

Knowledge Fusion Based on Cloud Computing Environment for Long-Term Care

Kai-Xiang Zhuang, National Formosa University, Taiwan

I-Ching Hsu, National Formosa University, Taiwan

ABSTRACT

Globally, aging is now a societal trend and challenge in many developed and developing countries. A key medical strategy that a fast-paced aging society must consider is the provision of quality long-term care (LTC) services. Even so, the lack of LTC caregivers is a persistent global problem. Herein, attention is called to the increasing need for identifying appropriate LTC caregivers and delivering client-specific LTC services to the elderly via emerging and integrative technologies. This paper argues for the use of an intelligent cloud computing long-term care platform (ICCLCP) that integrates statistical analysis, machine learning, and Semantic Web technologies into a cloud-computing environment to facilitate LTC services delivery. The Term frequency-inverse document frequency is a numerical statistic adopted to automatically assess the professionalism of each LTC caregiver's services. The machine learning method adopts naïve Bayes classifier to estimate the LTC services needed for the elderly. These two items of LTC information are integrated with the Semantic Web to provide an intelligent LTC framework. The deployed ICCLCP will then aid the elderly in the recommendation of LTC caregivers, thereby making the best use of available resources for LTC services.

KEYWORDS

Cloud Computing, LTC, Machine Learning, Semantic Web

INTRODUCTION

In many aging societies, long-term care (LTC) is now a strategic policy challenge (Rezgale-Straidoma & Rasnaca, 2016). Conceivably, the government of any country needs to integrate a variety of finite resources and talents in order to provide quality LTC for the elderly. The lack of human resources, such as LTC caregivers, is a common problem. The main focus of this research is to address how emerging information technologies (IT) may be leveraged and integrated to overcome the aforementioned restrictions faced in LTC servicing. Recently, many studies (Orciuoli & Parente, 2017; Bokhari, Shallal, & Tamandani, 2018; Mansouri, Alti, Roose, & Laborie, 2018; Portugal, Alencar, & Cowan, 2018) have used statistical methods, machine learning or semantic web technologies to improve the intelligence of information systems (IS), including those servicing LTC.

DOI: 10.4018/IJHISI.2020100103

This article, published as an Open Access article on January 29, 2021 in the gold Open Access journal, International Journal of Healthcare Information Systems and Informatics (converted to gold Open Access January 1, 2021), is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

A common thread among today's technological applications is the need to process big data. Today, cloud computing (CC) is deployed to handle massive data entries while improving the intelligent performance of database systems, including IS/IT targeting elderly care and LTC services (Verma, Mansuri, & Jain, 2016; Ramirez-Gallego et al., 2018). One of the most common IS architecture is the client-server model (Glushan, Lavrik, & Rybalchenko, 2015). Here, on the one end is the client who requires the service(s), and on the other end is the server that provides the servicing. This client-server model can also be conveniently applied in the LTC field. The client may be regarded as an elderly patient, whereas the server parallels a caregiver. Just as the caregiver provides the LTC service(s) to the elderly patient, the server provides the service(s) to the clients. Put simply, the technologies used in the client-server model to improve the efficiency of the client-server can be framed logically within the LTC field.

This paper argues for a novel Knowledge Fusion in Cloud Computing Framework (KFCCF) to integrate statistical methods, machine learning and semantic web technologies in CC environment to promote intelligence and facilitate big data processing for various application domains. Here, drawing on the KFCCF capabilities, the study evolves an Intelligent Cloud Computing Long-term Care Platform (ICCLCP) to recommend suitable caregivers in LTC to the elderly, thereby making the best use of available LTC resources. More specifically, the methodology integrates Term Frequency-Inverse Document Frequency (TF-IDF) statistical analysis with Semantic Web technologies to automatically assess the professionalism of each caregiver's LTC services. Additionally, it integrates Naïve Bayes Classifier (NBC) machine learning approach with Semantic Web technologies to estimate the LTC services needed for the elderly. The TF-IDF statistical analysis and NBC machine learning approach are time-consuming and computationally intensive. The suggested approach deploys the CC architecture to support parallel computing to deal with the big data and complex computing.

The remainder of this paper is organized as follows. Section 2 presents some related studies from the extant literature on big data analytics. Section 3 presents a novel Knowledge Fusion in Cloud Computing Framework (KFCCF). Section 4 developed an Intelligent Cloud Computing Long-term Care Platform (ICCLCP) based on the proposed KFCCF. In Section 5, the proposed system performance is evaluated and experimental results of the study reported. Finally, Section 6 closes the paper with a review of different data analytic capabilities when applied in managing the cluster testing with different Spark environments and provides future insights and research ideas for choosing among optimal analytical systems for LTC servicing for the increasingly aging populations.

RELATED WORK

Apache Hadoop (Apache, 2017b) is an open-source platform capable of effectively processing and storing big data. It features a distributed computing architecture and uses Hadoop Common primarily to provide the library needed by other modules. The core technologies of Hadoop consist of Hadoop Distributed File System (HDFS) and Map Reduce components. The Hadoop YARN (Rubiolo, Caliusco, Stegmayer, Coronel, & Fabrizi, 2012) cluster manager is imported into Hadoop 2 version.

Apache Spark (Apache, 2017a) is also an open-source platform for fast cluster computing. The major characteristic is that it uses in-memory technique, and finally stores the computing results in the hard disk so as to reduce the input-output (I/O) time in the hard disk. Spark architecture consists of the cluster manager, core and internal library. The Spark core includes the basic functions such as scheduling and memory management, as well as the API of the major Resilient Distributed Datasets (RDD). The cluster manager is a Standalone Scheduler built into Spark. The internal library provides Spark SQL, Spark Streaming, MLlib and GraphX. Some researchers (e.g., Domann & Lommatzsch, 2017; Yang, Ning, Cai, Liang, & Liu, 2018) focus on using Spark to improve computing performance. The distributed resource manager Apache Mesos (Foundation) is a distributed system core. Mesos considers the multi-node resources as a single high-performance computer, including the CPU,

memory, hard disk and other computing resources on the node. It shares the available resources on the node among different applications without mutual interference.

Term Frequency–Inverse Document Frequency (TF-IDF), which will be featured in this proposed study, is a weighting technique used for data searching or text mining (Akilan, 2015). It is essentially a complex statistical method. As shown in Equation (1), it is used to calculate the importance of Words in Document or Corpus.

$$tfidf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \times \log \frac{|D|}{|\{j : t_i \in d_j\}|} \quad (1)$$

$n_{i,j}$ is the times that i appears in document j , while $\sum_k n_{k,j}$ is the number of all words in document j . D is the number of documents. $|\{j : t_i \in d_j\}|$ is the number of documents that include word t_i .

The Naive Bayes Classifier (NBC) is based on the Bayesian theorem for constructing classifiers (Thulasi, Ninu, & Shiva, 2017). The main concept here is to infer, on the basis of known results, the probable occurrence of a certain reason. Typically, it is applied in the classification of data mining, which belongs to supervised learning. To classify the test data, it must calculate and multiply the posterior probability of each event, as shown in Equation (2), following which the eigenvector will be classified in the category with the highest event occurrence probability.

$$P(C | H) = \frac{\prod_{i=1}^n P(H_i | C) \times P(C)}{P(H)} \quad (2)$$

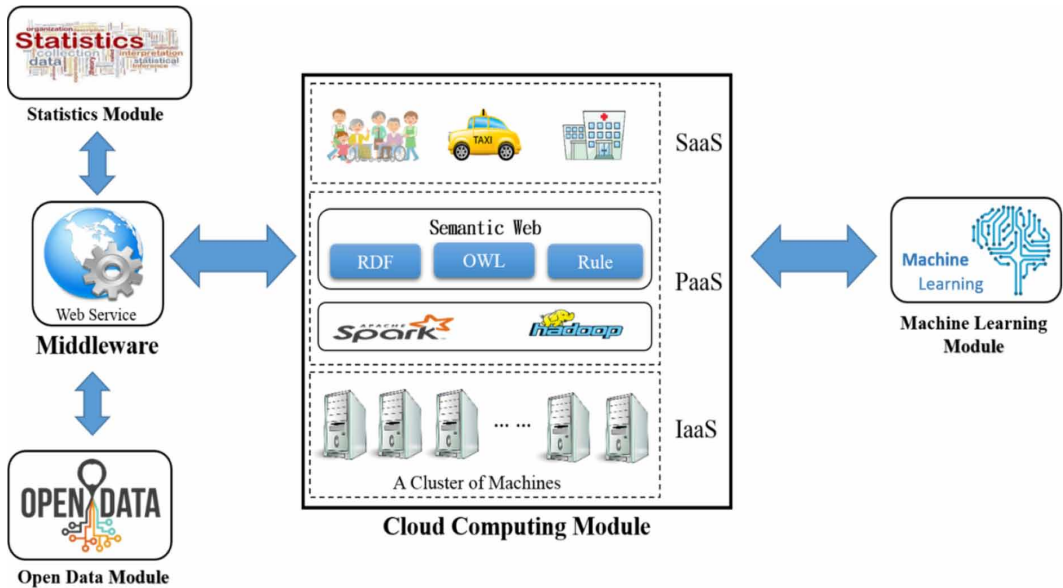
wherein, $H = \{H_1, H_2, \dots, H_n\}$ is a specific eigenvector, C is a certain category.

In term of LTC and nursing studies, Jheng-JheSie et al. (Sie et al., 2016) propose a system platform that integrates the Internet of Things (IoT), cloud, sensing network and community for use in elderly care. Through the use of real-time video surveillance, the system platform realizes activity monitoring, log, electronic nursing, and indoor positioning, so as to improve the life convenience for the elderly. Kurnianingsih et al. (Kurnianingsih, Nugroho, Widyawan, Lazuardi, & Prabuwo, 2016) adopt three supervised learning algorithms, namely, in-depth learning, K-Nearest Neighbor (K-NN) and Learning Vector Quantization (LVQ) to evaluate the health condition of the elderly with the aims of achieving prevention and reminding. Overall, the in-depth learning achieves the highest accuracy, 99.57%.

Jiayu Zhu et al. (Zhu, Min, & Wang, 2013) propose an intelligent remote medical care system. Their system builds the ontology for preliminary diagnosis and feedback for corresponding prescription so as to effectively relieve the burden for experts, especially in the level of primary health care. Available knowledge may be mined through big data analysis. In (Pramanik, Lau, Demirkan, & Azad, 2017), a big data enabled smart healthcare system framework (BSHSF) that offers a theoretical presentation of modeling in the healthcare business environment is proposed. However, there is no reference to using CC technology to improve big data processing in BSHSF.

As well, the remote healthcare system advocated by (Guan, Shao, & Wu, 2017; Zhou et al., 2018) monitors the condition of the elderly via the Internet. In (Zhou et al., 2018), a remote healthcare system using moving robots for the elderly care is envisioned. In (Gambhir, Malik, & Kumar, 2018), the authors recommend a model that adopts machine-learning technologies, including Artificial Neural Network (ANN), Decision Tree (DT) and Naive Bayes (NB), for an early detection and precise diagnosis of dengue disease.

Figure 1. KFCCF



Knowledge Fusion in Cloud Computing Framework

This study argues for a Knowledge Fusion in Cloud Computing Framework (KFCCF), which consists of Statistics Module, Machine Learning Module, Middleware Module, Open Data Model, and Cloud Computing Module, for generally intelligent CC applications. **Figure 1** displays the KFCCF, which represents the application framework at a high level of abstraction.

Statistics Module

Through the data analysis application, the system accumulates the original data and filters, and summarizes the available data, from which the correct information can be obtained. This study uses TF-IDF statistical analysis method to calculate the proficiency of the nurse aides in each service item.

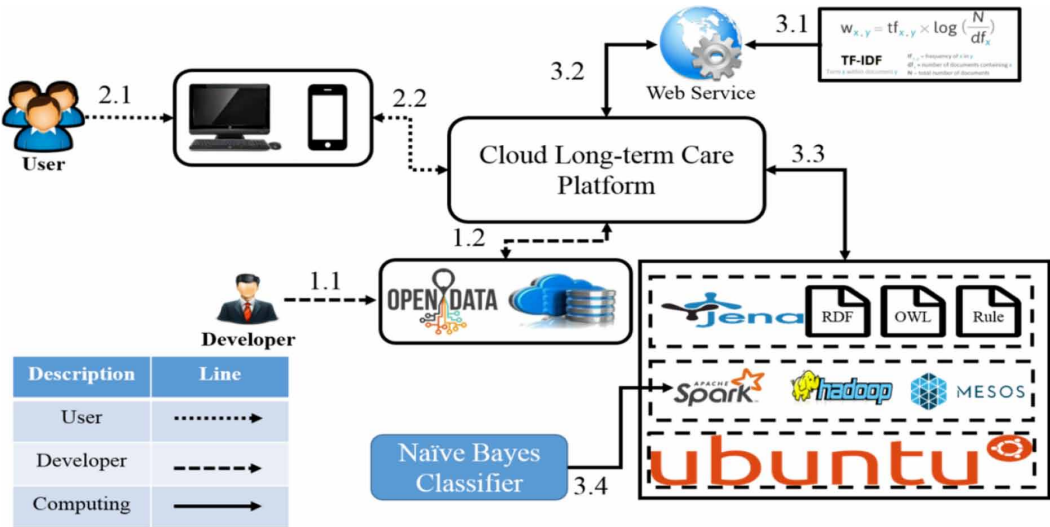
Machine Learning Module

The machine-learning module mainly adopts supervised Naïve Bayes Classifier (NBC), which is an extended application of Bayes' theorem. It follows the rationality of Bayes' theorem, that is, to infer, on the basis of known results, the probable occurrence of a certain reason. Placed in the context of the NBC model, the training data are used to estimate on the probability of the test data in each category, and then the data are classified into the correct category. The training and testing data in this study are presented in the pattern of objects so the category of each input object can be predicated by NBC.

Open Data Module

Open data is a global trend for data sharing. This module integrates the open datasets in the LTC field, for example, the geographic map of LTC provided by the Ministry of Health and Welfare¹, which is taken as the object needed by semantic inference; the Taipei City Assistive Technology Resources Center² and the three health services under the Department of Public Health, Taoyuan City³, which are taken as the recommended information of inference results obtained by machine learning.

Figure 2. Intelligent cloud computing long-term care platform



Middleware

This module is a cross-platform network service program developed according to Web Service standards, and written in PHP. It includes such services as database access, TF-IDF computing and automatic generation of objects. These services can be used to access the database, compute the professionalism of the service items provided by nurse aides, and output the service items as the category object needed by inference. Such a middleware module providing cross-platform service can then be used repeatedly, thus reducing the difficulty of system development, and further saving time and cost required for system development.

Cloud Computing (CC) Module

The bottom layer of this module includes the computer cluster Infrastructure as a Service (IaaS). The middle layer adopts Hadoop and Spark as a CC environment, and includes the machine-learning module with the NBC written in Python. The layer also has a Jena inference engine for the Semantic Web to perform CC. The Semantic Web first interprets the knowledge Ontology in the LTC field and then automatically collects large amount of objects generated in the machine-learning module to perform big data analysis and semantic inference. In this way, it recommends the LTC service items intelligently.

To validate the feasibility of KFCCF architecture proposed here, this study argues for the deployment of an Intelligent Cloud Computing Long-term Care Platform (ICCLCP) based on KFCCF, as shown in **Figure 2**. The illustration below is mainly based on data stream.

Developer

- 1.1 Import Open Data in the database of this system.
- 1.2 Display Open Data on the system platform, which can be viewed by the nurse manager, nurse aide and guardians of the cases.

User

- 2.1 The nurse manager, nurse aide and case guardian visit the system via mobile APP and webpage.
- 2.2 Mobile APP and webpage can access ICCLCP through PHP, so as to display the recommendation results.

Computing

- 3.1 The professionalism of nurse aide is calculated by TF-IDF algorithm, which is built in Web Service of this platform. It facilitates the cross-platform application of remote calling.
- 3.2 ICCLCP calls TF-IDF service of Web Service via the Internet, and searches the nurse aides under the same institution for computing so as to obtain the professionalism of the service items provided by them.
- 3.3 ICCLCP starts the Jena inference engine built in the Semantic Web of the CC Module, which will recommend the nurse aide suitable for the LTC case.
- 3.4 When performing step 3.3, it needs to call MLlib through Spark CC environment for NBC computing. It aims to calculate the target category of the object automatically. This step takes the NBC as supervised machine learning.

SYSTEM IMPLEMENTATION

Service Professionalism of Nurse Aides

This study adopts the TF-IDF algorithm in the LTC field, which is used to calculate the professionalism of service items provided by the nurse aides in an institution. tf_{ij} is calculated based on Equation (3), idf_i is calculated based on Equation (4):

$$tf_{ij} = \frac{\sum_i n_{i,j} \cdot w_i}{\sum_k n_{k,j} \cdot w_k} \quad (3)$$

$\sum_i n_{i,j} \cdot w_i$ is the time spent by nurse aide j in providing service item i for the case in the house, which is multiplied with the weight w_i rated by the case guardian for the said service item, to obtain a score. Finally, the scores obtained from all cases are summed up. $\sum_k n_{k,j} \cdot w_k$ is the time spent by nurse aide j in providing each service item for the case in the house, which is multiplied with the respective weight w_k . Then scores of each service item are totaled. Finally, the total scores obtained from all cases are combined:

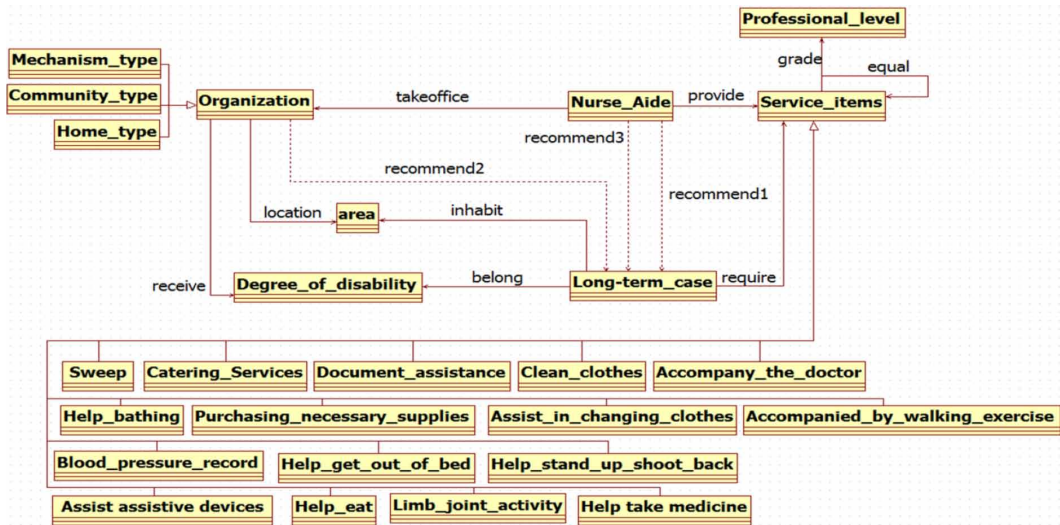
$$idf_i = \log \frac{|P|}{|\{p : t_i \in p\}|} \quad (4)$$

$|P|$ is the number of nurse aides in the same LTC institution. $|\{p : t_i \in p\}|$ is the number of nurse aides who provide service item i in the same LTC institution.

In case all nurse aides of the same institution provide the same service items, namely, $|P| = |\{p : t_i \in p\}|$, idf_i will be zero. To avoid such a situation, it is substituted by 1, as shown in Equation (5), and the professionalism of service item is only calculated based on tf_{ij} . Equation (5) also indicates longer service time will get higher professionalism:

$$idf_i = \begin{cases} 1, & |P| = |\{p : t_i \in p\}| \\ \log \frac{|P|}{|\{p : t_i \in p\}|}, & \text{Otherwise} \end{cases} \quad (5)$$

Figure 3. Ontology



Finally, tf_{ij} is multiplied with idf_i to obtain the professionalism of the nurse aide in the institution in terms of a certain service item. Moreover, the professionalism of the same institution is divided into three levels of A, B and C; A being the best, followed by B, and C, which represents the worst. Here, the professionalism of service item is taken as the object needed via semantic inference.

Analysis on Care Service Demands

The New Care Management Evaluation Scale proposed here for LTC policy not only draws on the Bayes' theorem to assess rationally on the need for case servicing items under changing circumstances vis-à-vis the ability of the case to communicate, short-term (ST) memory evaluation, the Activity of Daily Livings (ADL), and instrumental activities of Daily Living (IADL), but it also incorporates the cognitive function evaluation and rehabilitation demand and intention featured in the old scale as the data source.

More specifically, the analytic scale showcases a set H comprising four self-defined consecutive data fields, namely, diastolic blood pressure, systolic blood pressure, heart beat and blood oxygen. Each data field is also defined as discrete type. By calculating $P(B = Yes | H)$ and $P(B = No | H)$ respectively, the probabilities of set H under Yes and No situations can be ascertained. This then allows the set to be classified into the situation with higher probability to further determine if the set H is or is not required for service items.

Ontology and Rule

In **Figure 3**, OWL is used to construct the Ontology (Group, 2012). It describes the correlation among categories in the LTC field. With Jena-based rule (Hsu, 2013), the semantic inference effect in the way of if-then is accounted. Five rules are established in the study, which are used to recommend the suitable nurse aide or institution for the target case.

Figure 4 depicts Rule 1. The inference semantics of Rule 1 is explained as follows: the nurse aide (Nurse_Aide) provides care service (item), which belongs to service item (item_c). The user needs the service item (item_c). The nurse aide works in a certain institution (org), and both the user and the institution are located in the same area (area). When all of the above conditions are satisfied, ICCLCP will recommend the nurse aide for the user.

Figure 4. Rule 1

```
[rule1 : (?Nurse_Aide pre : provide ?item)(?item pre :
equal ?item_c)(?user pre : require ?item_c)
(?Nurse_Aide pre : takeoffice ?org)(?org pre :
location ?area)(?user pre : inhabit ?area)->
(?Nurse_Aide pre : recommend1 ?user)]
```

Figure 5. Rule 2

```
[rule2 : (?org pre : location ?area)(?org pre :
receive ?disability)(?user pre : inhabit ?area)
(?user pre : belong ?disability)->(?org pre :
recommend2 ?user)]
```

Rule 2 is shown in **Figure 5**. Here, the inference semantics is as follows: a LTC institution (org) located in a certain area (area) can accept the disability at some levels (disability). The user's disability level is assessed for accommodation by the LTC institution, and the user is located in the same area as the institution. When the above conditions are fully satisfied, ICCLCP will recommend the LTC institution for the user.

Rule 3, as shown in **Figure 6**, is the extension of Rule 1 condition. When the user is judged to be disabled at a certain level, including low, medium and/or high level, ICCLCP will recommend a suitable nurse aide based on the assessed disability level. Also, if the service provided by a nurse aide is of grade A, this nurse aide can then be assigned to LTC cases at any disability level.

Similarly, Rule 4, as shown in **Figure 7**, extends Rule 3. Here, if the service items the nurse aide provides are of grade B; then, the rule `notEqual(?disability,pre:serious_disability)` is used to judge whether the disability is or is not at high level. It indicates that the nurse aide can only provide service items for LTC cases with the disability not at a high level.

Figure 8 shows Rule 5 as another extension of Rule 3. Here, if the service items the nurse aide provides are of grade C; then, the rule `notEqual(?disability,pre:serious_disability)`, `notEqual(?disability,pre:medium_disability)` is used to judge whether the disability is or is not at high and/or medium levels. It indicates that the nurse aide can only provide service for LTC cases with the disability at low level.

Integrating Machine Learning and Semantic Web

Figure 9 shows how ICCLCP integrates the machine learning and semantic web in the CC environment.

Here, it shows how the ICCLCP is designed to recommend a suitable nurse aide for the LTC case. It first converts the evaluation form needed by the Naïve Bayes Classifier (NBC) into CSV format and imports the historical evaluation form into the NBC. Next, the system obtains the category

Figure 6. Rule 3

```
[rule3 : (?Nurse_Aide pre : provide ?item)(?item pre :
equal ?item_c)(?user pre : require ?item_c)
(?item pre : grade pre : A)(?user pre :
belong ?disability)(?Nurse_Aide pre :
takeoffice ?org)(?org pre : location ?area)
(?user pre : inhabit ?area)->(?Nurse_Aide pre :
recommend3 ?user)]
```


Figure 7. Rule 4

```
[rule4 : (?Nurse_Aide pre : provide ?item)
(?item pre : equal ?item_c)(?user pre :
require ?item_c)(?item pre : grade pre : B)
(?user pre : belong ?disability), notEqual
(?disability,pre : serious_disability)
(?Nurse_Aide pre : takeoffice ?org)(?org pre :
location ?area)(?user pre : inhabit ?area)->
(?Nurse_Aide pre : recommend4 ?user)]
```

Figure 8. Rule 5

```
[rule5 : (?Nurse_Aide pre : provide ?item)(?item pre :
equal ?item_c)(?user pre : require ?item_c)
(?item pre : grade pre : C)(?user pre :
belong ?disability), notEqual(?disability,pre :
serious_disability),notEqual(?disability,pre :
medium_disability) (?Nurse_Aide pre :
takeoffice ?org)(?org pre : location ?area)
(?user pre : inhabit ?area)->( ?Nurse_Aide pre :
recommend5 ?user)]
```

of each service item via CC. As noted in (a) of **Figure 9**, “1.0” indicates the need for the current service item, whereas “0.0” indicates the current service item is not needed. ICCLCP automatically generates the category object needed by Jena inference and displayed it in RDF format as captured in (b) of **Figure 9**. As depicted in (c) of **Figure 9**, TF-IDF automatically computes the service hours (also displayed in RDF format) for the nurse aide via the category object. Thus, the RDFs, OWLs and

Figure 9. Machine learning and semantic web integration architecture

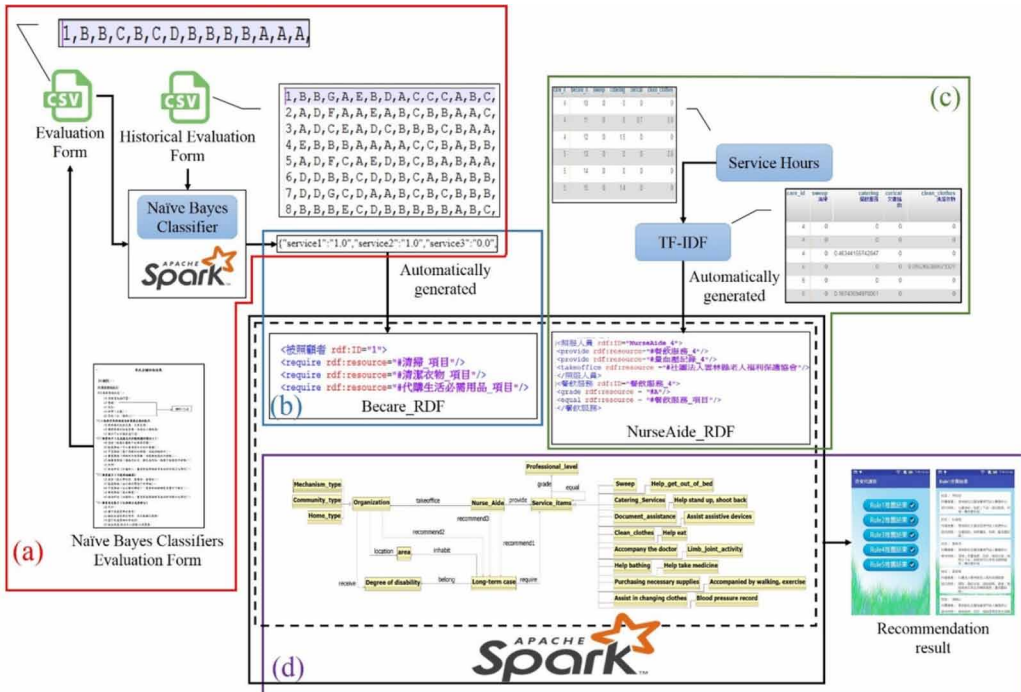


Figure 10. Service evaluation

Figure 10 consists of two mobile application screens. Screen (a) is titled '評估表' (Evaluation Form) and lists several assessment categories: '個案溝通能力' (Case Communication Ability), '短期記憶評估' (Short-term Memory Assessment), '基本日常生活活動能力 (ADL)' (Basic Activities of Daily Living), '工具性日常生活活動能力 (IADL)' (Instrumental Activities of Daily Living), '認知功能評估' (Cognitive Function Assessment), and '復健需求與意願' (Rehabilitation Needs and Willingness). Below these are input fields for '舒張壓' (Systolic Blood Pressure) with value 90, '收縮壓' (Diastolic Blood Pressure) with value 134, '心跳' (Heart Rate) with value 84, and '血氧' (Blood Oxygen) with value 98. A '送出' (Submit) button is at the bottom. Screen (b) is titled '貝氏評估結果' (Bayesian Evaluation Results) and displays a list of services with checkboxes: '清掃' (Cleaning) [checked], '餐飲服務' (Catering Service) [unchecked], '文書協助' (Document Assistance) [unchecked], '清潔衣物' (Cleaning Clothes) [checked], '陪診' (Accompanying to Clinic) [unchecked], '代購生活必需用品' (Purchasing Daily Necessaries) [checked], '協助沐浴' (Assisting with Bathing) [checked], '協助穿換衣服' (Assisting with Dressing) [checked], '協助上下床' (Assisting with Getting In/Out of Bed) [unchecked], '協助翻身、拍背' (Assisting with Turning, Back Patting) [unchecked], '協助使用日常生活輔助器具' (Assisting with Use of Daily Living Aids) [unchecked], '協助進食' (Assisting with Eating) [checked], '肢體關節活動' (Body Joint Activities) [checked], '協助服藥' (Assisting with Medication) [checked], '陪同散步、運動' (Accompanying for Walking, Exercise) [checked], and '量血壓記錄' (Blood Pressure Recording) [unchecked]. A '確認' (Confirm) button is at the bottom.

(a) Case evaluation form

(b) Bayes evaluation results

Rules are conducted with Jena inference set up on cloud. As shown in (d) of **Figure 9**, the proposed system platform will search for the suitable nurse aide and display the result on the mobile phone.

App Demonstration

In recent years, patient medication records and medical information are increasingly being accessed via mobile phones (Medhanyie et al., 2017). Accordingly, this study argues for the use of an App to help caregivers to assess the condition of the elderly.

After arriving at the case location, the care manager may import the case data quickly by scanning a QR code. Then, as shown in **Figure 10** (a), the “Service Evaluation” button can be clicked to evaluate the care service for the case. As portrayed in **Figure 10** (b), after performing the data entry procedure, the system will send the demands to ICCLCP for cloud cluster computing. This will initiate NBC computing and the results reverted back to the mobile phone user (care manager), who can then double-check if the inference is correct. Once approved (OK is pressed), the system will start Jena semantic inference. The five rules, as introduced in section 4.3 and depicted in **Figure 11** (a), will start to recommend a suitable nurse aide and institution, and the displayed results will be provided to the care manager for selection, as shown in **Figure 11** (b). The overall activity diagram is shown in **Figure 12**.

SYSTEM EXPERIMENTATION & RESULTS

In this section, the hardware of the cluster test environment being deployed is illustrated. At first, seven (7) servers are deployed in the machine room. The Spark cluster environment is built through the use of routers. The specification of each server is given in **Table 1**, for example, CPU i7-2600, 3.4 GHz, 16 GB memory, 1 TB hard disk, Ubuntu14.04 LTS edition, Hadoop 2.6, Spark 2.0 and

Figure 11. Semantic Inference results



Figure 12. Inference activity diagram

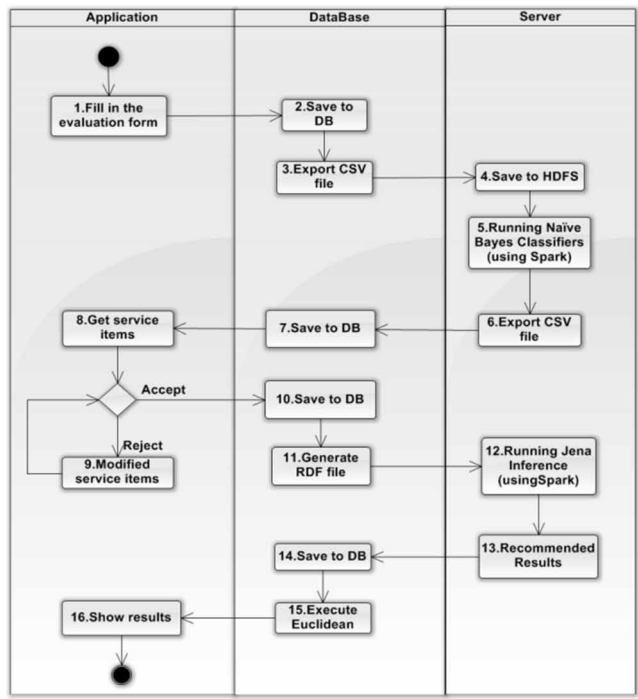


Table 1. Specifications of cluster

Host Name	CPU	Memory	HDD	OS
Master	i7-2600, 3.4 GHz	16 GB	1 TB	Ubuntu14.04LTS
data1	i7-2600, 3.4 GHz	16 GB	1 TB	Ubuntu14.04LTS
data2	i7-2600, 3.4 GHz	16 GB	1 TB	Ubuntu14.04LTS
data3	i7-2600, 3.4 GHz	16 GB	1 TB	Ubuntu14.04LTS
data4	i7-2600, 3.4 GHz	16 GB	1 TB	Ubuntu14.04LTS
data5	i7-2600, 3.4 GHz	16 GB	1 TB	Ubuntu14.04LTS
data6	i7-2600, 3.4 GHz	16 GB	1 TB	Ubuntu14.04LTS

Mesos1.3. The cluster specifications include the master as the parent node and data1~data6 as the child nodes. The architecture is shown in **Figure 13**. The study performs implementation and testing in three different cloud cluster computing environments, including Spark Standalone, Spark on YARN and Spark on Mesos.

Testing Environment 1: Spark Standalone

The study performs computing of NBC through Spark MLlib in the Spark environment, which takes the build-in Standalone as the cluster manager and HDFS as the file system. The architecture is given in **Figure 14**. This mode requires installing Hadoop and Spark on all nodes, and starting Spark and HDFS file system via commands.

Figure 13. Cluster architecture

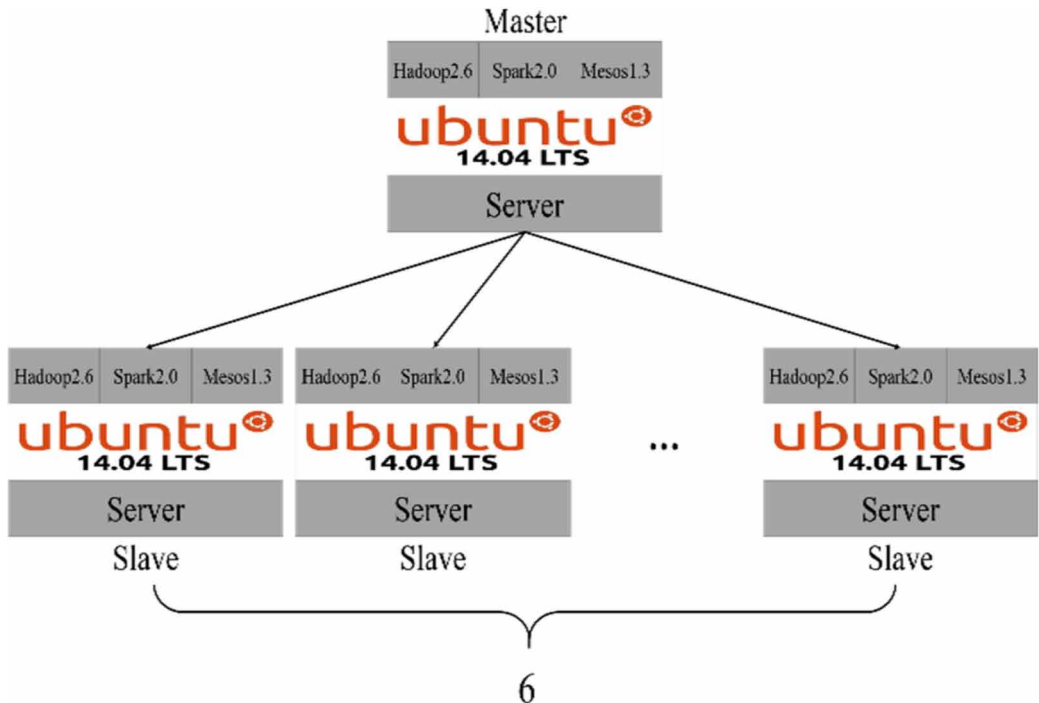
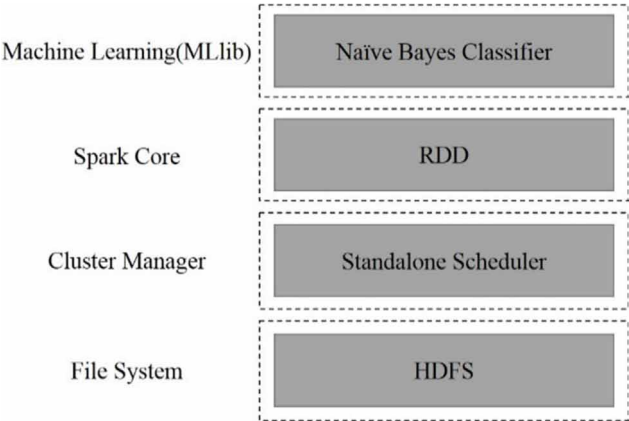


Figure 14. Spark standalone



Testing Environment 2: Spark on YARN

Under the Spark on YARN mode, the system takes Hadoop YARN as the cluster manager and HDFS as the file system. The architecture is provided in **Figure 15**. This mode requires installing both Hadoop and Spark on all nodes, and starting Spark YARN and HDFS file system via commands.

Testing Environment 3: Spark on Mesos

Under the Spark on Mesos mode, the system takes Apache Mesos as the cluster manager and HDFS as the file system. The architecture is diagramed in **Figure 16**. This mode requires installing Hadoop, Spark and Mesos on all nodes, and starting Apache Mesos, Spark and HDFS file system via commands.

Comparison of CC Cluster Environment

The test data used in this study draws mainly from the population with LTC demands, which is entered as the training data needed by NBC for module training. The size of a single file is about 1 KB, which is an evaluation scale. **Table 2** lists the data size of single service item under the different number of persons.

Figure 15. Spark on YARN

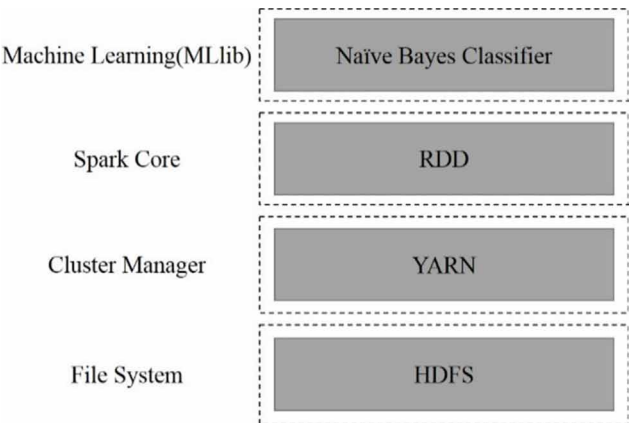


Figure 16. Spark on Mesos

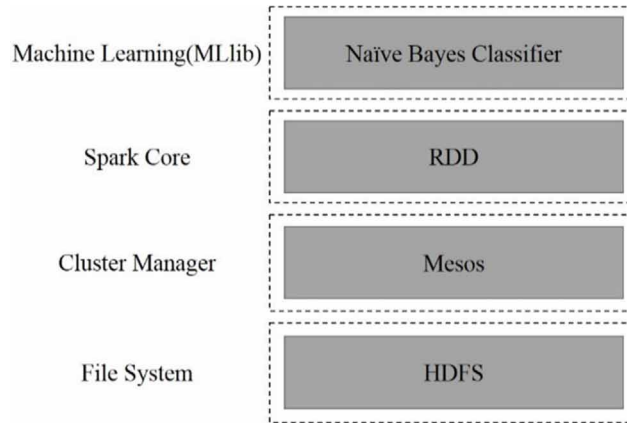


Table 2. Test data size

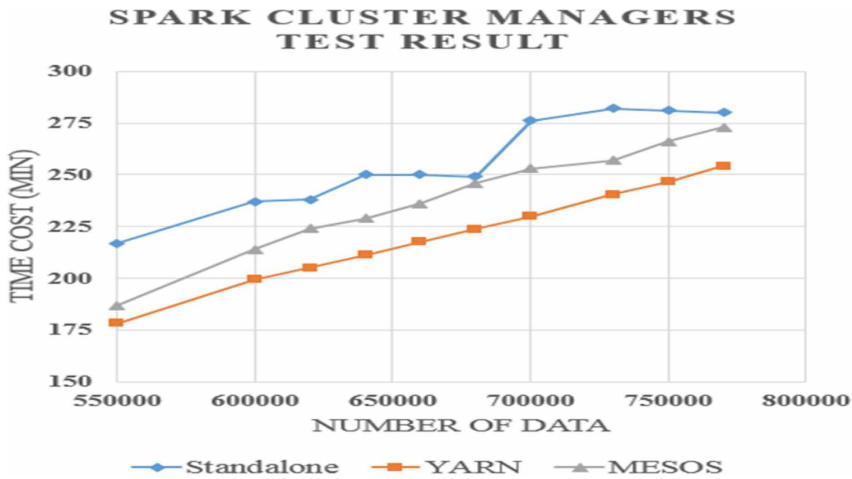
Number of Data	Single of Service Data size	All of Service Data size
550000	64.4 MB	1030.4 MB
600000	70.3 MB	1124.8 MB
620000	72.6 MB	1161.6 MB
640000	75 MB	1200 MB
660000	77.3 MB	1236.8 MB
680000	79.7 MB	1275.2 MB
700000	82 MB	1312 MB
730000	85.5 MB	1368 MB
750000	87.9 MB	1406.4 MB
770000	90.2 MB	1443.2 MB

Note that the study assumes the LTC institution to provide 16 service items in total. This study adopts an HDFS as file system, which can be set up to generate duplicates automatically, thus minimizing the risk of data loss. HDFS can also access large amounts of data over the network, which can be provided for computing in the entire cluster environment.

Figure 17 shows the integrated test results obtained from tests 1-3. Results indicate that the NBC under the YARN mode works more efficiently than either Standalone or Mesos. This is largely due to the fact that YARN can achieve better performance than Standalone in resource scheduling, which provides different schedulers for selection, such as Capacity Scheduler and Fair Scheduler. Moreover, YARN is suitable for running with large numbers of nodes and with highly complex data.

Mesos is mainly responsible for providing proper resources for the assigned task, which will be used by the original application to run executor. Hence, Mesos can be adopted to run multiple computing services. It can allocate proper sources dynamically through fine-grained mode, thereby avoiding idle allocated resources. Therefore, a CC environment Hadoop and Spark installed is suitable for performing test 1 in the development stage and test 2 in the official running stage. Finally, it

Figure 17. Spark cluster managers test results



can perform test 3 for running multiple services for long times, such as running Hadoop and Spark simultaneously.

The developed Intelligent Cloud Computing Long-term Care Platform (ICCLCP) is tested in three different cloud cluster computing environment, including Spark Standalone (test 1), Spark on YARN (test 2) and Spark on Mesos (test 3). **Table 3** shows the comparison results of these three (3) cloud-cluster computing environments. Based on the test results in **Figure 17** and **Table 3**, the authors observe the following points.

1. Experimental results showed that, for each dataset, the Spark on Mesos (test 3) has the shortest computing time. This is followed by results computed from running Spark on YARN (test 2) and Spark Standalone (test 1).

Table 3. Cluster managers compare

	Spark Standalone (test 1)	Spark on YARN (test 2)	Spark on Mesos (test 3)
Package	Spark	Apache Hadoop 2	Apache Mesos
Deployment time	Low	Middle	High
Deployment difficulty	Low	High	Middle
Availability	Low	High	High
Scheduler	FIFO	FIFO, Capacity Scheduler, Fair Scheduler	None
Master	Spark	Spark Hadoop	Spark Mesos
Slave	Spark	Hadoop	Spark Mesos
Suitable case	1.Low number of nodes 2.Low data complexity 3.Development stage	1.High number of nodes 2.High data complexity 3. Implementation stage	Run multiple interactive jobs

2. Experimental results also indicate a significant linear relationship between computing time and data size. These three tests show that the computing time is largely constrained by the data size.
3. The Spark on YARN CC environment and Spark on Mesos computing environment are suitable for handling complex data structures and computing nodes. However, Spark Standalone CC environment is suitable for processing low data structure complexity.

CONCLUSION

To cope with challenges faced by an aging population, the Taiwanese government promotes the LTC 2.0 Plan. This newer plan aims to improve the qualification restriction of the previous LTC 1.0 Plan and the low flexibility of service provision. Even so, the inadequate manpower of LTC basic service is a key issue confronted by the current Plan. This has driven us to argue for the use of a novel Knowledge Fusion in Cloud Computing Framework (KFCCF), which integrates the machine learning and semantic web in the CC environment and builds the Intelligent Cloud Computing Long-term Care Platform (ICCLCP) to validate the feasibility of KFCCF.

The proposed method quickly ascertains the professionalism of each nurse aide via the TF-IDF and Web Service and adopts the NBC to infer the care service possibly needed by the LTC cases. Moreover, it creates the Ontology and uses Rules to recommend nurse aide suitable for LTC cases. The study sets up the NBC on the Spark CC. Through memory computing, it reduces the data access time and explores the performance difference of Standalone, YARN and Mesos embedded in the system as and when applied in managing the cluster testing with different Spark environments. This would then allow the system administrator to build the optimal CC environment upon demands.

Future development will be to integrate various wearable physiological sensors (Corchia, Monti, De Benedetto, & Tarricone, 2017; Lo, Ip, & Yang, 2016) to obtain the individual status information, which can be taken as the data source. It can extend the knowledge ontology, which can be added with the professional knowledge of the physicians and nurses. Furthermore, it can extend rules to provide more individualized service, so as to increase the knowledge level. In this way, the LTC cases can obtain more complete care services. In the health care field, integrating national medical records is a kind of big data. Another research direction is to predicate individual's health and the required medical services using CC to process big data on medical records.

ACKNOWLEDGMENT

The author would like to thank the Ministry of Science and Technology, R.O.C, Taiwan, for financially supporting this research under Contract No. MOST 107-2637-E-150-009-

REFERENCES

- Akilan, A. (2015). *Text mining: Challenges and future directions*. Paper presented at the 2nd International Conference on Electronics and Communication Systems, ICECS 2015, Coimbatore, India. doi:10.1109/ECS.2015.7124872
- Apache. (2017a). *Apache Spark*. Retrieved Sep. 14, 2018, from <http://spark.apache.org/>
- Apache. (2017b). *Hadoop-Apache*. Retrieved Aug. 30, 2018, from <http://hadoop.apache.org/>
- Bokhari, M. U., Shallal, Q. M., & Tamandani, Y. K. (2018, January 1). Reducing the Required Time and Power for Data Encryption and Decryption Using K-NN Machine Learning. *Journal of the Institution of Electronics and Telecommunication Engineers*, 1–9.
- Corchia, L., Monti, G., De Benedetto, E., & Tarricone, L. (2017). Wearable antennas for remote health care monitoring systems. *International Journal of Antennas and Propagation*, 2017, 2017. doi:10.1155/2017/3012341
- Domann, J., & Lommatzsch, A. (2017). A highly available real-time news recommender based on Apache Spark. *Lecture Notes in Computer Science*, 10456, 161–172. doi:10.1007/978-3-319-65813-1_17
- Foundation, A. S. (n.d.). *Apache Mesos*. Retrieved Sep. 14, 2018, from <http://mesos.apache.org/>
- Gambhir, S., Malik, S. K., & Kumar, Y. (2018). The Diagnosis of Dengue Disease: An Evaluation of Three Machine Learning Approaches. *International Journal of Healthcare Information Systems and Informatics*, 13(3), 1–19. doi:10.4018/IJHISI.2018070101
- Glushan, V. M., Lavrik, P. V., & Rybalchenko, M. V. (2015). *Hypergraph model of hierarchical client-server architecture for distributed computing*. Paper presented at the 9th International Conference on Application of Information and Communication Technologies, AICT 2015, Rostov-on-Don, Russia.
- Group, W. C. O. W. (2012). *OWL 2 Web Ontology Language Document Overview* (2nd ed.). Retrieved Aug. 31, 2018, from <http://www.w3.org/TR/owl2-overview/>
- Guan, K., Shao, M., & Wu, S. (2017). A remote health monitoring system for the elderly based on smart home gateway. *Journal of Healthcare Engineering*, 2017, 1–9. Advance online publication. doi:10.1155/2017/5843504 PMID:29204258
- Hsu, I.-C. (2013). Integrating Ontology Technology with Folksonomies for Personalized Social Tag Recommendation. *Applied Soft Computing*, 13(8), 3745–3750. doi:10.1016/j.asoc.2013.03.004
- Kurnianingsih, N. L. E., Widyawan, L., & Prabuwo, A. S. (2016). Emergency alert prediction for elderly based on supervised learning. *Proceedings of 2016 1st International Conference on Biomedical Engineering: Empowering Biomedical Technology for Better Future*.
- Lo, B. P. L., Ip, H., & Yang, G.-Z. (2016). Transforming Health Care: Body Sensor Networks, Wearables, and the Internet of Things. *IEEE Pulse*, 7(1), 4–8. doi:10.1109/MPUL.2015.2498474 PMID:26799719
- Mansouri, K., Alti, A., Roose, P., & Laborie, S. (2018). Dynamic semantic-based green bio-inspired approach for optimizing energy and cloud services qualities. *Transactions on Emerging Telecommunications Technologies*, 29(5), e3305. doi:10.1002/ett.3305
- Medhanyie, A. A., Spigt, M., Yebo, H., Little, A., Tadesse, K., Dinant, G.-J., & Blanco, R. (2017). Quality of routine health data collected by health workers using smartphone at primary health care in Ethiopia. *International Journal of Medical Informatics*, 101, 9–14. doi:10.1016/j.ijmedinf.2017.01.016 PMID:28347452
- Orciuoli, F., & Parente, M. (2017). An ontology-driven context-aware recommender system for indoor shopping based on cellular automata. *Journal of Ambient Intelligence and Humanized Computing*, 8(6), 937–955. doi:10.1007/s12652-016-0411-2
- Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205–227. doi:10.1016/j.eswa.2017.12.020
- Pramanik, M. I., Lau, R. Y. K., Demirkan, H., & Azad, M. A. K. (2017). Smart health: Big data enabled health paradigm within smart cities. *Expert Systems with Applications*, 87, 370–383. doi:10.1016/j.eswa.2017.06.027

- Ramirez-Gallego, S., Mourino-Talin, H., Martinez-Rego, D., Bolon-Canedo, V., Benitez, J. M., Alonso-Betanzos, A., & Herrera, F. (2018). An Information Theory-Based Feature Selection Framework for Big Data under Apache Spark. *IEEE Transactions on Systems, Man, and Cybernetics. Systems*, 48(9), 1441–1453. doi:10.1109/TSMC.2017.2670926
- Rezgale-Straidoma, E., & Rasnaca, L. (2016). *Long-term elderly care: Quality assurance challenges for local governments* Paper presented at the 22nd Annual International Scientific Conference Research for Rural Development, Jelgava, Latvia.
- Rubiolo, M., Caliusco, M. L., Stegmayer, G., Coronel, M., & Fabrizi, M. G. (2012). Knowledge discovery through ontology matching: An approach based on an Artificial Neural Network model. *Information Sciences*, 194, 107–119. doi:10.1016/j.ins.2011.08.008
- Sie, J.-J., Yang, S.-C., Hong, Z.-Y., Liu, C.-K., Chen, J.-J., & Cimin, S. (2016). Integrating Cloud Computing, Internet-of-Things (IoT), and Community to Support Long-Term Care and Lost Elderly Searching. *Proceedings - 2016 International Computer Symposium*.
- Thulasi, K. S., Ninu, E. S., & Shiva, K. K. M. (2017). *Classification of diabetic patients records using Naïve Bayes classifier*. Paper presented at the 2017 - 2nd IEEE International Conference on Recent Trends in Electronics, Information and Communication Technology, Bangalore, India.
- Verma, A., Mansuri, A. H., & Jain, N. (2016). *Big Data Management Processing with Hadoop MapReduce and Spark Technology: A Comparison*. Paper presented at the Colossal Data Analysis and Networking (CDAN), Indore, India. doi:10.1109/CDAN.2016.7570891
- Yang, Y., Ning, Z., Cai, Y., Liang, P., & Liu, H. (2018). Research on parallelisation of collaborative filtering recommendation algorithm based on Spark. *International Journal of Wireless and Mobile Computing*, 14(4), 312–319. doi:10.1504/IJWMC.2018.093856
- Zhou, B., Wu, K., Lv, P., Wang, J., Chen, G., Ji, B., & Liu, S. (2018). A New Remote Health-Care System Based on Moving Robot Intended for the Elderly at Home. *Journal of Healthcare Engineering*, 2018, 1–11. Advance online publication. doi:10.1155/2018/4949863 PMID:29599949
- Zhu, J., Min, C., & Wang, F. (2013). *An ontology based intelligent telehealth system for long-term management of hypertension*. Paper presented at the International Conference on Health Informatics, ICHI 2013, Vilamoura, Portugal.

ENDNOTES

- ¹ <http://lctgis.mohw.gov.tw/Index/opendata.aspx>
- ² <https://data.gov.tw/dataset/29450>
- ³ <https://data.gov.tw/dataset/33611>

Kai-Xiang Zhuang is a master student working with Professor I-Ching Hsu in the Department of Computer Science and Information Engineering at National Formosa University of Taiwan. His current research aims at the cloud computing and machine learning.

I-Ching Hsu received a Ph.D. degree in the computer science department at the National Chung-Hsing University of Taiwan in 2007. Dr. Hsu has worked at Chung-Shan Institute of Science & Technology in the area of information engineering technologies since 1991. He is currently a professor of Computer Science and Information Engineering at National Formosa University, Taiwan. He has participated and directed projects in the area of Cloud Computing and Web Intelligence. His current research aims at the creation and study of Semantic Web Technologies, Cloud Computing, and Web Security.