Scalable Framework for Distributed Case-Based Reasoning for Big Data Analytics

Shaibal Barua(✉), Shahina Begum, and Mobyen Uddin Ahmed

School of Innovation, Design and Engineering, Mälardalen University, 721 23 Västerås, Sweden
{shaibal.barua,shahina.begum,mobyen.ahmed}@mdh.se

Abstract. This paper proposes a scalable framework for distributed case-based reasoning methodology to provide actionable knowledge based on historical big amount of data. The framework addresses several challenges, i.e., *promptly analyse big data*, *cross-domain*, *use-case specific data processing*, *multi-source case representation*, *dynamic case-management*, *uncertainty*, *check the plausibility of solution after adaptation* etc. through its' five modules architectures. The architecture allows the functionalities with distributed data analytics and intended to provide solutions under different conditions, i.e. data size, velocity, variety etc.

Keywords: Distributed analytics · Case-based reasoning

1 Introduction

The Internet-of-Things (IoT) paradigm is becoming key technologies for innovative products and applications. Smart objects can communicate and interact with each other in a defined IoT enabled environment and make autonomous decisions by appropriate correlation and association of data collected from the environment [1]. The envisioned I-IoT system senses the environment, and makes decision to sensed environmental changes dynamically and effectively with optimal resources, low cost and increased convenience. Such systems bring the following main generic research challenges:

- Heterogeneous data management that makes adequate and standard formats, models and sematic description of the data to support I-IoT systems in automated reasoning.
- Smart objects can autonomously respond to a wide range of circumstances without human interventions. Consequently, these systems must have the ability to reason efficiently about contextual information.
- These systems should interact with each other with the ability to operate distribute manner which is natural in IoT environment.
- Automated decision-making should be able to handle time critical situation and would provide fast and accurate decisions.

Today, the methodologies used for such systems are not particularly focused and are rather sparse. So, from system-level perspective, in implementation of such I-IoT

M. U. Ahmed et al. (Eds.): HealthyIoT 2017, LNICST 225, pp. 111–114, 2018.
https://doi.org/10.1007/978-3-319-76213-5_16

systems, the current need is to investigate and determine core principles and algorithms. Case-based reasoning (CBR) is a methodology that solves a new problem by remembering and using previously solved problems that are similar to the current problem [2, 8, 10]. CBR can avoid the re-invent the wheel approach by remembering and using knowledge base. Moreover, sub-tasks of CBR such as case base maintenance, cases retrieval, cases adaptation and retaining new cases [3, 9] are well fitted for aforementioned challenges. However, currently there is no such distributed platform for CBR system that can meet the requirements of I-IoT and big data. Thus, this paper proposes both the state-of-the-art disciplinary and the system-oriented CBR architecture targeting to meet the generic research challenges. The proposed distributed CBR can be achieved through multi-agent CBR architecture, which can be beneficial for organizing knowledge base within the system and processing knowledge by the system [4] and maximum efficiency in I-IoT and Big data.

2 Related Work

Today even though IoT has generated excitement among the research community but still there are a number of challenges that need to be emphasized [5, 6]. An Intelligent IoT expects to exhibit intelligent behaviour by gathering multiple data and information, data management to avoid collusion, sensor fusion for robust decision, and cloud for information sharing and so on. In IoT, the gathered data needs to be managed with proper order and classification of data is also a vital part. In the field of IoT, a layer based data management system is presented in [5]. Due to globalization of IoT systems, it needs Cloud service for information sharing. A Cloud service for IoT architecture has been developed for Vehicular Data Cloud Services in [7]. Design of a generic scheme for I-IoT systems need to deal with several challenges and it needs to be able to handle the number of things and objects that will be connected in IoT contextual intelligence is crucial. Another issue is exchanging and analysing massive amount of data and the vast amount of data also need to be processed, and presented in a seamless, efficient and interpretable form.

3 Proposed Architecture

Figure 1 shows a system-oriented overview of the proposed scalable Framework for distributed case-based reasoning that illustrates the main conceptual components and their connections to each other. The overall architecture consists of five modules. The *first module* is the input sources from various smart devices and sensors connected through Internet. In this module data gathered from heterogeneous sources will be stored. As a scalable generic architecture, it is not sufficient to store only the data or signals but also to store application domain knowledge. The main objective in this module is pre-processing the gathered data, so the data is clean and noise free, and feature extraction. However, a multidisciplinary application should also include cross-domain and use case specific knowledge for future decision support.

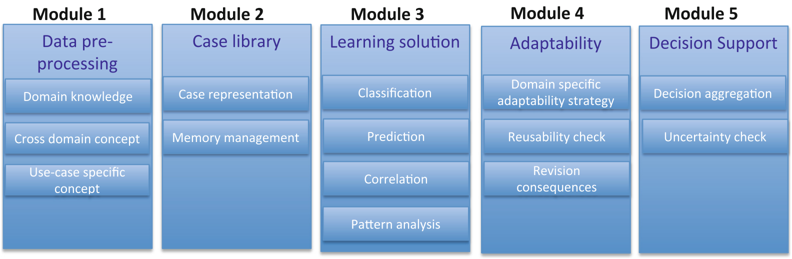

Fig. 1. Schematic diagram of the proposed case-based reasoning framework

To build a CBR system it requires to have a good knowledge base i.e., case libraries. An I-IoT system cannot be achievable without considering cloud environment. The *second module* creates collaborative case libraries in distributed manner for specific use cases and needs ability to reason efficiently about contextual information. Collaborative Big data management that makes adequate and standard formats, models and sematic description of the data using data and information processing and intelligent fusion algorithms, which will support I-IoT systems in automated reasoning. This will work with collaborative agent that can serve as the basis for automated systems that must collaborate with other local agents to interpret meaningful information from the data. One of the important modules is the *module 3* that performs various analytics for the CBR system. In the CBR, case retrieval is a major step, which seeks the most similar cases that corresponds to the target case. In the distributed multi-agent architecture, each CBR agent can be called to perform classification, prediction, and/or patterns analysis either on a single or multiple dataset. This will depend on how the case library is requested from the data storage and here one goal is to achieve the scalability through parallelism in cloud based environment.

Since knowledge bases are distributed across several nodes in clusters and a multidisciplinary application e.g., healthcare application often needs case adaption, hence *module 4* has design to perform domain specific adaptability strategy, uncertainty check in case adaption that ensures reusability of case in future learning and decision making. Last but not least, *module 5* needs to aggregate the results that are

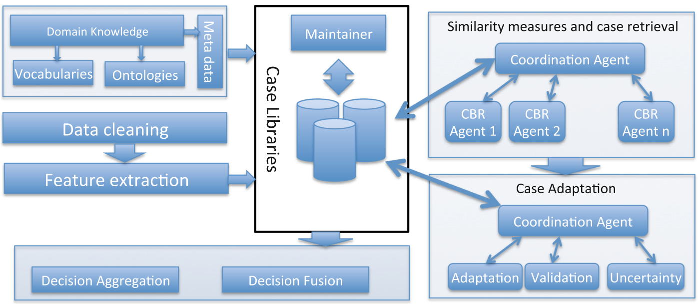

Fig. 2. Functional diagram of the proposed case-based reasoning framework

obtained from the multi-agents. This can provide the freedom to choose type of analytics i.e., using individualized knowledge or combination of knowledge for decision-making, and manage highest utilization of the cloud-based architecture. The functional connectivity among the modules in the proposed architecture is shown in Fig. 2. The functionalities should also be scalable depends of the application domain.

4 Summary

The IoT will be a basis for the future connected smart systems and sustainable infrastructures. Data management using IoT and Big data analysis on the cloud is going to be the most potential for a strong economic impact in the coming decades for the smart services. The paper has emphasized on the generic architecture for autonomous systems of future Intelligent Internet of Things (I-IoT). In addition, this system can be implemented on big data platform such as Spark with the existing support of distributed functionalities. Thus, cloud to connect intelligent things can be conducted by connecting sensors, data fusion and management through IoT; also apply them to big data analysis for automated decision-making by the scalable computing resources in the cloud.

References

1. Miorandi, D., Sicari, S., De Pellegrini, F., Chlamtac, I.: Internet of things: vision, applications and research challenges. Ad Hoc Netw. **10**(7), 1497–1516 (2012)
2. Kolodner, J.: An introduction to case-based reasoning. Artif. Intell. Rev. **6**(1), 3–34 (1992). (in English)
3. Aamodt, A., Plaza, E.: Case-based reasoning: foundational issues, methodological variations, and system approaches. AI Commun. **7**(1), 39–59 (1994)
4. Plaza, E., McGinty, L.: Distributed case-based reasoning. Knowl. Eng. Rev. **20**(3), 261–265 (2006)
5. Ma, M., Wang, P., Chu, C.H.: Data management for Internet of Things: challenges, approaches and opportunities. In: 2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, pp. 1144–1151 (2013)
6. Jianguo, M.: Internet-of-Things: technology evolution and challenges. In: 2014 IEEE MTT-S International Microwave Symposium (IMS 2014), pp. 1–4 (2014)
7. He, W., Yan, G., Xu, L.D.: Developing vehicular data cloud services in the IoT environment. IEEE Trans. Industr. Inf. **10**(2), 1587–1595 (2014)
8. Begum, S., Barua, S., Filla, R., Ahmed, M.U.: Classification of physiological signals for wheel loader operators using multi-scale entropy analysis and case-based reasoning. Expert Syst. Appl. **41**(2), 295–305 (2014)
9. Ahmed, M.U., Begum, S., Funk, P.: A hybrid case-based system in stress diagnosis and treatment. In: IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI 2012) (2012)
10. Begum, S., Barua, S., Ahmed, M.U.: Physiological sensor signals classification for healthcare using sensor data fusion and case-based reasoning. Sensors **14**(7), 11770–11785 (2014)