Title	A Double Helix Architecture of Knowledge Discovery System Based Data Grid and Knowledge Grid for Multimedia Communication Networks
Author(s)	Jing, He; Wuyi, Yue; Yong, Shi
Citation	
Issue Date	2005-11
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/3890
Rights	©2005 JAIST Press
Description	The original publication is available at JAIST Press http://www.jaist.ac.jp/library/jaist-press/index.html, IFSR 2005 : Proceedings of the First World Congress of the International Federation for Systems Research : The New Roles of Systems Sciences For a Knowledge-based Society : Nov. 14-17, 2100, Kobe, Japan, Symposium 3, Session 6 : Intelligent Information Technology and Applications Knowledge Management

A Double Helix Architecture of Knowledge Discovery System Based on Data Grid and Knowledge Grid for Multimedia Communication Networks

Jing He[1,3], Wuyi Yue[2] and Yong Shi[3]

[1]Institute of Intelligent Information and Communication Technology
Konan University, Kobe 658-8501 Japan
hejing@gucas.ac.cn

[2]Department of Information Science and Systems Engineering
Konan University, Kobe 658-8501 Japan
yue@konan-u.ac.jp

[3]Chinese Academy of Sciences Research Center on Data Technology
and Knowledge Economy, Beijing 100080 China
yshi@gucas.ac.cn

ABSTRACT

Knowledge Discovery in data resources managed in Grid is a challenging research and a development issue. Grid is a computation and management infrastructure that transforms science, business, health and society. This paper presents a double helix architecture of Knowledge Discovery based on both Data Grid and Knowledge Grid. This architecture utilizes the cooperating mechanism of the data nodes in Data Grid and the knowledge nodes in Knowledge Grid. By combining with an implementation for multimedia communication networks (MCNs), the knowledge discovery system for MCNs is put forward to explain the dynamic ascending process in the knowledge discovery system.

Keywords: double helix, Data Grid, Knowledge Grid, Knowledge Discovery, multimedia communication networks

1. INTRODUCTION

Knowledge Discovery employs a variety of tools and software systems to find useful patterns, models and trends from large volumes of knowledge. In many scientific and commercial applications, it is necessary to perform the analysis of large data sets, maintain geographically distributed sites by using the computational power of distributed and parallel systems. In this area, grid technologies may play significant roles in providing an effective computational support for the knowledge discovery applications. The extension of Data Grid and Computation Grid is Knowledge Grid [1].

Knowledge Grid can synthesize knowledge, search engine reasons, answer questionsand conclude. Constructing Knowledge Grid based on Computation Grid can provide conditions for the effective knowledge discovery process [1]. Knowledge Discovery in Grid is designed on the top of computation grid mechanisms, and is provided by the grid environment tools such as known tools Globus [2] and Legion [3]. Knowledge Grid uses the basic grid services such as communication, authentication, information, and resource management to build more specific knowledge discovery tools and services [4].

Based on Data Grid and Knowledge Grid, logic and cognizing architecture of Knowledge Discovery in Grid are proposed in this paper. This architecture is the dynamic double helix system that can dynamically integrate the new data and algorithm between Data Grid and Knowledge Grid during the process of the knowledge discovery.

Reviewing on the whole development process of Knowledge Discovery, the extension of Knowledge Discovery can be divided into the following three phases. (1) Knowledge Discovery based on the structural data.

When the conception of Data Mining was put forward in 1989, the main object of Knowledge Discovery was to mine the knowledge in the database. Many data mining softwares have been applied in academies and companies. (2) Data mining based on the complex data including graph, voice, video and other temporal or spatial data. The research works on this field have just begun in the past ten years. (3) Knowledge Discovery in Grid. During the development of Knowledge Discovery, new knowledge will be mined and Grid will contain more knowledge than the present data warehouse. Then the knowledge discovery based on Grid is put forward [5].

A knowledge discovery system should at least achieve one of the following three goals [6]. (1) Many kinds of knowledge discovery tools can be used in the open environments. (2) The knowledge discovery system can support the direct transfer among different resources and supply the knowledge transfer tools. (3) The knowledge discovery system can integrate new data and algorithm during the dynamic process of the knowledge discovery system.

This paper puts forward a constructive double helix architecture of the knowledge discovery system. This architecture can adaptively control the structure and the scale of Data Grid through Knowledge Grid.

In Section 2, basic elements, the definition and the architecture of the constructive double helix are introduced. In Section 3, the knowledge discovery system specially for the quality of service (QoS) management of multimedia communication networks (MCNs) is built to give a real-time example and implementation. The conclusion is put forward in Section 4.

2. DOUBLE HELIX ARCHITECTURE OF KNOWLEDGE DISCOVERY SYSTEM

2.1. Basic element analysis

Based on correlative researches of frame synthesizing on the knowledge discovery systems [7]-[12], a knowledge discovery system is essentially a systematic frame of processing data, information and knowledge.

A knowledge grid system is classified into four units shown as follows: (1) Knowledge conservation and distilling. (2) Knowledge transformation. (3) Knowledge application. (4) Knowledge making. The whole process is regarded as a self-cycle development system with these four units shown in Fig. 1. The knowledge conservation and distilling are the basis of the knowledge discovery system. The knowledge transformation is a necessary mode. The knowledge application is a realization mode of knowledge management value. And the knowledge making is the headspring of the new knowledge.

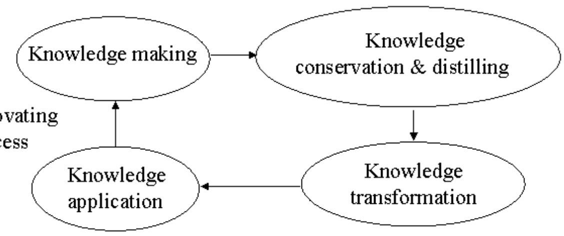

Figure 1: Self-cycle flow of knowledge grid system.

We can also get the self-cycle architecture of the data grid system shown in Fig. 2.

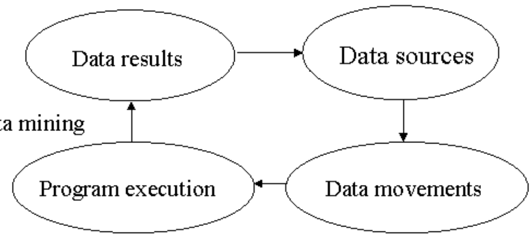

Figure 2: Self-cycle flow of data grid system.

The data grid system is classified into four units as follows: (1) Data sources. (2) Data movements. (3) Program execution. (4) Data results. The whole process is regarded as a self-cycle development system. The data sources are the beginning of the self-cycle flow. The data movements are the necessary modes. The program execution is the realization mode of data management value. The data results are the headsprings of the new data.

2.2. Basic relationship between Data Grid and Knowledge Grid

We will give out one to one corresponding relationship between data nodes in Data Grid and knowledge nodes in Knowledge Grid in Fig. 3, where [A] and [B] represent a data grid node and a knowledge grid node.

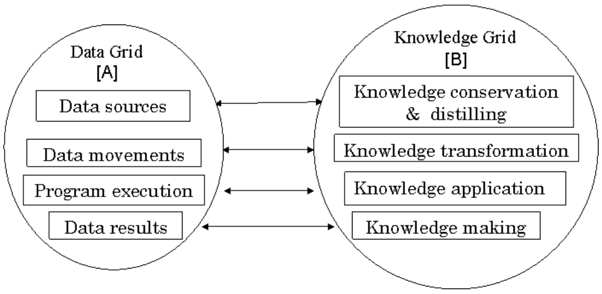

Figure 3: Structure corresponding relationship.

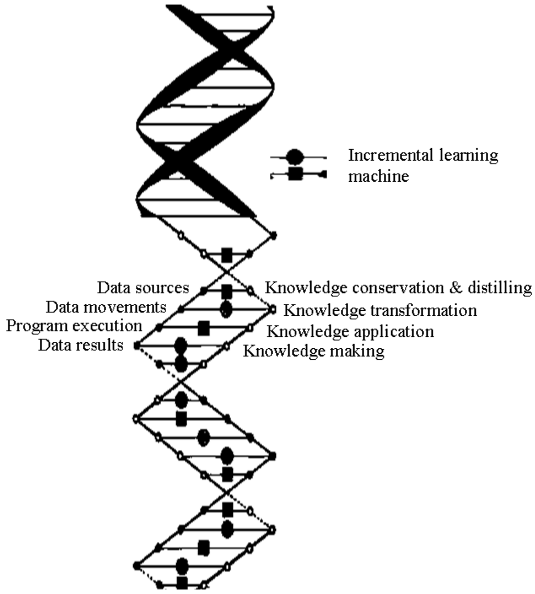

Figure 4: Double helix architecture of Knowledge Discovery based on Data Grid and Knowledge Grid.

Data Grid is also complementary for Knowledge Grid in the early stages of Knowledge Grid because it can help users to explore user's data, by focusing attention to important variables for identifying exceptions, or finding interactions. This is important because the better users understand user's data, the more effective the knowledge discovery process will be. To the domain of the knowledge discovery system, topology space of knowledge node [B] in Knowledge Grid is the same generalization space with topology space of data node [A] in Data Grid.

2.3. Double helix architecture of knowledge discovery system

The discovery of the double helix construction of deoxyribonucleic acid (DNA) was the most important biological discovery in 20th century, and it has laid the foundation for modern molecular biology successfully. The graph of DNA structure supplies good imagination of the architecture of the knowledge discovery system. The double helix architecture of the knowledge discovery system we presented in this paper is based on Data Grid and Knowledge Grid shown in Fig. 4.

The main chains of the double helix architecture presented in this paper are incremental learning machine between Data Grid and Knowledge Grid. There are two main chains convoluting the same axis in the twist shape along right hand direction. The two main chains are parallel and head for the reverse directions.

The incremental learning machine in the same surface connects the data grid node [A] in Data Grid and the knowledge node [B] in Knowledge Grid to format a twin grid. The twin nodes are combined with the corresponding parts in these two grids. These two corresponding nodes in the same surface head for the different directions. The situation results in the more interaction chances between data and knowledge.

This double helix architecture is a channel connecting Data Grid and Knowledge Grid which can use Knowledge Grid to check, supervise and drive the mining process of Data Grid. The architecture of the knowledge discovery system changes and formats the original operation mechanism to an open and optimized entity on the structure and the function.

The characters of the double helix architecture can be listed out as follows:

(1) Knowledge Grid depends on Data Grid. The new knowledge is mined from Data Grid and transferred to Knowledge Grid. The whole process is named ascending process.

(2) Data Grid depends on Knowledge Grid simultaneously. The knowledge discovered in Data Grid must be imported into Knowledge Grid. After it is validated, the knowledge is able to show the realizing meanings.

(3) A new knowledge grid depends on the original knowledge grid. The new knowledge discovered in Knowledge Grid must be imported into the original knowledge grid to check whether it is redundant, contradict or repeated.

The details of the above knowledge discovery process can be shown as follows: firstly, Knowledge Grid can acquire all the parameters and the errors of old Data Grid through learning. Then how to amend the parameters of Data Grid is specified. This kind of learning can be seen as batch learning. But in common sense, this kind of the batch learning cannot work smoothly. For example, we cannot get the enough big samples to learn. Real-time application requests the knowledge in the data grid update continuously and requests Knowledge Grid to learn from new samples on time.

Secondly, a double helix incremental learning (BHIL) is needed by the knowledge discovery system. BHIL can adjust Data Grid through analysis and recognition on new samples. The double helix architecture of Knowledge Discovery can learn new knowledge based on the original knowledge and rapidly adapt environment changes. BHIL can realize by adjusting the parameters and can increase or decrease the node of Data Grid. It changes the structure of Data Grid. Knowledge Grid determines the learning ability of Data Grid although this ability is not easily determined in advance. BHIL can avoid the quandary of the choice in Data Grid.

The adjustment of Knowledge Grid exerting on Data Grid has the basic three types: increasing, decreasing, or unite of the both. Generally increasing adjustment is easier to be worked out and to be realized than decreasing adjustment, because the decreasing adjustment is liable to the oscillation of the knowledge discovery system.

3. IMPLEMEMTATION FOR MCNs

The explosive growth of MCNs, the increased user demand for bandwidth, and the declining cost of technology have all resulted in the emergence of new classes of QoS integrated evaluation system with networks forwarding rates on the order of gigabits, or even terabits per second. So during the arranged processes of the real-time MCNs, an effective QoS centric integrated evaluation system must be designed and established.

Here, the double helix architecture of the knowledge discovery system based on Data Grid and Knowledge Grid of QoS integrated evaluation systems for MCNs is put forward. This knowledge discovery system meets the requirements of the next-generation MCNs which require structural simple, high-speed, high-operability evaluation system.

Grid frame tools such as known tools Globus [2] and Legion [3] supply the open frame and unit the resource management platform. The double helix architecture of the knowledge discovery system based on Data Grid and Knowledge Grid of QoS integrated evaluation systems for next-generation MCNs is constructed on Globus grid tools and services. These tools can unite the nodes in Data Grid and Knowledge Grid. These grid services can result in the parallel and distributed knowledge frames based on middle ware of Grid. Three layers shown in Fig. 5 organize the frame of the knowledge discovery system.

3.1. Ground layer for Data Grid

The ground layer for Data Grid includes a database, a information base, and a knowledge base of MCNs. The database and information base include real-time data monitored in MCNs. The knowledge base includes the knowledge from MCNs experts, the knowledge learned from the knowledge users and the knowledge mined during the knowledge discovery processes.

3.2. Middle layer for Knowledge Grid

The middle layer for Knowledge Grid gets knowledge from incremental learning in different structures, different technologies and different regions based on Data Grid and Computation Grid. The double helix

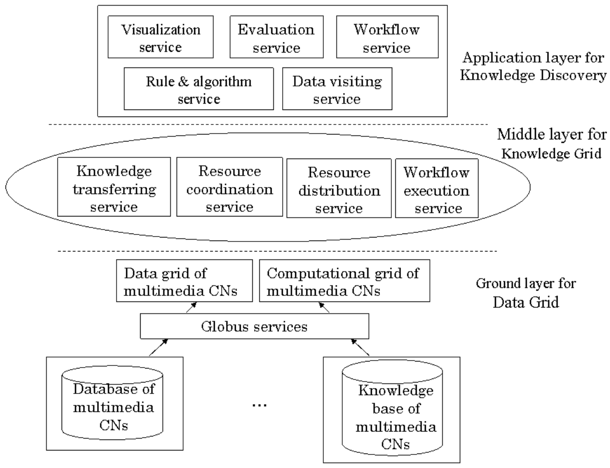

Figure 5: Frame of knowledge discovery system for MCNs.

architecture will realize the united organizing, execution, coordination of data mining tools and visualization tools.

There are four main service parts in this middle layer as follows:

(1) Knowledge transferring service. It takes the responsibility of semantic description in the grid resource including place of mainframe computer, place of mainframe monitor, source of data, tools, extracted algorithm, distributed knowledge workflow, results of mining.

(2) Resource coordination service. It administrates usage of the resources and responses the requests of the workflow. If usage of the resource is unbalanced and conflicted, this service will coordinate the mapping relationships.

(3) Resource distribution service. It determines the constraint condition and gives out the resource distribution scheme after the united analysis of workflow. This service is denoted by the abstracted semantic unit and the present resource.

(4) Workflow execution service. It executes the mapping between Data Grid and Knowledge Grid. This service will utter the invitation for the data mining process.

3.3. Application layer for Knowledge Discovery

The application layer for Knowledge Discovery includes incorporating, validating, and executing discovery service, supplying service with storing and mining knowledge. These services can be listed out as follows:

(1) Visualization service. It visualizes the extracted knowledge model such as correlation rules, colonial models and so on.

(2) Evaluation service. It compares with the QoS requests of MCNs or the original knowledge discovery models to decide the validity of the knowledge discovery system.

(3) Workflow service. It testifies the source of data, extracting tools, data mining tools, and visualization tools. It transfers relationship by graphing and defining the basic structure of resource applications and usage constraints.

(4) Rule and algorithm service. It finds, chooses and downloads data mining tools, rules and algorithms.

(5) Data visiting service. It searches, chooses, extracts, transfers and issues the data.

4. CONCLUSIONS

This paper presented a double helix architecture of Knowledge Discovery based on both Data Grid and Knowledge Grid. This architecture utilized the cooperating mechanism of the data nodes in Data Grid and the knowledge nodes in Knowledge Grid. The double helix knowledge discovery system can set up many positive-covers and negative-cover processes. The system can modify and optimize the parameters and structures of Data Grid continuously. Also the system can add the knowledge nodes acceding to the demands and prune the redundant nodes. The system not only keeps the advantages of the original knowledge discovery system, but also fits for incremental learning which could enhance the generalization capability of the knowledge discovery system. The situational results on MCNs showed that the double helix knowledge discovery system is not sensitive to the order of the samples and can learn quickly and steadily even if the performance of the original knowledge discovery system is not very good.

ACKNOWLEDGEMENTS

This work was supported in part by GRANT-IN-AID FOR SCIENTIFIC RESEARCH (No. 16560350) and MEXT.ORC (2004-2008), Japan and in part by NSFC (No. 70472074), China.

REFERENCES

[1] Z. Liu, A. Li and A. Zhu, "Review on Grid Technology," *Chinese Journal of Computers and Applications*, 40 (12): 1770-1774, 2003.

[2] http://www.globus.org/.

[3] http://www.legion.virginia.edu/.

[4] http://ecoweb.co.zw/computing/zimisis.aspEB /OL/.

[5] http://202.113.96.26/tjcbe/xueshubaogao/.

[6] J. He and G. Huang, "Logic and Cognizing Analysis of Knowledge Management: System Dynamics Approach," *Proceedings of the 5th International Symposium on Knowledge and Systems Sciences*, 392-396, 2004.

[7] M. Ackerman and C. Halverson, "Organizational Memory: process, Boundary Objects, and Trajectories," *Proceedings of the Thirty-Second Annual Hawaii International Conference on System Sciences*, 311-320, 1999.

[8] A. Tiwana, *Knowledge Management Tools*, Prentice Hall PTR, 2002.

[9] Y. Shi, "Humancasting: A Fundamental Method to Overcome Information Overload," *Information*, 12 (3): 127-143, 2000.

[10] http://www.natarch.mec.gov.na/isidaf/isidaf. log9805[EB/OL].2002-10-20/.

[11] http://si.deis.unical.it/ talia/euroglob.pdf/.

[12] H. Yu, L. Chen and W. Li, "Operations Management Faced to Service," *Research on Computers and Development*, 40 (12): 1770-1774, 2003.