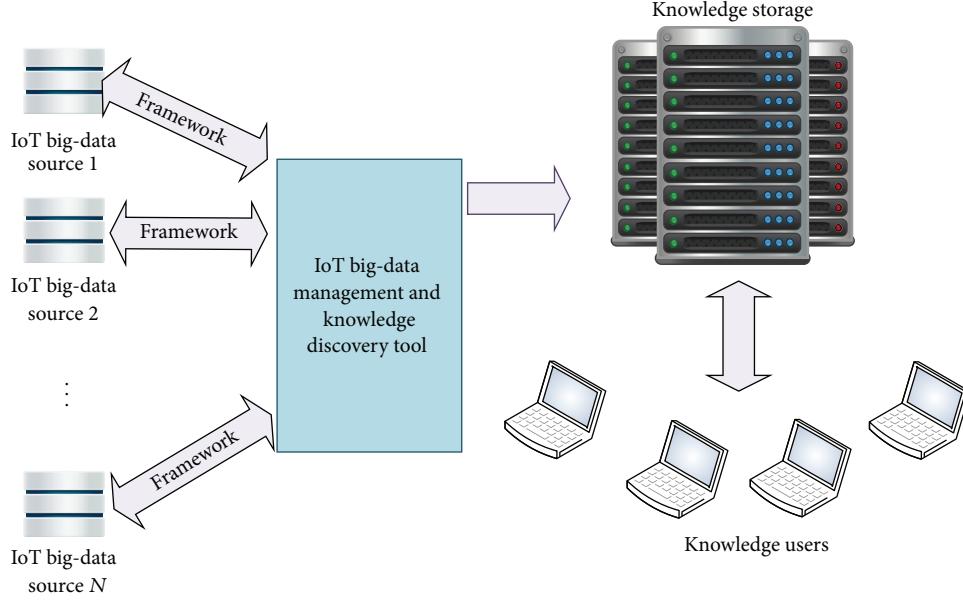


FIGURE 1: Knowledge depository system architecture.

So such IoT big-data needs suitable analytic framework to produce knowledge so as to measure the operational efficiency, load distributions, machine maintenance, and so on in industrial automation application. The digital technology is tremendously developed to scale the storage of IoT big-data; however, the data management and knowledge discovery speed may not be improved significantly to perform in a timeline basis. The IoT big-data management and knowledge discovery aims to a COIB-framework for real-time data management, so it remains a challenging issue for the knowledge mining researchers. For an effective data management and knowledge discovery, the COIB-framework always needs the suitable data management and analytic tools to transform large scale heterogeneous agile data streams into actionable knowledge, and the architecture is presented for the said purpose in Figure 1.

In heterogeneous knowledge depository framework, the IoT big-data sources are associated with different frameworks having different logical schemas. In each logical schema, the data structure is defined. The frameworks are heterogeneous because of using different models, name, scale, structures, and levels of abstraction. So it is tedious for the IoT big-data management and knowledge discovery tool to synchronise the data access so as to produce the knowledge of tactical values in a specific time scale. The IoT big-data frameworks use different formats and structures for data storage that may lead to data integration and transformation hazards. As a result, we need a comprehensive COIB-framework for the effective integration of heterogeneous IoT big-data sources in such a way that the higher layer can extract the right data and meta data for better data management and knowledge discovery prospective.

The remainder of this paper is organized as follows. In Section 2, we discuss related study on IoT big-data management and knowledge discovery. In Section 3, we propose

a COIB-framework for IoT big-data management and knowledge discovery aimed at regulating large scale automation applications. In Section 4, we present the implementation architectures, data organization, and analysis so as to stimulate the feasibility of the COIB-framework in real-time industrial automation applications. Finally, we conclude this paper in Section 5.

2. Related Work

The IoT big-data management and knowledge discovery is a key research challenge for the real-time industrial automation applications. So we study some existing system, models, or frameworks that are implemented in IoT big-data management and knowledge discovery prospective. The IoT big-data management includes several managerial activities such as data collection, integrations, cleaning, storage, processing, analysis, and visualizations that have been implemented through various systems, models, and frameworks [7–10]. The work in [11] gives a brief research direction in data management and knowledge discovery prospective in IoT big-data management platform, in which three activities are mainly associated, that is, data association, inference, and knowledge discovery. The work in [12] specially emphasizes a cognitive IoT framework for effective decision making and knowledge discovery. In order to enhance the decision making process, the work in [13] emphasizes designing an information management framework that performs data collections and information extraction operations in a real time basis so as to meet the need of the applications. Also, some ontology based modelling mechanisms are suggested to enhance the knowledge discovery from IoT big-data [14].

The traditional IoT big-data research uses noncognitive framework to perform numerous data management activities; however, the data management and knowledge

discovery always need a cognitive framework to perform the self-regulated operations in case of emergencies without human interventions. In industrial automation applications, the cognitive data management and knowledge discovery framework must have significant role towards self-regulated monitoring operations. So we propose a COIB-framework for effective data management and knowledge discovery within the industrial automation environment.

3. COIB-Framework Organization

To implement the COIB-framework, we must ensure that the automation environment has high industrial internet connectivity among the networked IoT objects that are fabricated in different machines so as to regulate the machines functional and operational efficiency. The functional aspects of the COIB-framework organization are presented in Figure 2. For a large scale parallel and distributed environment, each IoT segment has an independent control to application monitoring. So the entire industrial automation environment may be logically divided into several IoT segments so as to be compatible with the standard network configuration management requirements [15]. The IoT segments are responsible for producing IOT raw-data streams and act as raw-data sources for their corresponding segments. Those IoT big-data streams are highly unstructured with respect to the name, scale, abstraction level, and so forth and ensure high data inconsistency, redundancy, incompleteness, and other anomalies. So the big-data aggregators do the data fusion operations over those large IoT big-data streams. The data fusion operation may use some standard data semantics to eliminate those anomalies so as to produce the cleaned data with total quality management.

The IoT big-data classifiers split the cleaned data into multiple clusters based on data behaviours, characteristics, and domain to ensure it is easy to access, easy to understand, and easy to use. In the industrial automation application, the data domain includes operational data, production data, status data, maintenance data, and so forth. The HBase storage takes the responsibility to scale and store those data clusters into its multiple storage nodes. In HBase system, the tables can be set like relational database, such that each table contains rows and columns, and each table must have an element that defines as the primary key. For better supervision and access, the master node holds the control over the storage nodes. The storage nodes store the actual data clusters whereas the master node stores the metadata, that is, all access paths to the storage nodes. So in HBase system, the IoT big-data can be scaled effectively and efficiently. The IoT big-data analysis is an important aspect of regulating the data management and knowledge discovery issues through the cognitive and computational intelligent tools. For an industrial automation environment, several standard cognitive data analytic tools can be used to produce the cognitive decisions, plans, and actuations so as to control the total automation environment. Thus, in order to have the success of COIB-framework implementation in an automation environment, the IoT big-data generation,

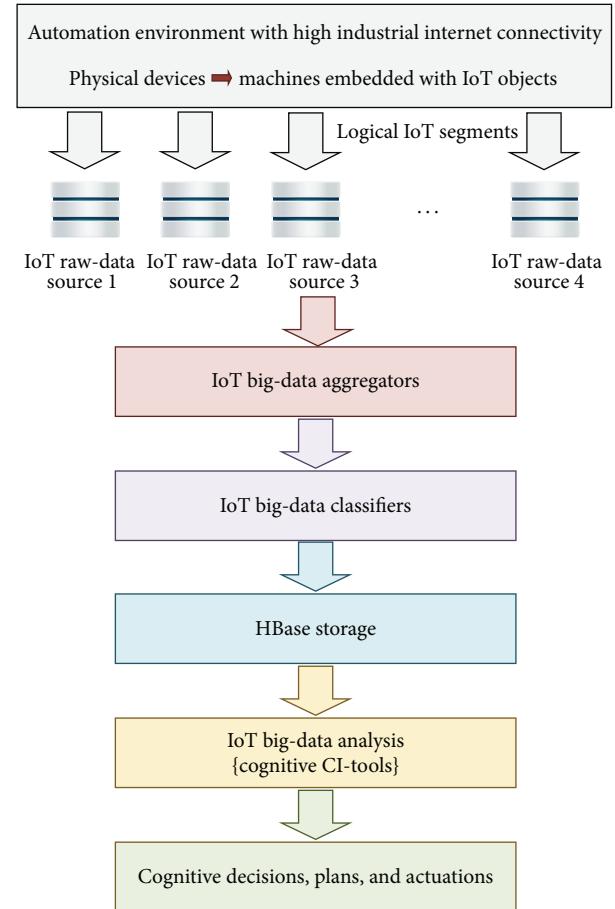


FIGURE 2: COIB-framework organization.

aggregations, classifications, storage, and the analysis must be synchronised in a real timescale basis.

3.1. Functional Analysis of COIB-Framework. Here, four important functions of COIB-framework are discussed and those functions are as follows. (i) IoT big-data aggregation, (ii) IoT big-data classification, (iii) IoT big-data storage, and (iv) IoT big-data analysis.

3.1.1. IoT Big-Data Aggregation. It encompasses the fusion of large scale data streams according to some standard data semantics. In each periodical time interval, the data aggregator checks the data availability in individual IoT raw-data source, if some availability finds then proceeds to perform the data integration process with an aim to achieve data integrity, consistency, completeness, and overly the total quality management. The data quality assessment is an important aspect of total quality management that further leads to achieving the desired quality of big-data analysis results. For a real-time monitoring application, once the data streams are successfully aggregated or fused in a periodic interval, the subsequent functional processes must be executed for effective data management and knowledge discovery within the critical timeline.

```

(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 to 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 ( $DS_i; i = 1, \dots, n$ );
(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;

```

ALGORITHM 1: Five clusters based data classifier system.

3.1.2. IoT Big-Data Classification. The real-time data classification and clustering for agile data streams are the challenging issues especially in IoT big-data environment. The data classifier initiates the classification process to formulate the fused data into multiple data groups, such that homogeneous data present in each data group. For an automation application, the fused data can be classified into multiple groups of having multiple event types such as machine status data, functional data, inventory data, production data, product quality data, and many more. Some nonlinear data classification process over the multilayer perceptron may be fit to the problem of large scale IoT big-data classification problem. For the time critical big-data classification and clustering, a simple “If...then...else” classifier can be used to formulate the aggregated data into multiple data clusters. For better clarification, let us design a five clusters based data classifiers as described in Algorithm 1. In Algorithm 1, we assume that aggregated IoT big-data streams = DS_i , for $i = 1 \dots n$, and each cluster stores the identical data of a specific event type.

3.1.3. IoT Big-Data Storage. In the current days, the technology is mounted to the large scale data storage in the IoT big-data environment. The HBase database management system takes the responsibility for IoT big-data storage across storage nodes that are headed by a control node. Based on application, the entire IoT space can be segmented into nonintersecting segments such that, in each segment, the HBase system can be implemented so as to ensure a distributed control over the large scale IoT big-data base. The control node of one HBase system must communicate with the control node of another HBase system for effective

data distribution and load balancing. Algorithm 1 ensures to formulate the data clusters based on event types and those data clusters can be effectively managed by the HBase system towards monitoring the automation applications. A standard IoT big-data storage engineering approach is described in following steps for the purpose of future big-data sensing and storage aimed at large scale IoT based automation applications.

Step 1. Recognize total IoT space.

Step 2. Identify the nonintersecting logical IoT segments.

Step 3. Observe the heterogeneous distributed data sources in each segment.

Step 4. Consider the data source of each segment as IoT big-raw data source.

Step 5. Set and configure the middleware in between the physical sensing and actual storage layer.

Step 6. Set and configure a storage management tool (HBase).

Step 7. Check the storage incompatibilities.

Step 8. Normalize the storage incompatibilities.

Step 9. Implement Algorithm 1, to place the data streams into the data clusters of corresponding storage servers.

In each IoT segment, a network of millions of smart IoT objects may be there to regulate the large scale automation

- (1) Input-identical data clusters based on event types.
- (2) Process-executes a machine learning system.
- (3) Output-generates cognitive decisions, plans, and actuations.
- (4) Validate the signals and event data.
- (5) Extract the identical event features.
- (6) Prepare training data sets.
- (7) Use GA (genetic algorithm) based optimization technic to extract selective event features from training data sets.
- (8) Set and configure an NN (neural network) architecture that fit with the problem.
- (9) Set and configure fuzzy weights from the event features.
- (10) Train NN-structure with fuzzy weights.
- (11) When training is over, test the NN-structure with the selective event features that consider as data sets.
- (12) Trace the cognitive outputs.
- (13) Diagnose those cognitive outputs to generate cognitive decisions, plans, and actuations so as to regulate the machine operations for the automation application.

ALGORITHM 2: Fuzzy-neuro-GA based real-time data analysis.

application. So the necessary storage configuration should be made carefully so as to ensure an effective synchronization in between the architectural layers.

3.1.4. IoT Big-Data Analysis. The data analysis takes the account of data modelling, visualization, and presentations of published data and knowledge in accordance with standard IoT big-data management and knowledge discovery strategy. The main aim of the analysis is to turn IoT big-data into knowledge, actions, and decisions in a real-time basis. The cognitive adopted computational intelligence tool acts as catalyst to data management and knowledge discovery in order to produce the cognitive decisions, plans, and actuations for the large scale industrial automation applications. We propose a fuzzy-neuro-GA based real-time data analysis approach that can be applied to future IoT big-data management and knowledge discovery for large scale industrial automation application. The detailed approach is described in Algorithm 2.

4. Analysis and Discussion

It considers the theoretical analysis of our proposed COIB-framework that accommodates heterogeneous IoT big-data sources so as to produce the higher knowledge abstracts that can further analyse for the industrial strategic, tactical, and operational decisions. The operational decisions have key role towards the industrial automation application; however, the strategic and tactical decisions are relating to the planning automations of the industrial applications. The COIB-framework mainly aims to IoT big-data management and knowledge discovery prospective, so a detailed review is presented in Table 1. The researchers have analysed numerous systems, models, and frameworks for the overall big-data management and knowledge discovery activities over IoT environment. The review reveals that the major real-time data mining and knowledge discovery activities can also be carried out over the IoT big-database so as to generate the meta data, information, and explicit knowledge for

regulating the automation applications. All big-data management and knowledge discovery activities of Table 1 are integrated into an unify knowledge discovery strategy, in which three diversified areas such data mining, data warehousing, and machine intelligence are integrated toward achieving a new prospective over IoT big-data based smart industrial automation application. Almost all automation applications are of business-critical, safety-critical, and mission-critical types, the failure of which lead to huge losses of business, society, and social property. So in-time data management and knowledge discovery have the key role to minimise those losses through resolving the possible threats and challenges that anticipate in the applications.

4.1. Implementation Architecture. The implementation architecture shows the way to implement the COIB-framework in large scale industrial automation applications and the architecture is shown in Figure 3. The data center is responsible to execute the data aggregation, classification, and storage operation. In each center, dedicated powerful servers can be used to carry on the respective operation. The knowledge production center access the precise data and proceeds to generate the explicit knowledge using cognitive computational intelligence (CI) tool.

The actuation center receives that knowledge, identifies the important actions, prioritizes those actions using an action queue, and sends to the regulatory center to take immediate actions. Based on actions, the regulatory center automatically generates regulatory instructions and transfers to the IOT based microcontroller objects to take actions against the operations. As the IoT big-data management mainly incorporate the data management and knowledge discovery process, so we focus the four subsystems (see Figure 4) that are to be managed under the COIB-framework. In order to improve the efficiency of data management and knowledge discovery, those four subsystems management play vital role.

The large scale industrial automation application needs a complete architecture to implement the COIB-framework,

TABLE 1: Comparison of various works based on the system/model/framework along with big-data management and knowledge discovery prospective over IoT platform.

Name of works/authors	Proposed system/model/framework	IoT big-data management and knowledge discovery prospective
Ma et al. [1]	IoT data management reference model	(i) Data cleaning (ii) Data storage (iii) Data processing (iv) Data analysis
Stankovic [11]	Data mining system	(i) Data association (ii) Data inference (iii) Knowledge discovery
Li et al. [16]	Discovery service model	(i) Data collections (ii) Data storage
Peng et al. [4]	Data processing model	(i) Data communication (ii) Data processing
Jin et al. [13]	Information framework	(i) Data collection (ii) Data evaluation (iii) Knowledge extraction (iv) Decision management
Wu et al. [12]	Cognitive IoT framework	(i) Data analysis (ii) Semantic derivations (iii) Knowledge discovery (iv) Intelligent decision making
Barnaghi et al. [3]	Web of things framework	(i) Data aggregation (ii) Query and discovery (iii) Knowledge extraction
Ding et al. [9]	IoT statistic DB model	(i) Data storage (ii) Data distributions (iii) Data access analysis
Zhou et al. [14]	Fuzzy ontology model	(i) Data modelling (ii) Data reasoning (iii) Data inference
Tang et al. [17]	PARHACS system	(i) Data analysis (ii) Data management
Zanella et al. [18]	Database and ERP system	(i) Data collection (ii) Data storage (iii) Data analysis
Stojmenovic [19]	CPS based M2M system	(i) Data acquisition (ii) Data communication
Nastic et al. [20]	PatRICIA framework	(i) Data storage (ii) Storage manipulations and analysis (iii) Data management
Tracey and Sreenan [21]	Holistic architectural framework	(i) Data abstraction (ii) Data aggregation (iii) Data storage
Perera et al. [22]	Service discovery and configuration framework	(i) Data acquisitions (ii) Data management
Mukherjee et al. [23]	Condor framework	(i) Data analysis (ii) Storage analysis
Ramaswamy et al. [24]	DQ based big-data architectural framework	(i) Data quality management (ii) Storage management (iii) Data analytics
Our work	COIB-framework	(i) Knowledge storage (ii) Cognitive decision, plan, and actuation management (iii) Data organization and knowledge exploration (iv) Apply for industrial automation application

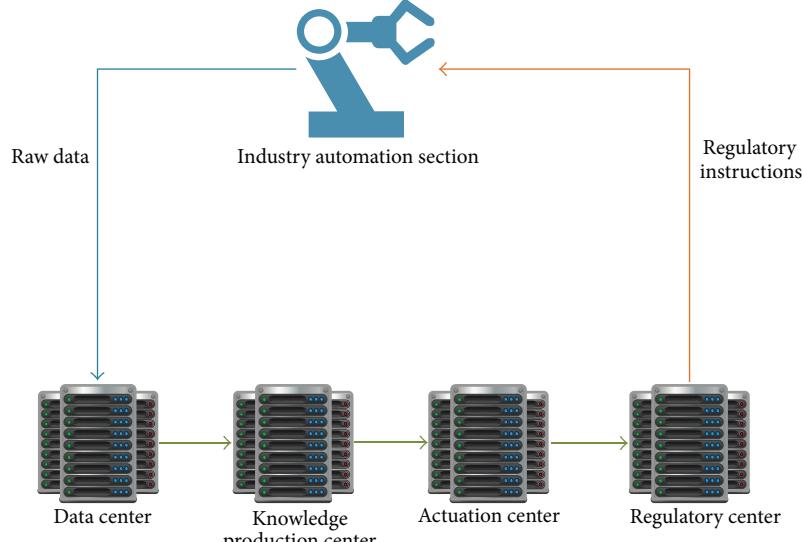


FIGURE 3: Implementation architecture.

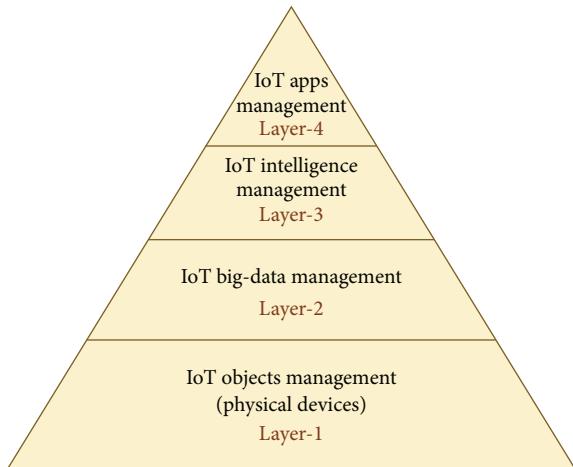


FIGURE 4: IoT big-data management subsystem.

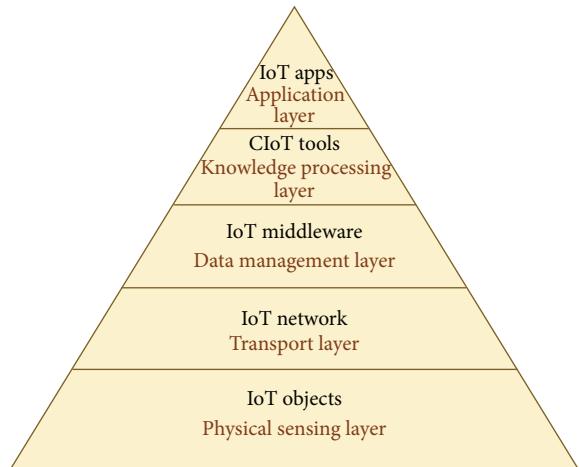


FIGURE 5: Overall IoT big-data layering architecture.

so a discussion is made on the overall IoT big-data layering architecture with a purpose of wide implementations.

An overall architectural representation is presented in Figure 5.

In this pyramidal architecture, the knowledge processing layer is specially introduced with a purpose of effective knowledge production that may be exploited in the application layer so as to manage various IoT applications. The effective knowledge production targets the performance of cognitive tools like fuzzy ontology, data semantics, neural computing, and so forth. In automation application, the IoT network uses the industrial internet to integrate the multi-faceted physical machines with networked IoT objects and intelligent tools so as to regulate the automated operations [25]. The industrial internet integrates the various diversified technology like machine to machine communication, machine learning, IoT, big-data, cloud, and computational

intelligence for industrial automation applications. The IoT middleware is mainly responsible for data processing, management, and communication in addition to maintaining the data perturbations in between the architectural layers [26].

4.2. Data Organization. In an IoT big-data system, bulk amounts of data are organized in the form of NoSQL databases [27]. The IoT big-data is a spatiotemporal database that depends on the time and location. In an IoT big-data, more numbers of rows are there along with less number of columns. So the column oriented data-depository can greatly improve the performance of IoT big-data in terms of data accessing and query processing. The heterogeneous IoT big-data cannot be stored in any relation database. So IoT big-data cell (NoSQL database) may be used to resolve the storage limitation and constraints of relational database [16, 28]. The data organization is an important aspect in

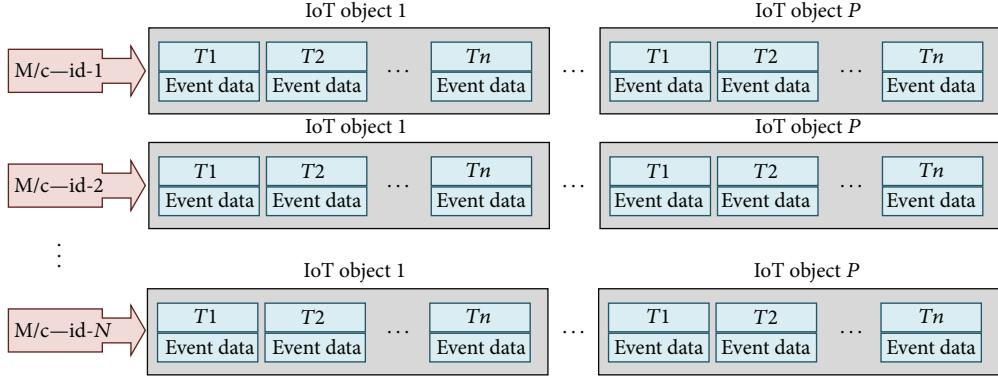


FIGURE 6: Data organization subsystem.

```

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-P) {
Column T1: value. event data
Column T2: value. event data
:
Column Tn: value. event data}
}

```

Box 1: HBase schema for data organization subsystem.

large scale industrial automation application. So in order to implement the IoT big-data system in large scale automation applications, a data organization subsystem is presented in Figure 6. In this framework, we assume N -number of machines are compounded with the IoT big-data system and regulated under the COIB-framework. In each machine, P -number of IoT objects are embedded in such a way that each IoT object can take the responsibility of sensing and sending n -number of events data to the COIB-framework, where Tn is the n th event data recording time stamp.

In order to ensure the database implementation, an HBase schema for the above data organization framework is described in Box 1. In this logical schema, we assume the arrival of events data from IoT big-data environment at n -number of time intervals to the COIB-framework.

4.3. Knowledge Exploration. The knowledge organization from IoT big-data always needs a database and knowledge base to

infer the cognitive outputs like decisions, plans, and actuations for monitoring time critical automation applications [29]. The IoT big-database consists of organised collection of facts that are organized in a standard NoSQL framework as illustrated in Box 1. Now the problem is designing a framework for IoT knowledge base that consists of predefined sets of rules to impart cognitive judgements and decision making for achieving desired intelligence. An IoT knowledge base framework inherits the features of natural language processing model, conceptual dependency model, logical inference model, and predicates model to construct a knowledge base information system (KBIS). The KBIS consists of rule-based, mathematical-based, statistical-based, textual-based, case-reasoning-based, and structure-based frameworks to incorporate human historical experience, analytical skills, logical reasoning, and innovative ideas, and ensures an effective knowledge management system. We highlight an IoT knowledge exploration subsystem as described in Figure 7. The framework consists of four knowledge management activities such as knowledge acquisition, knowledge storage, knowledge dissemination, and knowledge application. The knowledge acquisition implements a knowledge work station for knowledge discovery within the expert knowledge networks.

The knowledge storage performs the document management of IoT knowledge database within the operational framework of the pre-connected expert system. The knowledge dissemination uses search engines to spread the knowledge into the time critical automation application and the knowledge application includes the planned management of enterprise applications through explicit and tactical knowledge in order to achieve the core competencies of the IoT big-data based smart industrial applications.

4.4. Statistical Analysis. In future IoT big-data management and knowledge discovery prospective, the platform of sensors, RFID devices, wearable intelligent devices, and other smart technologies can be evolved into a large scale IoT big-data environment. So for analysis purpose we use some real-time data sets of specific event type and further transformed into fuzzified event data sets (EDS) in the range of $[0, 1]$ [30]. The event is taken as a load distribution study among

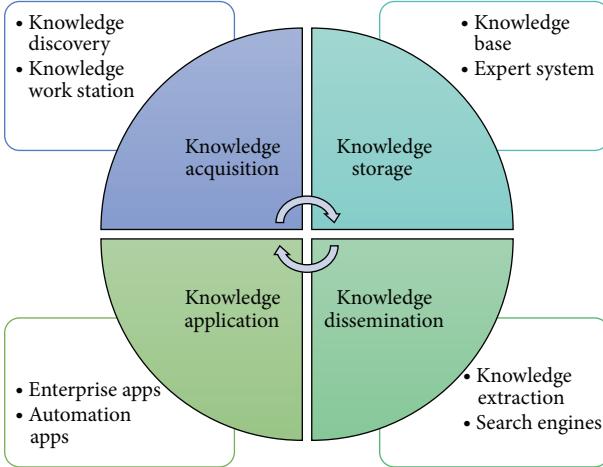


FIGURE 7: IoT knowledge exploration subsystem.

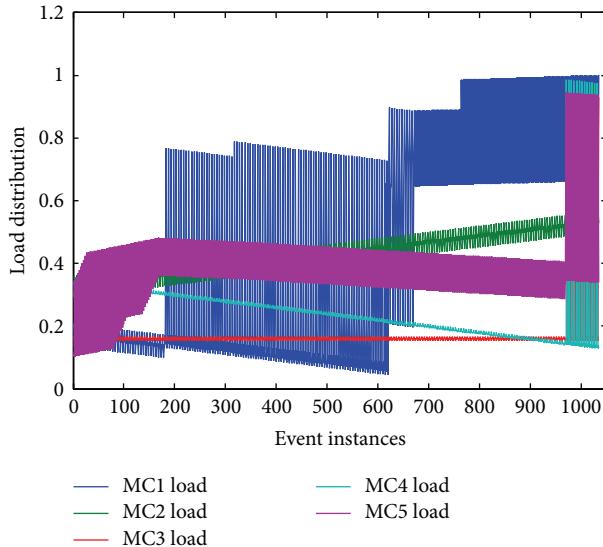


FIGURE 8: Load distribution analysis of MCs.

the five different machines (MCs) of an industrial automation application. A detailed statistical analysis is presented in Table 2. The analysis tells that in an industrial automation application, the load distribution is more balanced if the deviation is less with respect to set of event instances, so the MC-1 has the higher degree of abnormality in load balancing as compared to other machines. Here, more than one thousand event instances are studied for the analysis.

Figure 8 describes the load distribution analysis with respect to the event instances for an industrial automation application. In Figure 9, the MC's load data are taken in x -axis to analyse the reliability of MCs with respect to the set of event instances through survivor function with a confidence level of 95%. Based on the reliability values of individual MC, the MC's MTTF (mean time to failure) can be estimated to avoid probable failures and hampering of production without any human interventions. In a large

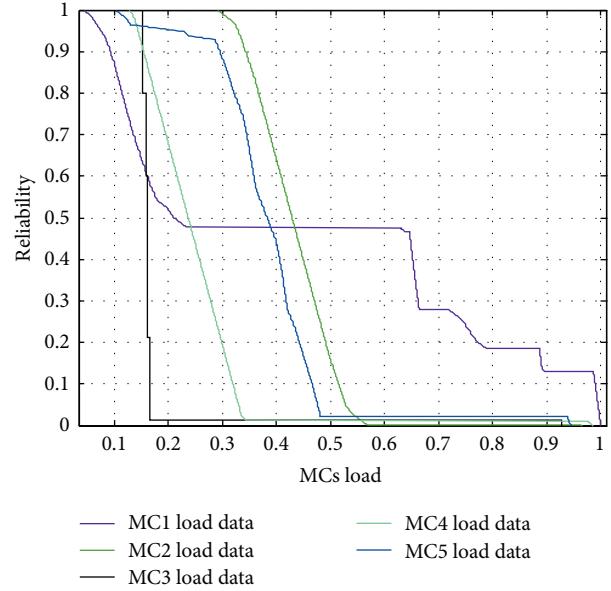


FIGURE 9: Reliability analysis of MCs.

scale IoT big-data environment, at a regular time interval such enormous sets of event instances are generated for the purpose of analysis and real-time decision making. Those enormous sets of event instances are highly unstructured, ambiguous, and even not adequate to discover the knowledge of strategic values. So in this situation, the machine learning tools can greatly support to turn such huge event instances into some explicit knowledge for effective business use such as production planning, scheduling, decision making, strategy building, and many more business uses.

4.5. Computational Analysis. This phase discusses the computational analysis of the COIB-framework by mapping the possible functions onto the neural network platform to minimise the framework's error. Here, we implement a computational machine learning approach based on Algorithm 2, in which the neurofuzzy mechanism is more effective in dealing with more tolerance and uncertainties that are actually faced by the time-critical applications [31, 32]. The machine learning approach aims at training and testing the COIB-framework to ensure functional and operational accuracy within the large scale automation environment. We use fitness parameter θ that mimics the behavioural aspects of genetic algorithm in order to classify the data sets into multiple generic data clusters. The higher fitness value θ associated with an event dataset implies higher quality, so to classify the quality of event datasets, we can define three fuzzy membership grades through varying θ values in the range of $[0, 1]$, to uniquely distinguish the average quality, good quality, and best quality event datasets (EDS) with an aim to extract selective event features.

- (i) $\mu_{(EDS)}^{(x)} : x \in X(\theta)$ and $\theta \rightarrow [0.001, 0.300]$; if and only if x is an event data element and X is an average quality EDS.

TABLE 2: Load distribution study of MCs.

Instance-ID	Least		Most		Values I001 (1), I002 (1), ⋮
	1999 (1)		1001 (1)		
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)	0.152	0.929	0.170	0.086	
MC-4 load (real)	0.128	0.987	0.245	0.012	
MC-5 load (real)	0.102	0.945	0.383	0.113	

(ii) $\mu_{(EDS)}^{(x)} : x \in X(\theta)$ and $\theta \rightarrow [0.301, 0.600]$; if and only if x is an event data element and X is a good quality EDS.

(iii) $\mu_{(EDS)}^{(x)} : x \in X(\theta)$ and $\theta \rightarrow [0.601, 1.000]$; if and only if x is an event data element and X is a best quality EDS.

Now the selective features can be extracted based on the θ value associated with an event dataset that can be used for COIB-framework learning. The fuzzy weights are also configured from the selective EDS ensuring better tolerance and precision. Here more than one thousand event instances are assumed in between the time intervals T_i , where $i = 1 \dots n$. Based on the event instances, we can create and train a neural network and estimate its performance through computing the errors (e); that is, $e = |T_o - C_o|$, where T_o is the target output and C_o is the computed output or the network generated output of the neural network. Those event instances can be mapped to a multilayer feed forward neural network with conjugate gradient back propagation algorithm as training algorithm [33]. In this algorithm, the weights of entire network may be considered as single input vector and their derivatives constitute the gradient vector [34].

The event instances are randomly allocated into five data samples such that 70% event instances are used for training, 15% event instances for validation, and the rest 15% for testing. In this analysis, we get absolutely zero error or minimum error $e = 0.04859$ among maximum number of event instances as depicted in Figure 10. The zero error or very low error indicates the satisfiable performance measurement of the learning system. As soon as the error computation is completed, the mean square error can be computed to determine the average square difference between the computed output and network generated output to analyse the overall performance of the preconfigured machine learning system and the system can be set for the operational aspects of the automation application.

5. Conclusions

In this paper, we propose a COIB-framework for the effective data management and knowledge discovery over the IoT big-data. We also propose implementation architecture along with the overall IoT big-data layering architecture to promote

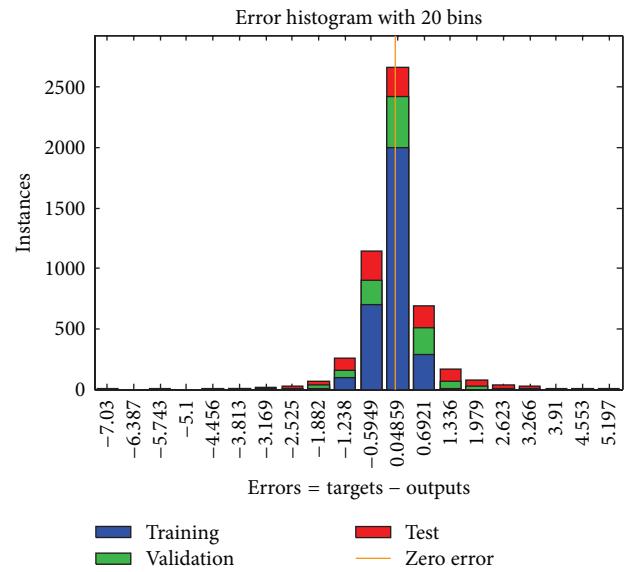


FIGURE 10: Error computation on event instances.

the implementation feasibility of the COIB-framework in large scale industrial automation environment. The COIB-framework follows the principles of data centric architecture to incorporate the large big-data streams by efficiently considering the data access paths and may be adequate in large spatiotemporal query management. A pyramidal IoT big-data management system is proposed to highlight the important subsystems rendering from IoT object management to real world applications management. Finally data organization and knowledge exploration subsystems are presented to implement the machines level big-data streams management. Also the integration of total architectures and frameworks gives a real time platform in IoT big-data management and knowledge discovery prospective for a large scale automation application. We add an absolutely new case in point for the functional analysis, in which the event data sets are normalized and transformed into standard fuzzified data sets having more than one thousand activity instances in the range of [0, 1]. We perform a statistical and a computational analysis to support the effectiveness of our proposed COIB-framework over IoT big-data platform for

an industrial automation application. In statistical analysis, we estimate two cognitive parameters such as load distribution and reliability analysis for industrial automation application. In computational analysis, we implement conjugate gradient back propagation algorithm as machine learning algorithm to compute errors between the computed output and the network generated output instances of the fuzzified event data sets so as to measure the overall performance of the said machine learning system.

In our future issue we would like to propose an IoT big-data engineering and reengineering framework for the IoT big-data applications. The future issues of the IoT big-data management may include time schedule based data transformation strategy, mining and analysis of large data streams generated by trillions of IoT objects employed in heterogeneous applications, and the security and data perturbation issues relating to IoT big-data management subsystems.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The financial support provided by the Ministry of Science and Technology, China, through Grants MOST 103-2221-E-182-003, MOST 103-2511-S-182-053, and CMRPD3B0033 of Chang Gung University is gratefully acknowledged.

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 (GreenCom '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] Smart Grid Big Data Management Team, Smart Grid Big Data Management Team, 2014, <http://www.accenture.com/us-en/blogs/technology-blog/archive/2014/04/28/the-right-big-data-technology-for-smart-grid-distributed-stream-computing.aspx>.
- [6] <http://www.accenture.com/us-en/blogs/technology-labs-blog/archive/2014/04/28/the-right-big-data-technology-for-smart-grid-distributed-stream-computing.aspx>.
- [7] J. Zhou, L. Hu, F. Wang, H. Lu, and K. Zhao, "An efficient multidimensional fusion algorithm for iot data based on partitioning," *Tsinghua Science and Technology*, vol. 18, no. 4, pp. 369–378, 2013.
- [8] M. Mozumdar, A. Shahbazian, and N.-Q. Ton, "A big data correlation orchestrator for Internet of Things," in *Proceedings of the IEEE World Forum on Internet of Things (WF-IoT '14)*, pp. 304–308, IEEE, Seoul, Republic of Korea, March 2014.
- [9] Z. Ding, X. Gao, J. Xu, and H. Wu, "IOT-StatisticDB: a general statistical database cluster mechanism for big data analysis in the internet of things," in *Proceedings of the IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing (GreenCom-iThings-CPSCom '13)*, pp. 535–543, IEEE, August 2013.
- [10] C. Wang, M. Daneshmand, M. Dohler, X. Mao, R. Q. Hu, and H. Wang, "Guest Editorial—special issue on internet of things (IoT): architecture, protocols and services," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3505–3508, 2013.
- [11] A. J. Stankovic, "Research directions for the internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 3–9, 2014.
- [12] Q. Wu, G. Ding, Y. Xu et al., "Cognitive internet of things: a new paradigm beyond connection," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 129–143, 2014.
- [13] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, "An information framework for creating a smart city through internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 112–121, 2014.
- [14] H. Zhou, Y. Wang, and K. Cao, "Fuzzy D-S theory based fuzzy ontology context modeling and similarity based reasoning," in *Proceedings of the IEEE 9th International Conference on Computational Intelligence and Security (CIS '13)*, pp. 707–711, December 2013.
- [15] S. Abdullah and K. Yang, "An energy-efficient message scheduling algorithm in Internet of Things environment," in *Proceedings of the 9th International Wireless Communications and Mobile Computing Conference (IWCMC '13)*, pp. 311–316, IEEE, July 2013.
- [16] M. Li, Z. Zhu, and G. Chen, "A scalable and high-efficiency discovery service using a new storage," in *Proceedings of the IEEE 37th Annual Computer Software and Applications Conference (COMPSAC '13)*, pp. 754–759, IEEE Computer Society, Kyoto, Japan, July 2013.
- [17] C.-H. Tang, M.-F. Wang, W.-J. Wang et al., "Large-scale data management and analysis for astronomical research," in *Proceedings of the 1st International Symposium on Grids and Clouds (ISGC '11), Held in Conjunction with the 31st Open Grid Forum (OGF '11)*, March 2011.
- [18] A. Zanella, N. Bui, A. P. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.
- [19] I. Stojmenovic, "Machine-to-machine communications with in-network data aggregation, processing and actuation for large scale cyber-physical systems," *IEEE Internet of Things Journal*, vol. 1, no. 2, pp. 122–128, 2014.
- [20] S. Nastic, S. Sehic, M. Vögler, H.-L. Truong, and S. Dustdar, "PatRICIA—a novel programming model for iot applications on cloud platforms," in *Proceedings of the 6th IEEE International Conference on Service-Oriented Computing and Applications (SOCA '13)*, pp. 53–60, December 2013.
- [21] D. Tracey and C. Sreenan, "A holistic architecture for the internet of things, sensing services and big data," in *Proceedings*

- of the 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing (CCGrid '13), pp. 546–553, May 2013.*
- [22] C. Perera, P. P. Jayaraman, A. Zaslavsky, D. Georgakopoulos, and P. Christen, “Sensor discovery and configuration framework for the internet of things paradigm,” in *IEEE World Forum on Internet of Things (WF-IoT '14)*, pp. 94–99, 2014.
 - [23] A. Mukherjee, S. Dey, H. S. Paul, and B. Das, “Utilising condor for data parallel analytics in an IoT context—an experience report,” in *Proceedings of the IEEE 9th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob '13)*, pp. 325–331, October 2013.
 - [24] L. Ramaswamy, V. Lawson, and S. V. Gogineni, “Towards a quality-centric big data architecture for federated sensor services,” in *Proceedings of the IEEE International Congress on Big Data*, pp. 86–93, IEEE, Santa Clara, Calif, USA, July 2013.
 - [25] GE Intelligent Platforms, Brochures: E-zine: lean manufacturing in the age of the Industrial Internet, 2014, <http://www.ge-ip.com/download/e-zine-lean-manufacturing-in-the-age-of-the-industrial-internet/13873/>.
 - [26] A. Gluhak, S. Krco, M. Nati, D. Pfisterer, N. Mitton, and T. Razafindralambo, “A survey on facilities for experimental internet of things research,” *IEEE Communications Magazine*, vol. 49, no. 11, pp. 58–67, 2011.
 - [27] M. Diaz, G. Juan, O. Lucas, and A. Ryuga, “Big Data on the Internet of Things,” 2012.
 - [28] W.-Y. Ku, T.-Y. Chou, and L.-K. Chung, “The cloud-based sensor data warehouse,” in *Proceedings of the 1st International Symposium on Grids and Clouds (ISGC '11)*, vol. 1, March 2011.
 - [29] N. Mishra, H.-T. Chang, and C.-C. Lin, “Data-centric knowledge discovery strategy for a safety-critical sensor application,” *International Journal of Antennas and Propagation*, vol. 2014, Article ID 172186, 11 pages, 2014.
 - [30] N. Mishra, C. C. Lin, and H. T. Chang, “Cognitive inference device for activity supervision in the elderly,” *The Scientific World Journal*, vol. 2014, Article ID 125618, 13 pages, 2014.
 - [31] F. Gaxiola, P. Melin, F. Valdez, and O. Castillo, “Interval type-2 fuzzy weight adjustment for backpropagation neural networks with application in time series prediction,” *Information Sciences*, vol. 260, pp. 1–14, 2014.
 - [32] D. Wu, “On the fundamental differences between interval type-2 and type-1 fuzzy logic controllers,” *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 5, pp. 832–848, 2012.
 - [33] S. Datta, A. Khasnobish, A. Konar, D. N. Tibarewala, and A. K. Nagar, “EEG based artificial learning of motor coordination for visually inspired task using neural networks,” in *Proceedings of the International Joint Conference on Neural Networks (IJCNN '14)*, pp. 3371–3378, IEE, 2014.
 - [34] J. Bilski, J. Smoląg, and A. I. Galushkin, “The parallel approach to the conjugate gradient learning algorithm for the feedforward neural networks,” in *Proceedings of the 13th International Conference on Intelligence and Soft Computing (ICAISC '14)*, pp. 12–21, Springer, January 2014.

