Research and Design of a Grid Based Electronic Commerce Recommendation System

Yueling Liang¹ and Guihua Nie²

Wuhan University of Technology, School of Economics, Wuhan, China
¹ liangyl1987@whut.edu.cn, ² niegh@mail.whut.edu.cn

Received 24 August 2009; received in revised form 2 December 2009; accepted 3 March 2010

Abstract

Current electronic commerce recommendation system is designed for single electronic commerce website and current recommendation technologies have obvious deficiencies. Centralized recommendation systems cannot resolve the contradiction between high recommendation quality and timely response, as well as that between limited recommendation range and ever rich information on the web. Distributed recommendation systems are expected to improve the recommendation quality while maintaining high performance. This paper analyses the problems of traditional electronic commerce recommendation system and clarifies the advantage of applying grid technology into electronic commerce recommendation system. It discusses the prototype of an intelligent recommendation system, namely grid based electronic commerce recommendation system (GBECRS). The paper analyses rationale and mechanism of GBECRS and designs its logical structure. It focuses on the design of grid services which are needed to be deployed into electronic commerce recommendation grid. Finally it does deep analysis of key technologies that are applied in the system.

Key words: Grid, Electronic Commerce Recommendation, Recommendation System, Grid Service, Prototype
1 Introduction

Electronic commerce recommendation system can enhance the competitiveness of electronic commerce websites and meets customers’ individual demands. At present almost all of the large-scale electronic commerce websites apply recommendation technology in their systems. Common-used electronic commerce recommendation technologies include content-based filtering technology, collaborative filtering technology, vital statistics based recommendation technology, effect-based recommendation technology, knowledge based recommendation technology and hybrid recommendation technology which uses two or more of the above technologies. These technologies have efficiency more or less. Content-based filtering technology must analyze text content information of particular resource, and thus it can not be applied in recommending music, graphs and videos. Besides it can not analyze the quality of recommendation information and produces good recommendation result. Collaborative filtering technology has the problem of “cold starting” [11] and bad expansibility.

Content-based filtering technology and collaborative filtering technology are two recommendation technologies that are most common in use [1]. However, their deficiency causes bad recommendation quality and sometimes even failure appears when recommending. Being unable to recommend what customers like or recommend what customers dislike only leads to customers’ dissatisfaction with electronic commerce websites. Besides, existing electronic commerce recommendation systems can only be applied in single electronic commerce website but not for large-scale distributed applications of electronic commerce recommendation. Existing systems can not meet the demand of collaborative recommendation among different electronic commerce websites. Recommendation precision and efficiency always contradict with each other. Existing recommendation systems can only ignore recommendation quality to ensure the real-time requirement. Poor efficiency, narrow recommendation range customers’ dissatisfaction with recommendation results, recommendation systems are now facing these problems that need to be solved urgently [2].

The next section introduces grid technology and related works covering grid technology, and clarifies the advantage of applying grid technology into electronic commerce recommendation system. Section 3 discusses the rationale and structure of grid based electronic commerce recommendation system, and designs grid services that need to be deployed into GBECRS. Section 4 deeply analyzes key technologies that are applied in the system and section 5 concludes this paper.

2 The Application of Grid Technology in Electronic Commerce Recommendation System

In the middle of 1990s, the concept of grid was brought forward in reference to electric grid. At first computation grid was the emphasis of grid technology research. Now grid technology, considered as the next generation Internet technology, is trying to realize complete connection of Internet resources, including computation resource, communication resource, data resource, software resource, knowledge resource, etc. In the research of grid technology, grid system architecture is scientists’ emphasis. It is about how to construct a grid, the science of definition and description of basic grid components and their function, the regulation of grid components’ relation and method of integration, and description of grid operation mechanism. Nowadays OGSA (Open Grid Services Architecture) is the most common adopted grid system architecture. It centralizes in service and tries to realize standardization of grid system architecture. The architecture adopts standardized and common-used service pattern, which means it views all resources as service and shares all resources through sharing services. Besides, the architecture provides different types of grid application with uniform support [24]. OGSA based service grid platform and its applications will realize integration and sharing of resources and applications. In hierarchy OGSA architecture incorporates four layers, down to the top, resources layer (physical resource and logical resource), Web service layer, OGSA based service layer and grid application program layer. In the four layers, Web service layer is the key layer of the architecture, providing bottom support for all service in the grid [3].

2.1 Related Works

Grid technology has to be applied into electronic commerce recommendation system to realize large-scale distributed recommendation. To date grid technology has been adopted by many organizations and projects. Combining grid technology and community technology, Earth System Grid (ESG) of American Energy Department provides seamless and great conditions for the next generation climate research. National Earthquake Engineering Research Center of America adopts grid technology to do collaborative research on earthquake engineering simulation grid (NEESgrid), which is the most integrated earthquake engineering grid system [23]. Zhaohui Wu proposed the knowledge grid architecture KB-Grid to support applications based on databases and repositories and applied it in the construction of TCM-Grid (Traditional Chinese Machine Grid) [7]. David De Roure, et al proposed the concept of semantic grid in 2001 by incorporating the technology of semantic web into grid and adopted it as the infrastructure of e-science [8]. The above researches view grid technology from different perspectives and mainly focused on its distributed application, namely fusion of distributed resources and tools. When applying grid
technology into electronic commerce recommendation system, the demand of large-scale distributed recommendation can be met in the same way as the above researches.

2.2 The Advantages of Applying Grid Technology Into Electronic Commerce Recommendation System

This section discusses the advantages of applying grid technology into electronic commerce recommendation system.

2.2.1 Elimination of “Cold Starting” Problem

Common recommendation technology, like collaborative filtering (CF) technology, faces the problem of “cold starting”. Because CF technology produces recommendation results through analyzing customers’ score on commodities, when a new customer comes into an electronic commerce website or a new product is added into the website catalogue, recommendation system based on CF technology can not produce recommendations to customers because of lack of users’ scores on particular products [6]. With the advantage of distributed application, grid technology realizes the resource fusion and share of several electronic commerce websites. When recommendation system of a certain website receives a new customer’s recommendation request, it can obtain the customer’s information, personal data and historical scores from node websites, which the customer has registered in, through grid technology. Then the problem of “cold starting” for lack of user’s score can be avoided. On the other hand, when a new product is added into the catalogue of electronic commerce website, the grid can obtain scores of similar products from other node websites. Then through analyzing obtained scores, recommendation system can finish recommending process and recommend products what customers will like.

2.2.2 Being Able To Produce Recommendations Timely and Efficiently for Users

Recommendation tools, algorithms and mechanisms of different electronic commerce websites are different. Recommendation system based on grid technology can aggregate the most integrated users’ information and then choose the most suitable recommendation tools from node websites registered in the grid, and finally provide the best recommendation result to users timely. The common contradiction between recommendation quality and efficiency of existing recommendation system can be eliminated, thus improving the efficiency of recommendation system and service quality of electronic commerce websites.

2.2.3 To Meet the Demand of Large-scale Electronic Commerce Recommendation

Up till now, most recommendation technologies and systems are designed for and applied in single electronic commerce website, which can not realize large-scale electronic commerce recommendation among multiple websites. In grid environment, users’ knowledge, products’ knowledge and recommendation tools can all be viewed as grid resources. Grid technology realizes resource fusion of multiple electronic commerce websites. Recommendation systems of different websites can collaborate seamlessly. Besides, with the advantage of high expansibility, many electronic commerce platforms, websites and systems, even heterogeneous systems can registered into the grid. Eventually the demand of large-scale electronic commerce recommendation can be met [10].

3 Rationale and Structure of Grid Based Electronic Commerce Recommendation System

Based on system engineering science, we design the prototype of grid based intelligent electronic commerce recommendation system (GBECRS). This chapter analyzes system rationale, structure design and grid services that are implemented in the prototype.

3.1 Rationale of GBECRS

GBECRS is structured on a four-layer model that includes, down to the top, electronic commerce websites, basic grid, electronic commerce recommendation grid and grid portal. Inside the system it builds functioning mechanism of automatic acquiring of users’ and products’ knowledge, integration and intelligent recommendation. It designs and deploys grid services in GBECRS to realize recommendation workflow. The theory and functioning mechanism of GBECRS is as follows. Firstly, Electronic commerce websites as source of users and products information, basic grid in GBECRS acquires relative knowledge from websites’ server and stores it into electronic commerce recommendation repository. Secondly, when receiving users’ recommendation request, electronic commerce recommendation grid acquires knowledge from recommendation repository and aggregates the obtained knowledge. Meanwhile, it chooses the best recommendation tool and algorithm from node websites to fulfill recommendation workflow. Lastly, recommendation system invokes electronic commerce recommendation grid services, invokes and accesses grid resources to execute intelligent recommendation workflow and eventually produces recommendation result.
3.2 Structure design of GBECRS

The structure of GBECRS is shown in figure 1. GBECRS incorporates four layers, down to the top, electronic commerce websites, basic grid, electronic commerce recommendation grid and grid portal. The function of every layer is elaborated as follows.

1. Electronic Commerce Websites. This layer incorporates several electronic commerce websites and every website has registered in the grid so they can be considered as grid nodes. User information, product information, recommendation tools, users’ score and cases rules summarized through analyzing historical recommendation cases are all stored in these grid nodes, specifically in node websites’ servers. Basic grid can obtain these distributed resources from this layer. Because of high expansibility of grid, this layer can incorporate many electronic commerce websites, even heterogeneous electronic commerce systems, like mobile electronic commerce systems.

2. Basic Grid. It is the basic platform on which grid realizes its function. Basic grid is constructed based on OGSA, which views all grid resources as services and adopts WSRF (Web Service Resource Framework) to define a common and open architecture [4]. Basic grid provides bottom support for electronic commerce recommendation grid and includes some sub-layers, resource layer, web service layer, OGSA enabled service layer and grid application programming layer. The function of basic grid is to realize knowledge obtaining service, grid catalog service and MDS (monitor and discovering service), etc [5].

3. Electronic Commerce Recommendation Grid. This layer is the core component of GBECRS, where intelligent recommendation workflow is realized. Data, knowledge, recommendation tools and application program in the basic grid can all be viewed as resources. Electronic commerce recommendation grid acquires user information, product information, scores and cases rules from the basic grid and invokes recommendation algorithms to create recommendation workflow. Then it invokes and distributes grid resources to execute recommendation workflow and eventually produces recommendation result. Grid services are developed and deployed into electronic commerce recommendation grid, which is also constructed based on OGSA and accords with WSRF criterion [13]. Main grid services in electronic commerce recommendation grid are resource allocation and management service, resource operation service, knowledge aggregation service, resource access service and workflow service, etc.

4. Grid Portal. This is interaction interface between grid and users. After registering in the portal website and obtaining identity authentication, users can log in the electronic commerce website they are interested in.
Recommendation request will be triggered and GBECRS will acquire users’ relative information according to their identities, and then fulfill recommendation workflow. After the recommendation workflow is finished, grid portal shows users recommendation results. Single login technology is adopted when building grid portal to obtain users’ exclusive authentication and track users' log in different websites. Besides identity authentication, grid portal provides management interface and grid resource management interface for system administrator [14].

### 3.3 Grid Services

In electronic commerce recommendation grid, the management, distribution and operation of grid resources, aggregation of users and products information, invocation and access of data and recommendation algorithm, and execution of recommendation workflow, all of these tasks have to be realized through implementation of grid services. Electronic commerce recommendation grid is constructed based on OGSA. The realization theory of grid service is similar with Web service [15]. Grid services deployed in electronic commerce recommendation grid are as follows.

- **Resource Allocation and Execution Management Service (RAEM).** The service realizes the mapping from Electronic Commerce Recommendation Grid (ECRG) to basic grid. It aims at mapping resource request in ECRG resource space into grid resource request presented in Resource Specification Language (RSL) to meet application demand and grid restriction. Resource request of every recommendation task is presented in RSL. Analysis and processing of recommendation task will eventually turn out to be common uniform resource request, which is then translated into RSL request of local GRAM (Grid Resource Allocation and Management) service [16]. Realization of RAEMS depends on GRAM service provided by Globus, one of the most famous associations that do grid research.

- **Metadata Catalog Service (MCS).** Metadata is the information about stored data and semantic description about grid resources’ characteristics. MCS’s function includes creation, reuse, extraction and maintenance of metadata. To exchange information and knowledge among different virtual organizations, MCS has to define an open framework of metadata interaction.

- **Resource Operation Service (ROS).** The service operates on grid resources, accesses metadata base, converts grid resource operation sentences into RSL and then delivered it to RAEMS to execute. Actually ROS realizes mapping from electronic commerce recommendation grid resource operation to basic grid resource operation, guaranteeing resource operation’s logical independence between electronic commerce recommendation grid and basic grid.

- **Knowledge Aggregation Service (KAS).** KAS aggregates knowledge in grid, both users and product knowledge [11], and produces uniform knowledge view. The key point of knowledge aggregation is how to effectively solve the problem of mass knowledge confliction and redundancy clearing in knowledge grid. KAS provides users with uniform logical view as entry of accessing and acquiring knowledge in grid. Aggregated knowledge and users knowledge view are both stored into repository [17].

- **Knowledge Access Service, Data Access Service and Software Resources Access Service.** These services are called by a joint name, resource access service (RAS). RAS realizes long-distance operation on repository and database, including querying, insert, modification and delete of content in repository and database, and revision and maintenance toward the pattern of repository and database. Besides, KAS is responsible for search, choosing, downloading and invoking software and algorithms, such as recommendation tools, KDD tools and reasoning machine, etc [19].

- **Workflow Plan Service (WPS).** After recommendation system obtains suitable grid resources, recommendation workflow is automatically generated and mapped into service sequence, forming workflow plan preparing to execute, which is stored in Workflow Repository (WFR). To the same workflows, workflow plan in WFR can be reused. When some services in the service sequence have got a change, WPS will do service mapping again to form a new service sequence.

- **Workflow Execution Service (WES).** WES starts up workflow plan in WFR and does management and monitoring on execution status of workflow to ensure the whole workflow plan can be successfully executed. WES and RAEMS cooperate to finish execution process of workflow. WES is in charge of correspondence and execution of the whole workflow. RAEMS takes charge of allocation of relative resources according to specific grid environment and monitoring on operation process of particular service. RAEMS also provides WES with feedback of workflow execution status so that WES can initiate backup service when some service fails, to guarantee the integrity and validity of workflow’s execution process [12].
4 Key Technologies

GBECRS adopts four key technologies, which are elaborated as follows.

4.1 Single Login

To ensure users' identity authentication to be exclusive, single login technology is adopted to track users' identity in grid portal. This technology realizes ID mapping over multi-systems, making users' knowledge more convenient to acquire and avoiding uncertainty of users' identities because of different ID mapping mechanism of different website. GBECRS adopts single login technology to create identity authentication for users to allow them to use grid services and ensures security of users' identity authentication with application of X.509 security mechanism. A user submits registration request and after system auditing will obtain an exclusive authentication that is stored in proxy server. The user logins into the grid portal, inputs his ID and password. Then with user ID the grid portal will get a proxy authentication from the proxy server. After the authentication being scrutinized, the user will get the right to access grid service [20]. The mechanism of single login technology is shown in figure 2.

4.2 Users' Knowledge Acquirement and Aggregation

Users' knowledge is information about users' interest and preference, and it reflects what users will decide to buy and their behavior in a specific occasion. There are two kinds of users' knowledge. The first one is user summarization which reflects users' interest and preference. The second one is user occasion which reflects external affected factors when a user makes buying decision. The source of user occasion knowledge is statistics information that users input, user's IP address and users behavior analysis. The process of obtaining users' knowledge is to discover users' interest and preference. Many literature have done the research of the method to discover users' interest and modeling [21].

The key point of obtaining users' knowledge is learning and update of user summarization. Learning of user summarization usually adopts data mining technology and machine learning technology. User customization is also a learning channel of user summarization but it is a method not so automatically, which means that users have to update their summarization themselves. User customization is mainly applied in occasion of information filtering. Now Web mining is used most often to discover users' preference. The method does not require users to input their preference information so it has high automatization.

Users' knowledge can only be effectively used after aggregation and then help the system to make reasoning and judgment that are correct and effective. Knowledge aggregation in GBECRS is the process of knowledge redundancy and consistency validation, redundancy elimination and knowledge confliction clearing on several knowledge resource items. At first it checks redundancy status and then validates on the redundancy. The validation contents cover rule redundancy, condition redundancy, rule confliction, circulation, unreachable aim and dead ending. There are several kinds of validation methods, such as Boolean calculation and transformation rules, etc. In practical application, the redundancy clearing strategy is stored in rule repository [9]. Knowledge aggregation service validates on aggregated knowledge and gets clearing strategy from rule repository in inconsistent situation to clear inconsistency [22].
4.3 Recommendation Strategy's Self-adaptation Selection

Recommendation strategy is a series of rules that define what recommendation technology should be used by recommendation system in certain condition. These rules are stored in repository, inputted by experts and knowledge engineers and can also be learnt by the system [18]. Self-adaptation of recommendation technology means that recommendation system can learn from recommendation history and assess on qualities of different recommendation technologies used in similar user-product environment. After assessment recommendation system chooses recommendation technology with best quality to recommend products to users [25].

Data source that recommendation strategy learns from can be data of one website or multi-websites. Recommendation technology adopted by different websites may be different. A website may adopt several recommendation technologies [26]. The system will collect data to form a recommendation set and study on the set using Adaptive Resonance Theory. When recommending products to users, GBECRS will firstly clustering on all the user-product situations, which different situations represent different recommendation environment and user-product items in the same cluster will be considered to be in the same recommendation environment. Next the system analyzes and compares the results of different technologies in the same cluster using the method of statistics analysis, and finally gets the best recommendation technology [27].

4.4 Task Management

In GBECRS there may be several parallel tasks, like parallel operation of several recommendation algorithms and parallel execution of several workflows. To improve system operation efficiency, GBECRS must manage these parallel tasks effectively. Task management includes arrangement of task priority, reasonable distribution of resources that a task needs and task lifecycle management, etc [28].

One of the most famous grid research association, Globus, provides a toolkit to realize task management and some other grid services. One of them is Grid Resource Allocation and Management (GRAM) [29]. GRAM takes charge of task submission, long-distance application resource request processing and long-distance task invoking processing, etc. in a way of Web service interface. GRAM is task execution center in grid computing environment. GRAM provides mechanisms of querying task status, which can execute operations like task termination. Application program maybe provide feedback to users, eliminate and release resource and do some other operations with the above functions of GRAM in necessary. For example, if a task fails, other tasks depending on the result of the failed task have to be terminated lest consuming too much resource in vain. However, GRAM just provides mechanisms for task management but not specific task management component [30].

5 Implementation and Comparison

Based on the design of GBECRS, we have developed a prototype of the system that aims to recommend movies to users. The prototype collects recommendation resources from grid nodes, and calls grid service to fulfill recommendation tasks. Finally it provides recommendation results to users based on their interests, comments and scores on movies. The feature of the prototype is shown in figure 3. Left column of the feature is user control center where users can log in and out and input some personal information like their age, sex, and interests, etc. Right column in the feature is a movie list where users can view and score on movies they are interested in. Users can also put interesting movies into a collection. It is just a simple feature of the prototype which can not show all functions of the prototype for the limit of the passage.

![Figure 3: Prototype of GBECRS: For Movie Recommendation](image)
We did a test on 30 users, most of whom are college students. After collecting their personal information, such as gender, age, interests, favorite directors, actors/actresses, etc, and their scores on specific movies, we did a comparison of recommendation results between GBECRS and traditional recommendation. Part of the collected user information is shown in Table 1.

Table 2: Sample user information collected for system test

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Gender</th>
<th>Age</th>
<th>Diploma</th>
<th>Major</th>
<th>Interest</th>
<th>Favorite Director</th>
<th>Favorite Actor/Actress</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Xiaoming, Chen</td>
<td>Female</td>
<td>20</td>
<td>Undergraduate</td>
<td>Electronic Commerce</td>
<td>Reading, Music</td>
<td>Joe White; Steven Sperberg</td>
<td>Nicolas Cage; Julia Reberts</td>
</tr>
<tr>
<td>2</td>
<td>Li, Shen</td>
<td>Female</td>
<td>23</td>
<td>Postgraduate</td>
<td>Electronic Commerce</td>
<td>Reading, Music, Movie</td>
<td>John Woo; Xiaogang Feng</td>
<td>Sammo Hung Kam-Bo; Stephen Chow;</td>
</tr>
<tr>
<td>3</td>
<td>Yawen, Zou</td>
<td>Male</td>
<td>19</td>
<td>Undergraduate</td>
<td>International Economics</td>
<td>Music, Video Game</td>
<td>Eric Bress; J. Mackye Gruber;</td>
<td>Matt Damon; Will Smith</td>
</tr>
<tr>
<td>4</td>
<td>Yan, Shu</td>
<td>Female</td>
<td>20</td>
<td>Undergraduate</td>
<td>Electronic Commerce</td>
<td>Sport</td>
<td>Jing Wong; David Wong</td>
<td>Shu Qi; Hanyu Zhang;Tom Hanks</td>
</tr>
<tr>
<td>5</td>
<td>Meiguang, Yin</td>
<td>Female</td>
<td>25</td>
<td>Postgraduate</td>
<td>Computer Science</td>
<td>Movie, Jogging, Badminton, Calligraphy</td>
<td>Peter Jackson; Xiaogang Feng; Yimou Zhang</td>
<td>Andy Lau; Stephen Chow; Jet Li; Betty Sun</td>
</tr>
</tbody>
</table>

After putting these sample data into GBECRS and traditional recommendation algorithm (A-priori) respectively, we get the time that are consumed in these two tests and users' satisfaction degree mining from recommendation result data. Then we compare time consumed and satisfaction degree and finally get the comparison result. The comparison result is shown in figure 4. The horizontal axis represents time consumption for recommendation and the vertical axis represents users' satisfaction level about the recommendation result. The figure shows that compared with traditional recommendation, GBECRS can provide better recommendation results for users in less time.

Figure 4: Comparison of Recommendation Result between GBECRS and Traditional Recommendation

6 Conclusions

The application of grid technology into electronic commerce recommendation system can meet the demand of large-scale electronic commerce recommendation, which means collaborative recommendation among different electronic commerce websites. Besides, it can also improve efficiency and quality of recommendation system. This paper analyzes rationale of electronic commerce recommendation system (GBECRS) and designs system's logical
structure and grid service modules in the system. Finally, the paper does deep analysis of key technologies that are applied in the system, which paved the way for the realization of the system. The research on GBECRS can help to solve the problem of traditional recommendation algorithm and optimize traditional recommendation result. E-commerce websites can benefit a lot from our research. However, large scale distributed application requires support of expensive servers to provide running environment of thousands of grid nodes, we only did some test to validate our assumption in simulated large scale grid environment. To guarantee the validity of our research we should further enlarge our research environment. Besides, we still have not put our research result of knowledge grid into practice, so based on the prototype, GBECRS, in the future we will construct a knowledge grid to run electronic commerce recommendation system.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant 70572079.

References


