

Multisource Remote Sensing Data Mining System Construction in Cloud Computing Environment

Dong YinDi¹, Liu ChengJun¹

¹Department of Information Engineering, Chongqing City Management College, Chongqing, 401331, China

²College of Electronic Information and Automation, Chongqing University of Technology, Chongqing, 400054, China

Keywords: Cloud Computing; Remote Sensing Information; Data Mining; System Construction

Abstract. In allusion to the problem that the multisource remote sensing data are not fully used and the data processing efficiency is relatively low, the multisource data mining system based on the cloud computing service mode is designed according to 3-layer 4-mode cloud computing service hierarchy system. Firstly, the technical process of the multisource remote sensing data mining is elaborated in this article; secondly, the cloud service hierarchy system is introduced into the system construction to design the multisource remote sensing data mining system framework in cloud computing environment; finally, the main characteristics and functions of the system are introduced in detail in order to lay a foundation for further researching the multisource remote sensing data mining system.

Introduction

Since the 80s of the 20th century, the remote sensing technology has been gradually concerned by people [1]. In recent over 20 years, the application research of the remote sensing technology in China has been mainly carried out under the guidance of geoscience thought, but the remote sensing application level significantly lags behind the development of the spatial remote sensing technology in the following aspect: a lot of useful information in the gradually accumulated multisource satellite remote sensing data has not yet been fully extracted and used, and the massive data reception & processing and the information reception depend on superior data processing technology and data mining system [2]. As a result, the problem regarding the design and realization of the multisource remote sensing data mining system and the data processing efficiency has become the hot issue for fully using remote sensing information.

Data Mining

As an advanced information extraction technology, data mining mainly aims at extracting hidden predictive information from large database [4]. The data mining technology can not only identify the potential mode among the data, but also extract the information and knowledge valuable for the users, thus to further guide practical application and research [5].

General data mining process

Generally speaking, the data mining process is a continuously adjusted, modified and circulated process, as shown in Fig.1. In this article, the general data mining process is divided into 4 steps:

The 1st step ---- data preprocessing: this step mainly includes such processes as data cleaning, data integration, data selection and data conversion;

The 2nd step ---- data mining: this step mainly refers to the knowledge discovery process for the preprocessed data achieved through a series of data mining algorithms, and this step is actually a key step of the data mining process;

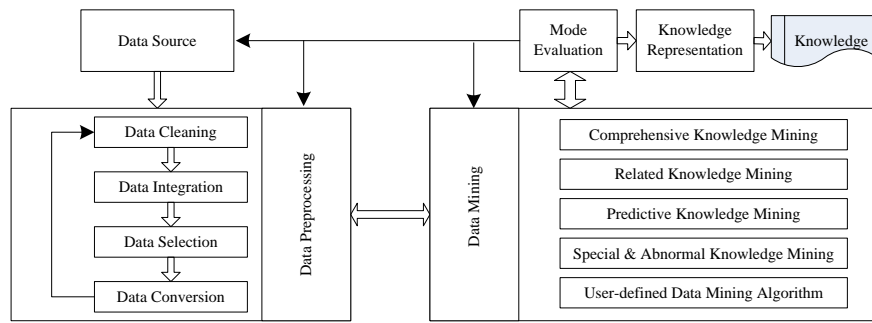


Fig.1. General data mining process

The 3rd step ---- mode evaluation: this step aims at identifying the real interesting mode of the knowledge through a certain interesting measurement;

The 4th step ---- knowledge representation: this step aims at enabling various users to observe the knowledge extracted thereby through knowledge representation technology, visualization technology, etc.

General data mining method

According to the data mining objective of the multisource remote sensing data mining system, the data mining methods are generally summarized in the article; meanwhile, for the further application of the subsequent multisource data mining system of this system, an interface - user-defined data mining algorithm is specially set for the data mining method part in order to improve the expansibility and the multifunctionality of the system.

In conclusion, the data mining methods are classified into 5 types as follows:

(1) Comprehensive knowledge mining: after preliminary data mining, the extracted comprehensive knowledge can be visually displayed for the users in a diagrammatic form through visualization technology.

(2) Related knowledge mining: this method mainly includes correlation, clustering and classification, wherein the common clustering algorithms mainly include 5 major clustering methods respectively based on division, hierarchy, density, network and model; the common classification methods include decision-making tree, neural network, genetic algorithm and evolutionary theory, Bayesian classification, support vector machine, correlation classification, analogical learning, rough set, fuzzy set, etc.

(3) Predictive knowledge mining: this method mainly includes neural network, machine learning, typical statistical method, etc.

(4) Special & abnormal knowledge mining: this method mainly aims at disclosing the special rule for the normal state deviation of objects, and includes sequential exception analysis, isolated point analysis, special rule discovery, etc.

(5) User-defined data mining algorithm: users can define the algorithm or other functional algorithms through this setting, and these user-defined algorithms can be saved in the system platform for future use, improvement, selling, etc.

Cloud Computing

The cloud computing system mainly includes three layers, namely: application layer, platform layer and infrastructure layer [1]. However, such problem as low data utilization rate universally exists in the application of the multisource satellite remote sensing data. At present, the cloud computing will provide deep data analysis and data mining for the users, and this is called RdaS in this article (namely: the remote sensing data are regarded as a service). In conclusion, the multi-mode cloud computing construction method is proposed in this article according to literature [10], and in allusion to the specific problem regarding the multisource satellite remote sensing data and the services provided thereby, 3-layer service hierarchy of the cloud computing is divided into 4 service modes, namely IaaS, PaaS, SaaS and RdaS, as shown in Fig.1.

(1) Infrastructure layer: the infrastructure is mainly regarded as a service (IaaS). IaaS is located at the bottom and functions as a standardized service to provide relevant storage and computing

ability in the network. Virtualization is the typical representative technology thereof, and the most common basic services thereof include hardware, storage, database, etc.

(2) Platform layer: the platform is regarded as a service (PaaS) for abstractly encapsulating the development environment and encapsulating the effective service carrier; in other words, the server platform or the development environment is regarded as the service to be provided to the users or the system.

(3) Application layer: the application layer mainly includes two service modes: the software is regarded as a service (SaaS) and the remote sensing data are regarded as a service (RaaS). This layer is the highest layer in the 3-layer cloud computing service system, wherein SaaS aims at providing the software to different users or systems through the Internet, and the users only need to use the webpage based software through the commercial lease mode rather than independently purchase the software, and meanwhile the users do not need to maintain the software; RaaS aims at providing the original multisource remote sensing data and the processed intermediate data to the users as a service.

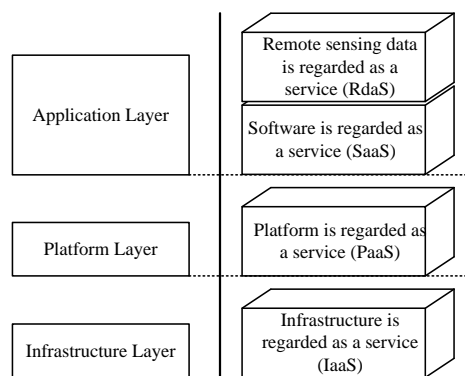


Fig.2. 3-layer 4-mode cloud computing system structure

Multisource Remote Sensing Data Mining System in Cloud Computing Environment

System structure design

According to the above analysis, a multisource remote sensing data mining system framework in cloud computing environment is designed in this article, as shown in Fig.3. This system is mainly composed of 4 subsystems, 1 database cluster and 1 server cluster. While using the cloud computing service mode, this system also takes the remote sensing data as a resource sharing service. In this way, the system can not only better realize multisource remote sensing data mining tasks, but also continuously improve and update itself through the open system, thus to improve data mining efficiency.

(1) Remote sensing data mining subsystem

This subsystem as a main part of this system aims at adopting the data mining process for the knowledge discovery in the system data and the user data so as to finish the preset data mining tasks. Specifically, this subsystem mainly includes 8 modules: problem understanding and definition, collection and extraction of multisource remote sensing data, remote sensing data purification, remote sensing data mining engine, remote sensing data algorithm engine, operating data mining algorithm, result evaluation and data & problem refinement, and result application. Additionally, the subsystem is provided with remote sensing data display interface.

(2) System account management subsystem

System account management subsystem is mainly used for recording the use condition of relevant services of the system and the incomes & expenses of other accounts for different users including management users and common users. Specifically, the management users can use and process the remote sensing data under the supervision of the subsystem in order to obtain useful information; the common users can share the remote sensing data, the data mining algorithms and various database resources through payment and meanwhile share other useful remote sensing data information so as to obtain corresponding remuneration.

(3) Data mining algorithm subsystem

This subsystem is mainly used to manage the typical algorithms and the algorithm models needed by the remote sensing data subsystem. Meanwhile, this subsystem can assist to realize the function of the “user-defined mining algorithm” module in the remote sensing subsystem.

This subsystem respectively includes the following contents: classic mining algorithm, user-defined mining algorithm, algorithm use record, algorithm selling record, algorithm sharing, help information, etc. Specifically, the classic mining algorithm refers to the classic mining algorithm set in the system and relevant functional introduction thereof; the user-defined algorithms refer to the user-defined mining algorithms based on different requirements or other improved algorithms; algorithm history records are made by different users for collecting the information regarding the use and search conditions and other relevant conditions of these algorithms in the system.

(4) Extracted data management subsystem

This subsystem is used to manage different data resources of various users, including the multisource remote sensing data obtained through various remote sensing satellites, system data uploaded by other individual users, purchased system data and sold data. The design thought for the subsystem is proposed according to DaaS service mode in the cloud computing environment: the common users can purchase the needed data from the master system through this subsystem, and also can sell their data to the master system to obtain corresponding benefits.

(5) Server cluster and database cluster

The server cluster is mainly used to provide efficient system access, data operation, lease, algorithm processing and other services to the users. The server cluster adopts virtual technology and distributed technology to integrate and share the local and remote server resources to realize efficient computing and service so as to significantly improve the ability for processing massive data. Additionally, the database cluster is mainly used to store various data resources generated from the interaction with the system, and includes all databases related to the data mining system, such as the database used for storing remote sensing data, user account database, the algorithm library used for storing data mining algorithms and other relevant databases.

Characteristic function of the system

In allusion to the specific application problem of the multisource remote sensing data processing, the 3-layer cloud computing service mode is expanded for the system designed in this article, and meanwhile the remote sensing data are taken as a resource to realize the sharing service, thus to more or less improve the utilization rate and the processing efficiency of the multisource remote sensing data and conquer some defects of the data mining system. Compared with other systems, the system proposed in this article has the following characteristics:

(1) This system has strong openness favorable for continuous updating and expansion of the system. Various users can adopt different terminal equipment to obtain the remote sensing data through the Internet and execute corresponding operations, thus to significantly increase the system openness; meanwhile, the user-defined data mining algorithm function of the system can enable the users to improve and share the algorithms provided by the system, thus to not only constantly expand algorithm type and performance, but also rapidly promote algorithm research.

(2) This system is convenient to use and operate. This system provides various help functions to the users, including detailed algorithm description, remote sensing data analysis, result evaluation, etc., and can enable different users to rapidly understand and master the remote data mining process.

Conclusion

In this article, the multisource remote sensing data resources are introduced into the cloud computing environment; through the analysis of the cloud computing structure, a 3-layer 4-mode service system is proposed for the remote sensing data mining system in order to design the multisource remote sensing data mining system in the cloud computing environment. The subsystems of this system are combined with corresponding cloud service modes to flexibly provide corresponding data mining services to various users through the loosely-coupled association mode. As a result, through data sharing and distributed data storage & processing technology, this system can more or less solve such problems as the low utilization rate and the low processing efficiency of the remote sensing data, thus to have better expansibility and openness and provide good solution for the remote sensing data mining system.

References

- [1] Yishuang Geng, Kaveh Pahlavan, On the Accuracy of RF and Image Processing Based Hybrid Localization for Wireless Capsule Endoscopy, IEEE Wireless Communications and Networking Conference (WCNC), Mar. 2015
- [2] Jie He, Yishuang Geng and Kaveh Pahlavan, Toward Accurate Human Tracking: Modelling Time-of-Arrival for Wireless Wearable Sensors in Multipath Environment, IEEE Sensor Journal, 14(11), 3996-4006, Nov. 2014
- [3] Lv, Zhihan, Liangbing Feng, Haibo Li, and Shengzhong Feng. "Hand-free motion interaction on Google Glass." In SIGGRAPH Asia 2014 Mobile Graphics and Interactive Applications, p. 21. ACM, 2014.
- [4] Zhong, Chen, Stefan Müller Arisona, Xianfeng Huang, Michael Batty, and Gerhard Schmitt. "Detecting the dynamics of urban structure through spatial network analysis." International Journal of Geographical Information Science 28, no. 11 (2014): 2178-2199.