A Smart Cloud-Based Energy Data Mining Agent Using Big Data Analysis Technology

Hsueh-Yuan Lin and Sheng-Yuan Yang

*Department of Electrical Engineering, St John’s University, New Taipei City, Taiwan; †Department of Information and Communication, St John’s University, New Taipei City, Taiwan

ABSTRACT

The aim of this paper was the development of a smart cloud-based energy data mining agent, OntoDMA, within the cloud environment Web service-based Information Agent System. Big Data analysis technology was used, embedded in a cloud-based active multi-agent system, to proactively provide fast and appropriate, real-time, domain information prediction. A case study was done to outline and explore the feasibility of the system architecture and the energy saving information system. A preliminary system presentation and experimental verification were prepared. The cache performance of the Solutions Pool was increased by 19.82% and the query workload of the Prediction Rules was reduced by 66.51%, a considerable decrease of the back-end servo system workload.

ARTICLE HISTORY

Received 6 February 2019
Accepted 19 March 2019

KEYWORDS

Data mining agents; big data analysis; web services; energy saving agents

1. Introduction

The era of data economy began in 2014 when Gartner predicted the advent of 10 major technology trends. These included, among other things, several aspects of the cloud and big data. Data mining, a common term now in general use, refers to finding potential rules and patterns as well as reasons and relationships that have significance, from massive amounts of data to provide a critical basis for information system decision-making.
processes that use statistical analysis, informational classification, machine learning, and other related technologies. The system proposed here combines Big Data analysis, classification and other related associations, to construct relevant information mining rules using the support of a case-based system. Mining of the relevant energy saving information, operation knowledge or rules, is then done autonomously to rapidly provide an appropriate, real-time operation basis or pattern for active back-end energy saving information systems.

There are many domestic and foreign references that use data mining technology (even combined with open databases) to research and develop the paradigms of related information systems. Syu [1], for example, proposed the exploration of educational data, as based on the Big Data-based history environment, to provide parents with more in-depth information of about the learning of history. These include diagnostics and warnings, explore related technologies that improve participation by parents, increase parent–teacher interaction and educational data mining, and support learning analysis. Hsu [2] also proposed corresponding fuzzy data mining technology and explored relevant technologies for valuable health care pharmaceutical combinations, to the benefit of sales, by considering health care product prices, manufacturing equipment, the personnel payroll, competitive pharmaceutical companies, etc. Lausch et al. [3] integrated data mining with linked open data technology to explore new application perspectives and technologies related to environmental research. Vaish and Srivastava [4] discussed modern databases that support Big Data, and explored the associated challenges, architecture, and analysis tools.

Most of these studies relied on traditional data mining tools or methods, with added self-improved or self-developed mining concepts to highlight the importance of newly added mining mechanisms. The system developed in this paper also has its own relevant data mining mechanism. Specifically, a Web service-based Information Agent System (WIAS) [5], built to provide related web services. An Ontological Data Mining Agent (OntoDMA) [6,7] can initiate web services via WIAS to conduct periodic processing. This is based on case variation in an Ontological Case-based Reasoning Agent (OntoCBRA) [8], which generates Prediction Rules corresponding to case variation. Based on these Prediction Rules, OntoDMA can then process the response of related infrequent or non-existent information inquiries. Finally, an Ubiquitous Interface Agent (Ubi-IA) [9] provides corresponding query and information browsing services, which highlight self-developed data mining technologies as more suitable for application fields, and demonstrates that they are not restricted by the applied tools.

2. Background and Technology

Figure 1 shows the ontology related to energy saving information, which mainly defines the layer relationships and features related to the basic knowledge and concepts of various energy saving equipment, and supports the overall operation of the back-end information agent system. Reference was made to Protégé (http://protege.stanford.edu/) when the basic architecture of cloud environment energy information ontology at St John’s University was built. The related ontology services, including the semantic distance conversion of search terms, and conversion of the hypernyms, hyponyms, synonyms, as well as antonyms of the corresponding search terms were implemented. Then, WordNet (http://wordnet.princeton.edu/) is used as the basis of pattern comparison, while Jaccard similarity [10,11] was introduced to estimate the consistency of the ontological concepts. The basic approach was use of the consistency between the concepts of the search terms and the corresponding concepts of WordNet, as well as their related positions, to index the domain concepts. Finally, the identifier Synset_ID in WordNet is used as the basis to access the domain concepts and support the overall system operation. Most importantly, users can use SQL and JWNL (Java WordNet Library, http://jwordnet.sourceforge.net/handbook.html) to access the WordNet database, which is why the SQL

![Figure 1. Part of the energy information ontology.](image-url)
database was used in this study to construct the ontology and Java was used to develop the system. The hasURI and hasConsistency attributes of this representative domain concept were used to index its position in the WordNet ontology and consistency of the domain concept was calculated using the Jaccard similarity which was then stored in the corresponding hasConsistency attribute to complete the corresponding processing. Finally, the Synset_ID in WordNet was used to access the position distance of the domain concept in the Ontological Databases (OD) and to support the overall system operation. This is the operational foundation of support for subsequent time-series analysis by Big Data ontological index technology.

This study also proposes that the parallel reduction mechanism [10], as based on Big Data analysis technology, is divided into four major steps: (1) preprocess output operations for individual websites corresponding to keyword sets; (2) use of domain ontology and Jaccard dissimilarity to obtain the map operations of three keyword sets that represent individual websites; (3) organizing the shuffle operations corresponding to the three optimal keyword sets of individual websites; (4) application of the average of Jaccard dissimilarity to output the reduced operations of the three closest corresponding keywords to the user query. The process is shown in Figure 2.

Based on the above and related literature reviews, this study explores the realization of ‘R + Hadoop = Big Data Analytics’ [12] in the open source framework of Hadoop (such as Dropbox), builds the MapReduce parallel reduction mechanism, integrates a Jaccard dissimilarity calculation of keywords supported by the domain ontological service, and supports the various information services of WIAS, as based on the Big Data analysis technology. The actual operation process is illustrated in Figure 3, where OutputFormat is used to construct the OD of real data relationships according to Jaccard dissimilarity in the domain ontological service index. Thus the system realizes the operational foundation of supporting the subsequent time-series analysis technology using the Big Data analysis technology of the domain ontological index.

Figure 4 shows the complete architecture of the WIAS system [5]. User information requests are sent through various network channels to the back-end information system via Ubi-IA [13], which acts as a control center that saves energy conservation information solutions through the Solution Finder. Ubi-IA is responsible for providing the processing and conversion of cloud query information, as well as intelligent query decision-making. At the start of system operation, the cloud information ontology is constructed by domain experts and matched with corresponding default rules. The useful information of word frequency in response to the query information is then collected, and the support and confidence of the Prediction Rules corresponding to the query information are initialized. At the same time, similarity calculations between relevant cases, as supported by WIAS web service corresponding to the cloud information ontology, is done to initialize operation of the system. The system makes periodic requests in response to query information, and collects the most frequent and most infrequent queries using time-series analysis techniques.

![Figure 2. Concept of proposed parallel reduction mechanism.](image-url)
OntoCBRA generates relevant case information, along with the two-stage time-series prediction algorithm, to trigger OntoDMA to revise to the corresponding Prediction Rules. With the support of domain ontology, it also compares and extracts the appropriate corresponding information to effectively increase the quality of cloud information consultation and sharing. This enhances the accuracy, authenticity, and completeness of the provided information. Should neither of the aforementioned two processes provide appropriate cloud information solutions, the system will trigger OntoIAS [11] to complete the information search, classification, and presentation (or sorting). Other technologies can also be employed, including preprocess, map, shuffle, and reduce operations supported by the parallel reduction mechanism. Appropriate cloud information solutions can even be sought from outside the Internet. The default rules can be updated through field experts, and the learning cycle can be fully reconstructed in response to the query information. Finally, the system will find an optimal energy saving information solution through the three-phase intelligent decision-making architecture, OntoDMA, OntoCBRA, and OntoIAS. In this paper focus is on the research
and development achievements of OntoDMA under the support of WIAS.

### 3. Proposed System Architecture

Figure 5 illustrates the system architecture of OntoDMA [5,6]. First, from information provided by the Case Base, as constructed by OntoCBRA, and with the support of the aforementioned system ontology, as based on Information Entropy (such as ID3, C4.5, and C5.0), the Rule Maker calculates the relevant Object-Action Pairs. Construction of the appropriate Prediction Rules is then based on the related semantic position distance in this ontology. When the source information determines a prediction solution, it is usually abnormal energy saving information from infrequent or non-existent query information, referring to the Infrequently Historical Information from WIAS. The Prediction Monitor accordingly outputs the corresponding prediction solution through the Prediction Rules to the Solutions Pool, and provides the corresponding prediction to the interface agent system Ubi-IA according to the system threshold. If the solution is successful, it becomes learning material for the case-based reasoning agent system OntoCBRA, and is gradually reflected in the Case Base. It is then used to provide the corresponding Case Information [8]. The Prediction Rules are revised according to the response to further enhance the prediction robustness of the agent system OntoDMA.

The Rule Maker mainly serves to generate Prediction Rules, which provides the system with a knowledge processing core that can process abnormal energy saving query information online. The construction of Prediction Rules is based on system ontology and related ontological services, and generated during periods when the system is offline. The material for rule generation comes from the Frequently Historical Information, which is periodically provided by the Data Monitor in WIAS [7], and based on the case generated by OntoCBRA, this triggers OntoDMA which then refers to the web service DM_TransCaseToPred provided by WIAS. The generation of the Prediction Rules for case variations in Case Bases is done offline with the support of system ontological indexing. The Case accessed by Case Base is constructed by a set of common energy saving query scenarios corresponding to the query type and specific space in a certain period. The Prediction Rules are defined as: ‘Behavior patterns that address the operation of infrequent or non-existent information in specific periods, locations, and query types’ (Prediction Rules can be viewed using DM_ViewPrediction, the web service provided by WIAS). The Rule Maker processes comparisons of the common energy saving inquiries during the same period on different days to give the most and least frequent energy saving inquiries. These two values are central to the construction of the Prediction Rules. The design philosophy of rule construction lies in values other than the most frequent query (beyond the range), which is an abnormal phenomenon of the inquiry range. In other words, OntoCBRA processes normal inquiries within the frequent energy saving query range; however, OntoDMA processes the abnormal
inquiries outside the frequent energy saving query range, to process all the variations of all the energy saving information inquiries. However, the output material of the prediction rules comes mainly from case variations, so if the rule output time period is shorter than, or equal to, the case output time period, there will be too little data for analysis and the prediction rules output will be inaccurate. Therefore, the rule output time period must be longer than the output time period to ensure accuracy of the Prediction Rules. The two-phase process for outputting Prediction Rules based on the source of case variation is detailed in Equations (1) and (2) [10].

\[
P(G_1) = \{C_1(V_{\text{DOWN}}), C_1(V_{\text{up}})\} \\
P(G_2) = \{C_2(V_{\text{DOWN}}), C_2(V_{\text{up}})\} \\
P(G_3) = \{C_3(V_{\text{DOWN}}), C_3(V_{\text{up}})\} \\
\vdots \\
P(G_n) = \{C_n(V_{\text{DOWN}}), C_n(V_{\text{up}})\} \\
\]

\[
\text{Sorting by the value of } V_{\text{DOWN}} \text{ and } V_{\text{up}}, \text{ and to be the set of } SV_i \\
P(SV_i) = \{SVi(V_{\text{MAX}}) = \text{Max}(\sum_{i=1}^{n} SV_i), \\
SVi(V_{\text{MIN}}) = \text{Min}(\sum_{i=1}^{n} SV_i)\}, \\
P_{\text{Ans1}}(SV_1) = \{SV_1(V_{\text{MAX}}), SV_1(V_{\text{MIN}})\} \\
P_{\text{Ans2}}(SV_2) = \{SV_2(V_{\text{MAX}}), SV_2(V_{\text{MIN}})\} \\
P_{\text{Ans3}}(SV_3) = \{SV_3(V_{\text{MAX}}), SV_3(V_{\text{MIN}})\} \\
\vdots \\
P_{\text{AnsN}}(SV_N) = \{SV_N(V_{\text{MAX}}), SV_N(V_{\text{MIN}})\} \\
\]

\[
\text{Sorting by the value of } V_{\text{MAX}} \text{ and } V_{\text{MIN}}, \text{ and to be the set of } SSV_i \\
P(SSV_i) = \{SSV_i(V_{\text{MAX}}) = \text{Max}(\sum_{i=1}^{n} SSV_i), \\
SSV_i(V_{\text{MIN}}) = \text{Min}(\sum_{i=1}^{n} SSV_i)\} \\
\]

4. System Presentation and Verification

4.1. System Presentation

The main operation of OntoDMA is divided into two parts: The Rule Maker is responsible for the automatic offline output of Prediction Rules and monitoring the conversion of Case Base into Prediction Rules. The Prediction Monitor is responsible for online simulation of the interaction with OntoDMA in the three-phase intelligent operation of Decision Maker in Ubi-IA as well as support of the decision-making process. The DMA conducts online operations based on Date, Area Type, Area Size, Sensor Type, and Value of raw energy saving data, as input from the interface agent system IA and through the operation process queried by the DM_Solutions of the WIAS web service. These include input query range, query prediction rule, temporary query prediction rule, sorting, and valuing and judgment results, as shown in Figures 6–9.

4.2. Experiment and Verification

First, classroom E108A in the Electrical Engineering and Information Building at St John’s University was searched. More than 68,000 monitored energy saving records from March to June 2012 were collected [6]. These were converted into semantic format and stored in the system database. The first experiment focused on the performance of OntoDMA using all the previously recorded data as a training database. Table 1 shows results of five experiments, as reviewed by domain experts, and solved by the returned information. The table shows that in 175.6 queries, the Prediction Rules reached 43.6 queries (about 22.96%) on average; and the Solutions Pool of the system could process up to 132 queries (about 77.04%).

The second experiment focused on the learning performance of OntoDMA. This used the same Solutions Pool and Prediction Rules as in the first experiment. More than 34,000 ‘new’ monitoring records from June to September 2012 were collected and converted into internal system record format before being stored in the system database. The number of new cases generated by the system case base were gradually reduced after each experiment, and replaced by the Solutions Pool of the system. Table 2 shows the results of the five experiments reviewed by domain experts and solved by returned information. The table shows that quick cache technology can process an average of 84.68% ((77.04% + 92.31%)/2) of the data queries through the Solutions Pool of the system. To summarize: in the two OntoDMA experiments, the Solutions Pool could increase query performance by 19.82% ((92.31% – 77.04%)/77.04%) on average while the Prediction Rules a reduction of 66.51% ((22.96% – 7.69%)/22.96%) query performance on average. The performance of the Solutions Pool was improved and the query performance of the Prediction Rules was reduced. This demonstrates an effective reduction of the workload on the back-end servo system, and a great improvement in performance of OntoDMA with new material.

Table 1

| Experiment | Query Performance (%) | Average Queries |
|------------|-----------------------|-----------------|
| 1          | 22.96%                | 43.6            |
| 2          | 77.04%                | 132             |

Table 2

| Experiment | New Case Base Reduction | Average Queries |
|------------|-------------------------|-----------------|
| 1          | 50%                     | 132             |
| 2          | 20%                     | 132             |

The table shows the results of the five experiments reviewed by domain experts and solved by returned information. The table shows that quick cache technology can process an average of 84.68% ((77.04% + 92.31%)/2) of the data queries through the Solutions Pool of the system. To summarize: in the two OntoDMA experiments, the Solutions Pool could increase query performance by 19.82% ((92.31% – 77.04%)/77.04%) on average while the Prediction Rules a reduction of 66.51% ((22.96% – 7.69%)/22.96%) query performance on average. The performance of the Solutions Pool was improved and the query performance of the Prediction Rules was reduced. This demonstrates an effective reduction of the workload on the back-end servo system, and a great improvement in performance of OntoDMA with new material.
Online process:
Step 1: Enter the interval of the query; Step 2: Query prediction rules; Step 3: Query prediction rules and sort their values; Step 4: Get value and judge the result

Figure 6. Query interface of online DMA prediction rules.

Online process:
Step 1: Enter the interval of the query; Step 2: Query prediction rules; Step 3: Query prediction rules and sort their values; Step 4: Get value and judge the result

Figure 7. Qualified prediction rules queried by online DMA prediction rules.

Online process:
Step 1: Enter the interval of the query; Step 2: Query prediction rules; Step 3: Query prediction rules and sort their values; Step 4: Get value and judge the result

Figure 8. Temp prediction rules and sorting of online DMA prediction rules.
5. Conclusions and Discussions

In this paper a description of the development of OntoDMA, a cloud data mining and energy saving information agent system based on web services, Big Data analysis, and ontological support, has been presented. In addition, details of the related R&D technologies and results are also presented as well as details of the preliminary system development interface and experimental verification. The quick cache mechanism in the Solutions Pool of this OntoDMA system can increase the average system operation performance by 19.82% while the Prediction Rules of the system allow a reduction of the workload by 66.51% on average. In other words, the quick cache efficiency of the system Solutions Pool is improved, while the query workload of the Prediction Rules was reduced, which effectively alleviates the back-end servo system workload.

In practice, this study builds a cloud operating environment, as based on web services and Big Data analysis technology, and pragmatically refers to the agent technology in the development of an intelligent energy saving information processing and decision-making support multi-agent system. This can effectively implement the practical application of OntoIAS and the information agent shell through ontology technology, and integrate data mining to achieve the purpose of learning and system application amplification. This system offers economic benefit and has a prospect of broad application by the effective implementation of continuity of related research and development results. The capabilities, architecture, and technology of the proposed system are related to the application and integration of data mining technology in the field of artificial intelligence, and involve the related technologies of intelligent environment applications. Consequently, it is unique and innovative for related industries and national development. The authors hope to examine the pragmatic perspectives of technical and vocational colleges, and make measurable academic contributions in practical application and technology integration.

Acknowledgments

The authors wish to thank Ming-Yu Tsai and Guo-Jui Wu for their assistance in earlier system implementation and preliminary experiments. The authors are deeply indebted to the Department of Information and Communication Engineering and Department of Electrical Engineering, St.
John’s University, Taiwan, for all aspects of assistance provided.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**Funding**

This project was sponsored under grant 106-2221-E-129-008 and 107-2632-E-129-001 from the Ministry of Science and Technology, Taiwan, ROC.

**References**

[1] Syu SW Mobilized App “Parent-Teacher Ark”: exploring the potential of teacher-family communication and educational data mining [Doctoral Dissertation]. Taoyuan (Taiwan): Graduate Institute of Network Learning Technology, National Central University; 2015.

[2] Hsu CH Pharmaceutical companies use data mining technology to enhance health care drug sales [Master Thesis]. Taichung (Taiwan): Dept. of Information Management, Chaoyang University of Technology; 2016.

[3] Lausch A, Schmidt A, Tischendorf L. Data mining and linked open data – new perspectives for data analysis in environmental research. Ecol Modell. 2015;295:5–17.

[4] Vaish P, Srivastava S Mining big data: study of tools of open source revolution. Proceedings of 4th International Conference on Recent Innovations in Science Engineering and Management; 2016 Mar 20; Lodhi Estate, New Delhi, India, pp. 479–486.

[5] Yang SY. A novel cloud information agent system with web service techniques: example of an energy-saving multi-agent system. Expert Syst Appl. 2013;40(5):1758–1785.

[6] Wu GJ A study and application on a web service-supported information agent system – example on an energy-saving information system [Master Thesis]. New Taipei City (Taiwan): Dept. of Computer and Communication Engineering, St. John’s University; 2013.

[7] Yang SY, Wu GJ Cloud smart energy saving data mining agent system. Proceedings of 2013 National Workshop on Information and Communication Technology; 2013 Apr 24, Taipei, Taiwan, pp. 1–6.

[8] Yang SY. Developing an energy-saving and case-based reasoning information agent with web service and ontology techniques. Expert Syst Appl. 2013;40(9):3351–3369.

[9] Yang SY, Lee DL, Deng LY A web services-based cloud interface agent for intelligent YOHO information processing. Proceedings of 2013 International Conference on IT Convergence and Security; 2013 Dec 16–18, Macau, China, pp. 310–313.

[10] Chen WC, Chen WH, Yang SY. A big data and time series analysis technology-based multi-agent system for smart tourism. Appl Sci Open Access J. 2018;8(6):947.

[11] Yang SY. OntoIAS: an ontology-supported information agent shell for ubiquitous services. Expert Syst Appl. 2011;38(6):7803–7816.

[12] Chiu JC The application on big data analysis of R and Hadoop [Master Thesis]. New Taipei City (Taiwan): Dept. of Statistics and Information Science, Fu Jen Catholic University; 2014.

[13] Chen KY, Chen FH, Yang SY. An upgraded intelligent mobile information multi-agent system with universal application interfaces based on open data of Taiwan government. J Comput Appl Sci Educ. 2017;4(4):1–23.