

The analysis of Intelligent Agent-based data mining in the railway information integration program

Yanran Hu ⁺

School of Economics and Management, Beijing Jiaotong University, Beijing, China

Abstract. China's railway information business system are basically developing independently , mutual independence makes it difficult to exchange and share of information resources,also can't realize comprehensive and integral benefit.The railway information integration program through established the Information Integration Center ,solved the integration needs of the information of the ministry of Railways, Railways and Stations. And then established of the whole road to be an information integration platform.Using Intelligent Agent-based data mining tasks to ensure the integration of dynamic collaborative process smoothly.

Keywords: Share; Integrated application; Information integration; Intelligent Agent; Dynamic cooperation

1. Introduction

Western countries abroad faced on transportation, passenger travel services, scheduling and other established information management system as early as the 1980s, and moving toward an integrated, comprehensive development. Part of the domestic information technology industry that started earlier has begun the practice of information integration.China's railway information has nearly 30 years' construction,builed a lot of changing, distributed business management system.Such as financial management systems for financial, operational management for the management of systems.China's railway construction information developed a lot of different information subsystems based on time, the business,but these systems are serious heterogeneity.Therefore ,it need to build railway information system to provide integrated information resources sharing and ensure a reasonable distribution of information.In order to solve the information integration process in the railway about how to effectively use massive data problems,this paper proposed Intelligent Agent-based data mining system.This system can support decision-making railway management.

2. Overview of Agent-based Mining

2.1. What Is the Intelligent Agent

Current academic still have no uniform definition of Agent.Now more accepted definition is the theory proposed by the British Agent Wooldridge and Jennings:Agent is an autonomous program, which based on its understanding of the environment, have the ability to control its own decision-making behavior, thus seeking to achieve the target^[1].The Intelligent Agent is an intelligent entity on behalf of all an abstract noun.

Intelligent Agent's characteristics are as follows:1) Autonomy. Intelligent Agent is an independent entity,which has certain ability of control,can solve a given number of issues without human's direct intervention, and then respond to the surrounding environment in a particular way;2) Initiative. Intelligent Agent should have the initiative to act in a rational way of reasoning and forecasting under the direction of

⁺ Corresponding author
E-mail address: huyanran0407@163.com

the target;3) Interaction. Intelligent Agent should be able to collaborate with each other, rather than independently exist in the environment. In addition, the Intelligent Agent also has persistent, adaptive characteristics.

2.2. What Is Data Mining

In general, data mining is to extract valuable knowledge from a large number of data or to find interesting patterns. Data mining can be used for any type of information repository. Databases, data warehouse or other repository are commonly used.

Data mining process including: data preprocessing, correlation analysis, classification and clustering, etc. The Intelligent Agent-based data mining mainly uses data preprocessing. Data pre-processing is do some of the data processing before the main processing. Because the data in the world is dirty, incomplete and inconsistent in general. Pre-processing of the data can improve data quality, and thus help to improve the accuracy of the subsequent mining process and performance^[2].

2.3. Intelligent Agent Data Mining Model

Agent-based data mining model contains many different types of Agent set to achieve different functions. The commonly used types are as follows: 1) Agent which is responsible for interpersonal interaction; 2) Agent used to select data pre-processing method; 3) Agent used for data pre-processing; 4) Agent used to select data mining algorithms; 5) Agent used to achieve the specific data mining process; 6) Agent used to express the results and for post-processing. The Model can contain multiple functions with the same Agent. The structure of the mining model M and the information interaction^[3] between different types of Agent are shown in Figure 1.

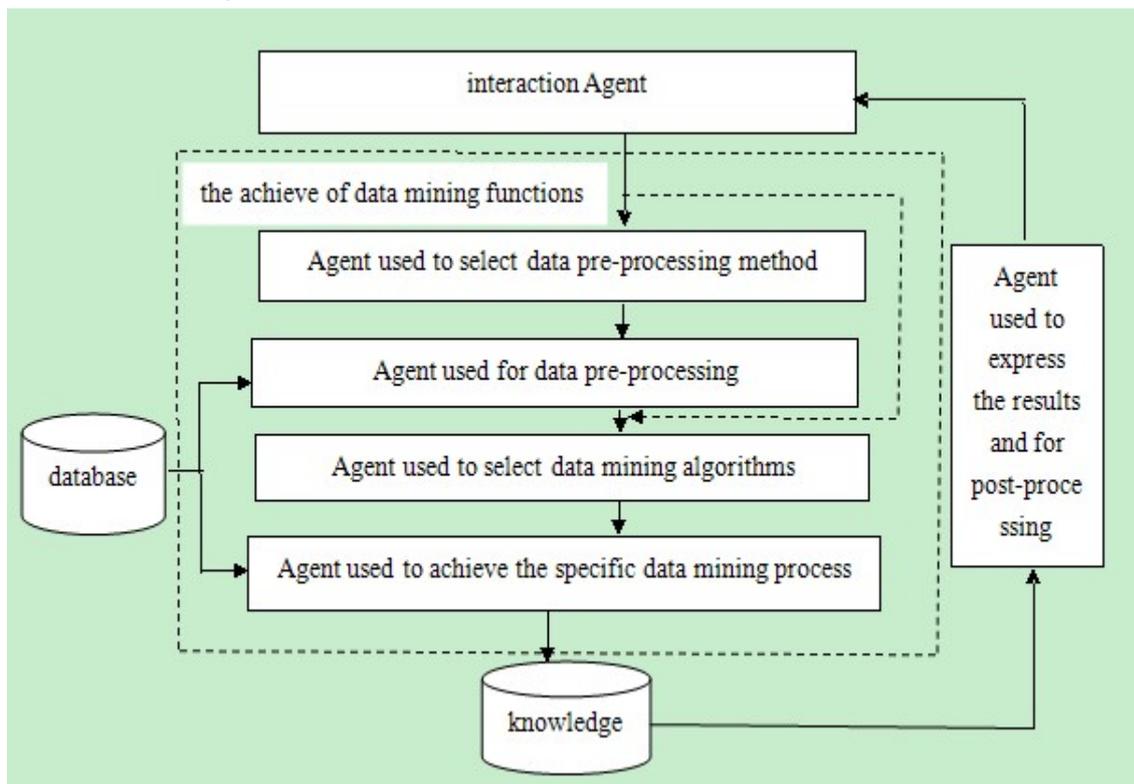


Fig. 1 Intelligent Agent-based data mining model and the interactions between the various members

2.4. The Specific Process of Agent-Based Data Mining

Use this model for data mining is a process of human-machine cooperation, through the interaction between the user and the model. The Model can self-complete data mining preprocessing and the select of Mining algorithms based on user needs. And use the corresponding mining algorithms finally get the goal data, and output it in form that meet the users' preferences. The specific steps include:

(1) The Agent that is responsible for human-computer interaction generate the mining task according to the users' requirements. This step can use the graphical interface to allow the user to input information about

implementing mining. For example: the location that the database stored, the type of mining results and the forms of expression, etc;

(2) After receiving the mining task, the interactive Agent passed the information to the Agent that used to select the method of data preprocessing. Methods Agent determine whether conduct the data pre-processing according to the received information. If necessary, it then select the appropriate pretreatment method based on their knowledge, after that the Agent pass the pre-processing tasks to the appropriate data pre-processing Agent; if it doesn't need to conduct the pretreatment, then directly choose the mining algorithms;

(3) The data pre-processing Agent is mainly used to complete the data pre-processing, and sent the results to the mining algorithm selection Agent;

(4) The Agent that used to select mining algorithms receives the data, combined the information needs of users and their expertise to select the appropriate mining algorithm, the task then pass to the corresponding data mining algorithms Agent;

(5) Data mining algorithms Agent receives the information, call the corresponding algorithm and started mining tasks, mining results will eventually be outputed to the Agent used to express the results and for post-processing;

(6) The Agent that used to express the results and for post-processing receives the results, then choose the form that meet the users' preferences and sent it to the interaction Agent to display to the users.

The model proposed above is just a simplified model, it need further analysis. If want to get really effective knowledge, such as the Agent requires the expression of knowledge, collaboration rules between the different Agent, Agent's autonomic computing mining algorithm design.

2.5. The Specific Realize of Data Mining in the Railway Information Integration Program

The data mining model proposed above is based on intelligent Agent, it achieved the whole process of data mining. In this integration program, also use the data mining Agent to specific realize.

On the basic of the knowledge in the context of rail transport, spatial information and services on the basis of information, knowledge that the Ministry of Railways management decision-making information needed will be gained through the data mining^[4]. The valuable information that can help to improve the transport efficiency also can gained through the excavation of the railway business information.

2.6. The Specific Procedure of Data Mining in the Railway Information Integration Program

The specific procedure of data Mining in the railway information integration program can be divided into the following four steps:

(1) Embed the function of the interaction Agent and the representation and post-processing Agent into the Ministry of Railways, Railways and Stations of the User Agent group, in the integration task of all User Agent group, reflect the user's mining needs through a specific, appropriate manner.

(2) In the program of develop specific solutions, the task analysis Agent use the knowledge of the field mission to determine whether exist integration of data mining-related tasks, if there exists, instruct it with an effective specific form in the program.

(3) After determined the integration program, using the evaluate Agent to evaluate the program, and converted it into executable tasks. For the data mining tasks, use the production management Agent to establish the basic function-perform bodys that are used to perform the corresponding data mining tasks. Each body performs different parts of the task. These function-perform bodys are composition to create global and local data model Agent.

(4) When performing the data mining tasks, function-perform bodys finish the responsible data mining tasks based on timing logic within the appropriate time. The information obtained is the effective integration of data model components. This information is used to support the implementation of decisions of Railways, Ministry of Railways, Railways and Stations of the integration with the inner information.

When the need of the Ministry of Railways proposed requires data mining of the Railways and Stations' information, there should include planning of the data mining tasks. This task uses corresponding dynamic mission planning Agent to complete. The specific processes are as follows:

(1) Information integration ^[5]Agent establish the total integrated solutions and create a global data model, then publish the abstained relevant standards of the data model. Put the standards into the data center,and use the dynamic mission planning Agent to evaluate;

(2) After completed the evaluation,establish the task that used to realize the global data model, also, to determine whether there exists the requirements of data mining, if exists, according to the specific location of resources,sent the relevant data mining task to the information integration Agent that belongs to the class of Railway.The task that contains the need to mining the data sources also be send to this Agent;

(3)After received relevant information of Brightest Bureau, Integration Agent firstly analysis tasks.If the data source that is required for mining is in the Railway Administration's internal information systems, Integration Agent perform mining tasks.If the data source that is required for mining is in the lower integration Agent,it separated on this task and the lower interaction Agent perform mining tasks.

(4) the same to step (3),after the stations integration Agent received tasks that is from the higher stations integration Agent,it also firstly analysis tasks.After determine the data source is located within the system, it perform mining tasks.Then Agent is used for information integration pass information of mining and meeting any requirements to its data center.Railway Administration Information Integration Agent pass it to data center of ministerial-level information integration Agent. Finally,it completed a global data model generation.

The Integrated Railway Information System conduct data preprocessing on different kinds of data sources that included in Railway scheduling and dispatch related systems management information system.And centralizly process the real-time and non real-time data to get the full, real-time, unified data center of the whole Railway .

Having abstained a unified data center, It becomes easy to integrate various systems together in the application level,the original distributed systems can now form a unified whole.

3. References

- [1] L .M. Jia, G. Liu and Y. Qin: Intelligent Agent Based Dynamical Collaboration Mechanism for Complex Task Solving[M] .Science Press2007:8-14
- [2] J. W. Han, M. Kamber: Date Mining Concepts and Techniques,Second Edition[M].China Machine Press, 2008:146-155.
- [3] N. Saiwaki, J. Kawabata, H.Tsujimoto: Cooperative performance system based on virtual agents[C].IEEE Conf. USA:1996.3230-3234
- [4] S. Bergamaschi, G. Cabir, F. Guerra, L. Leonardi, M. Vincini and F. Zambonelli.Exploiting agents to support information integration[J].*Internation Journal on Cooperative Information Systems*,2002,11(3)
- [5] M. R. Genesereth, A. M. Keller and O. M. Duschka.Infomaster:an information integration system[C].*Proceeding of SIGMOD 97*,New York,May 1997,539-542