International Journal of New Technology and Research (IJNTR)
ISSN:2454-4116, Volume-2, Issue-2, February 2016 Pages 01-05

Study of Cloud Services Recommendation Model Based on Chord Ring

Chen Li, Qing-Tao Wu, Jing Chen

***Abstract*— The advent of cloud computing era, the amount of application data increasing sharply, To solve the problem that the combination of personalized recommendation technology and cloud computing, which facing prolonging the recommended time delay and on the large of network overhead, a cloud services recommendation model based on chord ring is proposed. In this model, cloud-based distributed storage model into Chord ring, Sort recommendation process a greater impact on service metadata collection system to ensure quick retrieval of candidate recommendation service sets; proposed recommendation algorithm based on weighted bipartite graph on this basis, predict the target user top-k set of recommendations. Experimental results show that the mechanism can effectively improve the recommendation accuracy and recommend efficiency.**

***Index Terms*—cloud service, Chord ring, personalization recommendation, weighted bipartite network.**

I. Introduction

Highlight with the rapid development of Internet technology, online data increasing rapidly, on the one hand, people could exchange information and share data remains within doors, on the other hand, with the sharp increasing of network information, users are difficult to quickly find valuable information from these vast amounts of information, so the rate of information utilization decline, which brings the problem of information overload [1]. Personalized recommendation, which as an important information filtering measure, actively recommends to users with its potential interesting in the item by analyzing the interests and historical behaviors of users, which effectively solve the problem of Internet information overloading [2]. The recommended results of personalized recommendation system are closer to the users' individual needs, which is different from the "one to many" server offered by search engines (that is, search engine presents the same search results to all users, which cannot provide the corresponding services by different users' interests [3]).

Personalized recommendation develops rapidly in recent years. A complete recommend system consists of three components: recording module to collect user behavior information, analysis module to analysis potential interests of user and recommendation algorithm module to real-time filter the information which is user interests from target collection. In which, the recommendation algorithm module is the core of the recommendation system [4]. Personalized recommendation system is mainly divided into collaborative filtering recommendation system, recommendation system based on content, mixture recommendation system and recommendation system based on network structure of user-product according to different recommendation algorithms. In which, network structure recommendation algorithm of the user- project bipartite network has received extensive attention of researchers [5]. Bipartite network as a kind of special network, the principle of personalized recommendation for users is using the process of complex network dynamics such as substance diffusion, heat conduction on bipartite network, etc. Bipartite network is the simple heterogeneous information network, which contains only two types of nodes, edges only exists between heterogeneous nodes. Recommendation algorithm based on bipartite network does not consider the content features of user and project, the information used in all algorithms are hidden in the selection relationship between the user and project [6], so bipartite network recommendation algorithm in cloud computing environment has wide application prospect.

On the other hand, cloud service providers in order to improve the parallel computational efficiency, constructs the structure without a frame of server cluster [7], each node in the server cluster are independent of the local storage, data is distributed storage to each node parallel processing, the same user score according to also be dispersed storage in different nodes, so in the process of recommendation and need to retrieve data from each storage node, which will lead to the problem of high delay[8].cloud service providers need frequent updates the user of service evaluation data set, so it will lead to inter node data transmission rate of increase, increasing the network overhead.

In literature [9], Aggarwal was first proposed recommendation algorithm based on network structure (bipartite network) in KDD 99. In literature [10], Huang, etc. proposed a network diagram recommendation algorithm based on the allocation of resources. Using product bipartite network to build product correlation, and put forward recommendation algorithm of bipartite network, and open up a new direction of researching recommendation algorithm. In literature [11], Zhou, etc. proposed network recommendation algorithm (Network-Based Inference, NBI), using bipartite network to allocate resource, which has better results than collaborative filtering algorithm CF, but in this paper,

Chen Li, computer application technology, College of Information Engineering, Henan University of Science and Technology, Luoyang, 471023, China, +8618238883541,

Qing-Tao Wu, information safety, College of Information Engineering, Henan University of Science and Technology, Luoyang, 471023, China.

Jing Chen, computer application technology, College of Information Engineering, Henan University of Science and Technology, Luoyang, 471023, China.

bipartite network has no weighted, resource allocation is evenly distributed among projects, without considering the problem of edge and weight. In literature [12], I.S.Dhillon was first applied spectral clustering to vocabulary-document bipartite network, the algorithm is a clustering algorithm based on graph partitioning, which regard the problem of clustering as a multiple segmentation problem on undirected graph, its essence is using eigenvalue and eigenvector of Laplacian matrix as a tool of describing the connectivity of graph. Tveit ,etc. proposed flooding strategy recommendation algorithm, the target user transmits score vector to calculate candidate service set, the algorithm will have a huge network overhead [13].

This paper proposes cloud services recommendation model based on chord ring to improve the recommendation accuracy and recommend efficiency. The main work includes the following two aspects: adopting chord ring model to improve the speed of retrieving service candidate sets, optimizing bipartite network recommendation algorithm to improve the recommendation accuracy.

II. CLOUD SERVICES RECOMMENDATION MODEL

A. *Recommendation model based on Chord ring*

Cloud service recommendation model based on chord ring is divided into 4 layers, which are basic data layer, data processing layer, service recommendation layer and user access layer. As shown in the figure 1 below.

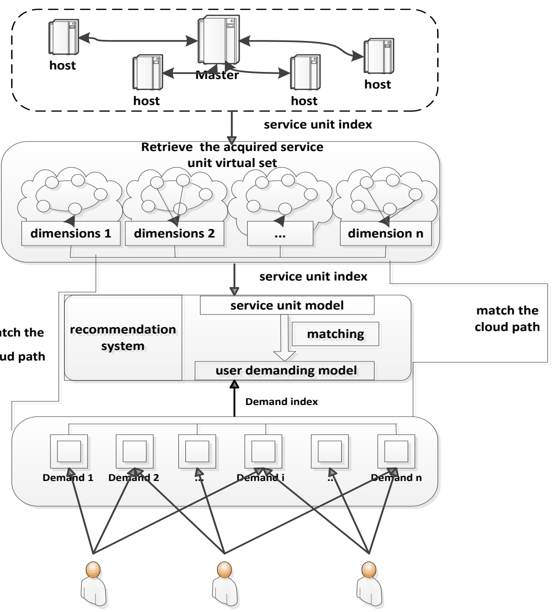

Figure 1: The cloud service recommendation model based on chord ring

1) Data pre-processing layer: To data pre-processing on the need of the recommended target set, that is, when user recommendation service resources, server cluster retrieved each storage node, then found out all selected history service unit by users, which provide support to 2).
2) Data processing layer: Searching the same attributes of each storage node in server cluster according to the service resource attributes, generating some chord rings, in which each chord ring has a same service element attribute.
3) Recommender system layer: Responsible for the corresponding recommendation request triggered by users, calculating the similarity between service unit with the same attributes, predicting service unit ratings of users and generating recommendation sets.
4) User access layer: According to the recommendation result, the end-users select the service resources, and extract demand model of each attribute in service resources.

B. *Cloud service organization structure*

According to the function and attribute of the cloud service, the service resource is organized by the Hashi function based on SHA-1, which is shown in figure 2. The different attributes of the service resource is generated the corresponding attribute based on the hash function. Service resources by SHA-1 operations on their own property and get an M bit flag number, denoted by ID, given the following correlation definition:

Definition 1(Sub Chord Ring) According to the service of each attribute tag and the size of the ID order to service form a logical ring, that is, the sub Chord ring.

Definition 2(Main Chord Ring) A ring with the services including highest reputation on this attribute in all sub chord ring.

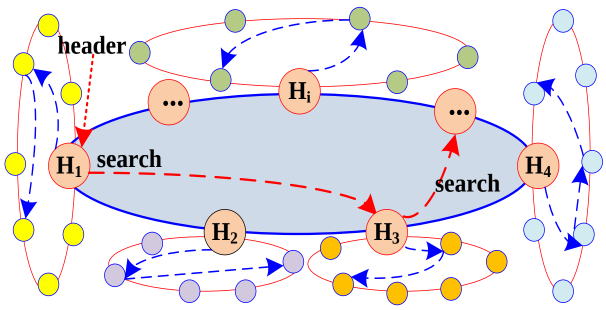

Figure 2: Service Natural Virtual Organization Structure

Using dynamic “the stronger” clustering according to the different attributes of service resources, which realizing that mapped distributed service resources into corresponding properties of the virtual collection. In which, “the stronger” service resources refers to service resources with the highest score in the same class. Then, represent the generated corresponding virtual collections in the form of a "ring". According to the different attributes of service resources to generate a series of sub-ring, and all the sub-rings make "stronger" as head node to build a more advanced main ring. When searched service resource, searching from the main ring to all the sub-ring, generating a virtual service unit set D that conform the requirements of a search service.

III. RECOMMENDATION ALGORITHM BASED ON WEIGHTED BIPARTITE NETWORK

In this paper, *Q* represents user requirements, *U* represents all users, *u* represents individual users. Data processing layer retrieves storage nodes on each host based on the needs of users, which uses *D* to represent it. In cloud computing environment which contains massive services, describing services recommendation as $G=(U\cup Q\cup D,E)$, there are links between user requirements and service unit in some way to abstract into edge set E. Therefore, the fundamental problem of the information retrieval can be expressed as the process of user u sending a query of Q to seek which meet the cloud services of user demand.

A. The Basic Thought of Bipartite Network Recommendation Algorithm

If there are more than one type of nodes or links in information network, which called heterogeneous information network. Further, dividing heterogeneous information network into simple heterogeneous information networks and complex heterogeneous information network. Simple heterogeneous information network mainly contain certain network diagrams such as bipartite network, tripartite network, double-type information network and star information network and so on. In cloud environment, vast and heterogeneous cloud service resources from different suppliers and users of various *QoS* requirements can be regard as a heterogeneous information network [14].

Bipartite network also called binary network, which is the most simple heterogeneous information network. Edge exists only between heterogeneous nodes, and there has no edge in homogeneous nodes. Many applications in information retrieval can be abstracted as a bipartite network, such as click relationship between query and document, the relationship between author and publish articles, marked relations between images and labels and so on.

The traditional bipartite network is no weighted network, that is in the process of allocating resources among projects, allocating project resource evenly to users, while users allocating the allocated resources to project. This algorithm only determine whether the user has chosen the project or not, which has no consideration in browsing but no buying, in which also contains the tendency of the intending to buy of the user, which may cause the loss of information. Therefore, in practice, the weight of edges between user and project plays an important role. Introducing weight to value weight *v* of the edge of user-project, if user y_i has selected project x_j, $v_{ij}=1$, if user y_i has browsed but no bought $v_{ij}=\lambda$, in which the value of λ between 0 and 1, if user y_i has neither browsed nor bought, $v_{ij}=0$. Any project *j* allocating resource to project *i* is by all the edges connected to protect *i* and *j* to allocate, the weight calculation formula is

$$w_{ij}=\sum_{l=1}^{m}\frac{a_{il}a_{jl}}{k(y_l)k(x_j)} \tag{1}$$

In which, $k(y_l)$ represents as the degree of user y_l, that is, which is the number of projects connected to user y_l, $k(x_j)$ represents as the degree of project x_j, that is,which is the number of users connected to project x_j, a_{il} is the value of *i*th row *l*th column in the adjacency matrix $A=(a_{il})_{n\times m}$ of $n\times m$, λ represents as the weight between project x_i and user y_l in bipartite network. Calculating grade $v_{\alpha\delta}$ formula of target user x_α to not rated project y_δ is as follows:

$$a_{il}=\begin{cases}0\ , & x_i y_l\notin E, and\ \ user\ \ has\ \ no\ \ browsed\ \ this\ \ project\\ \lambda\ , & x_i y_l\notin E, and\ \ user\ \ has\ \ browsed\ \ this\ \ project\\ 1\ , & others\end{cases} \tag{2}$$

Calculating all the not rated projects for a target user, in this paper, it is not feedback directly the project which higher *a* value *r* to user, but save and as candidate service set S. As shown in formula (6).

$$f'(o_i)=\sum_{j=1}^{m}w_{ij}a_{jl} \tag{3}$$

B. Efficient cloud retrieval based on Chord ring

This paper uses peer-to-peer retrieval method of Chord protocol, designs quick resource locate method based on Chord, Chord chooses SHA-1 as hash function to ensure repeatability of the hash. SHA-1 produces the space of 2160, each item is big integer with 16 bytes (160bit). We can think these integers end-to-end to form a ring, which called the Chord ring. Each node in the Chord ring maintains a Chord routing table, which could locate data block resources more conveniently, and then reduce the resource search time.

Each routing table has m items. The routing table structure of node n in the Chord as shown in table 1.

Table 1 Chord routing table structures

Expression	Definition
$finger[i].start$	$(n+2^{i-1})\bmod 2^m, 1\le i\le m$
$finger[i].interval$	$[\,finger[i].start,\ finger[i+1].start]$
$finger[i].node$	The first node is greater than $finger[i].start$
successor	Successor of node n
predecessor	Precursor of node n

To seek the successor of node k, while there has no key value of node k in the node n routing table, then performing the following actions, the node n selects node m which nearest k, then find the node more closer k node by node m. By repeating the operation, we can locate the needed resources finally.

The algorithm 1 gives corresponding pseudo-code of storing nodes resource location. This algorithm judges whether node n is the precursor of k. It has not find the required resources in routing table, then find node n' which nearest the node with target resources, and repeat the algorithm by node n' until find the target resources.

From the structure of Chord routing table and storage nodes resources locate algorithm, we can find that every turn it find the nearest node, the distance between new node and resource object usually less half than the original. In general, it could be successful positioning less than logN times, thus it could accelerate data block positioning of the storage node.

Algorithm 1 resource location algorithm

```
n.findSuccessor(k)
   if (k ∉ (n,n.successor ] )
return n.successor;
   end if
   for (i=m;i>1;i--)
     if (finger[i].node in (n,k))
```

```
                return finger[i].node;
            end if
        end for
        n' =n;
        return  n' .findSuccessor(k);
```

When m=4, n=2, k=14, find the example diagram shown in figure 3:

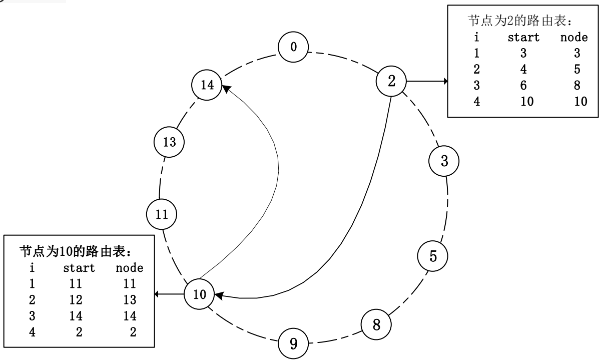

Figure 3: Map of resource location

C . Algorithm description

The specific steps of cloud service recommendation algorithm based on Chord ring are as follows:

Step 1: Calculating the candidate set.

Input: Score matrix *R* and target user u_a of users and service units.

Output: The set of u_a is the list of recommended for selecting projects.

1) According to the score matrix *R* of the user and service unit to construct the bipartite network *G* which represents the relationship between the user and the service unit. In which, the weight of the edge λ determined by user whether or not to purchase or browse the project. If the user has purchased the project, so *v*=1, if the user has browsed the project, $v=\lambda, 0<\lambda<1$, if the user has never purchased nor browsed the project, *v*=0.

2) If the user has just browsed the project, regulating the value of λ, which could be adjusted λ has the impact on formula (1), and to get a more accurate rating matrix w_{ij} .

3) According to the various attributes of the service unit to generate the corresponding attribute expand name by hash algorithm, in which expand name contains the attribute information of the project, then uses "the stronger service unit" clustering algorithm to map the distribution service unit to the virtual set of the corresponding attributes to generate a series of sub-chord ring. In which "the stronger service unit" refers to the highest rate service unit in the same class service unit, each sub-chord ring has a same attribute service unit, then regard highest rate service unit as head node to constitute the main chord ring.

4) According to the formula (3) to get corresponding attributes by the user history selected service unit as forecast rating set of the service unit, then classify into each sub-chord ring by attributes, select *Top-N* forecast highest rating projects in sub-chord ring as candidate set *O*.

Step 2: Recommending corresponding attribute service unit set O' according by the selected service of the target user.

Retrieving the service unit which has the same attributes from main chord ring according by the selected service of the target user until to the all sub-chord ring, returning the final recommendation service unit list to user.

IV. EXPERIMENTATION

A. Data set

Using standard data set MovieLens to test the feasibility of the algorithm [15]. This data set contains 1682 movies, 943 users, a total of 100000 movie ratings with users. The rating value from 1 score to 5 score, the degree of liking the movies increments from 1 score to 5 score, the rating value more than 3 represents recommend this movie, selecting score records more than 3 are 82520 as edges which constitute as users-film bipartite network. The experiment platform configuration is as follows: 2.0 GHz Intel CPU, 2 GB of memory, Windows XP, and uses the Java programming language. In the experiment, selecting randomly 90% of the data in the data set as training set to design and construct algorithm, the other 10% as test set to test the feasibility of the recommended algorithm. Selecting 10 average value of the evaluation result as final recommendation result to do comparison test.

B. Algorithm Evaluation Index

Using average rank score [11] to measure the recommended accuracy of the algorithm in this paper, using Hamming distance [11] to evaluate the diversity, and using recommended time delay to evaluate the execution efficiency of the algorithm.

Average Rank Score

For the target user, recommendation algorithm could be given the user a sort of length L recommended list, if user has selected service o_j, and o_j at the $R_{i,j}$ in the recommended list, arguing that the relative position of project o_j is as follows:

$$r_{ij}=\frac{R_{ij}}{L} \tag{4}$$

Service o_j in the test set is the selected of user, the more accurate algorithm of relative position and the at rank, the smaller of r_{ij} .

Diversity

For any two users u_i and u_j , which distance between its recommended list is

$$H_{i,j}=1-\frac{Q_{i,j}}{L} \tag{5}$$

In which, *L* represents the length of the recommended list, $Q_{i,j}$ represents the same project number in length *L* recommended list of users u_i and u_j . Calculating the Hamming distance between any two users, then calculating the average value *H*, using *H* to measure the diversity of the algorithm. Values *H* is between 0 and 1, *H*=1 represents all the user recommended lists are not the same, which has the best recommended diversity, and *H*=0 represents all the user recommended lists are the same.

Novelty

Because of the less popular products, the more novel users fell, so the algorithm should be able to recommend the project of difficulty to find for users, rather than fashion products. Therefore, the degree of novelty evaluated by the average degree <K> of L projects in user recommend lists. It is more practical significance to recommend less popular commodity than popular products, so the smaller the average degree of <K>, the better of the algorithm novel degree, that is not a popular project could also be recommended.

C. *Analysis of Experimental Results*

This experiment mainly comparison accuracy, diversity popularity and so on, set v =0.5 in this experiment, L represents the length of the recommendation list. The experimental results are shown in figure 4 and figure 5 and figure 6.

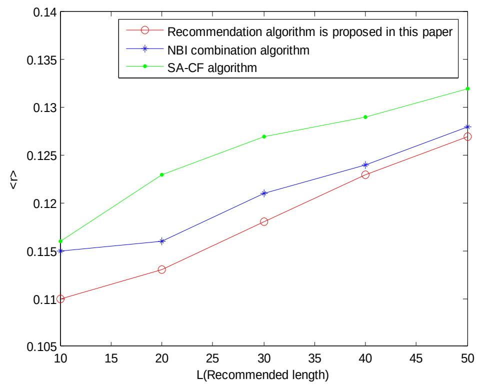

Figure 4: The accuracy

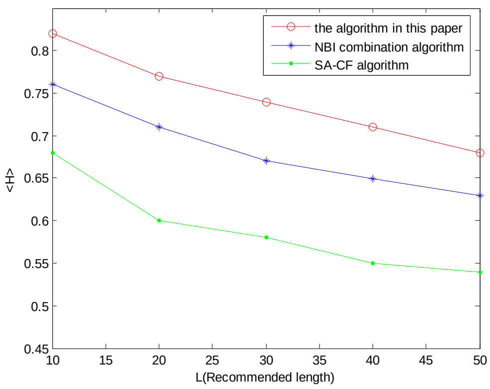

Figure 5: The diversity

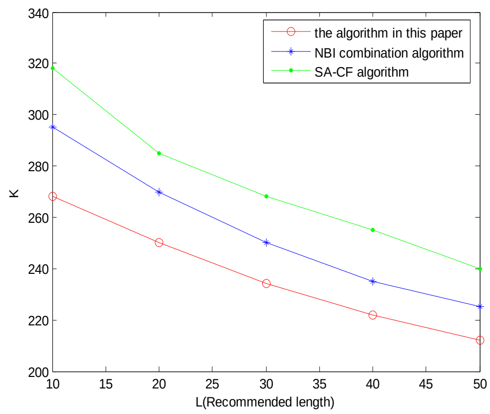

Figure 6: The popularity

From figure 4 and figure 5 and figure 6, we can see that recommendation algorithm based on the weighted bipartite network can improve the recommendation accuracy and diversity, and reduce popularity of the recommended project. This indicates that the degree of personalized has been enhanced, and the less popular service also has the chance of recommendation.

V. Conclusion

This paper presents a bipartite network recommendation model based on chord ring in cloud computing environment, which is used to solve the problems of the traditional recommendation algorithm applied to the cloud computing environment, such as the loss of the score and the extension of the recommendation. Firstly, this paper introduces chord sorting to select service resource which has greatly influenced on the process of recommendation, and convenient recommendation system rapidly retrieve the candidate sets, then introduces weight in the bipartite network to improve the accuracy of recommendation algorithm. The experimental results show that the recommendation algorithm in this paper have significantly improved than NBI algorithm and SA-CF algorithm in the recommendation accuracy, the degree of personalization and recommendation of unsought goods.

References

[1] Tunkelang D. Recommendations as a conversation with the user[C]//Proceedings of the fifth ACM conference on Recommender systems. ACM, 2011: 11-12.

[2] XU Hai-Ling, WU Xiao, LI Xiao-Dong, YAN Bao-Ping. Comparison Study of Internet Recommendation System [J]. JOURNAL OF SOFTWARE, 2009, 20(2): 350-362.

[3] HUANG ZAN, DANIEL ZENG, HSINCHUN CHEN. A comparison of collaborative-filtering recommendation algorithms for e-commerce[J]. IEEE Intelligent Systems, 2007 (5): 68-78.

[4] DHILLON INDERJIT S. Co-clustering Documents and Words using Bipartite Spectral Graph Partitioning[J]. 2001: 269-274.

[5] JIAN-GUO LIU, BING-HONG WANG, QIANG GUO. Improved collaborative filtering algorithm via information transformation[J]. International Journal of Modern Physics C, 2009, 20(02): 285-293.

[6] YI-CHENG ZHANG, MARCEL BLATTNER, YI-KUO YU. Heat conduction process on community networks as a recommendation model[J]. Physical review letters, 2007, 99(15): 154301.

[7] Mell P, Grance T. The NIST definition of cloud computing[J]. 2011.

[8] Zhang Y C, Blattner M, Yu Y K. Heat conduction process on community networks as a recommendation model[J]. Physical review letters, 2007, 99(15): 154301.

[9] Wang J C, Chiu C C. Recommending trusted online auction sellers using social network analysis[J]. Expert Systems with Applications, 2008, 34(3): 1666-1679.

[10] ZAN HUANG, HSINCHUN CHEN, DANIEL ZENG. Applying associative retrieval techniques to alleviate the sparsity problem in collaborative filtering[J]. ACM Transactions on Information Systems (TOIS), 2004, 22(1): 116-142.

[11] TAO ZHOU, JIE REN; MATÚŠ MEDO; YI-CHENG ZHANG. Bipartite network projection and personal recommendation[J]. Physical Review E, 2007, 76(4): 046115.

[12] KOH SU MING, LIANG-TIEN CHIA. Web image clustering with reduced keywords and weighted bipartite spectral graph partitioning. Advances in Multimedia Information Processing. PCM 2006. Springer Berlin Hei, 2006: 880-889.

[13] TVEIT AMUND. Peer-to-peer based recommendations for mobile commerce[C]. Proceedings of the 1st international workshop on Mobile commerce. ACM, 2001: 26-29.

[14] ZHANG Yun-yong, LI Su-fen, WU Jun, FANG Bing-yi. Research on the cloud services provider-oriented services selection method [J]. Journal on Communications, 2012, 33(9): 66-76.

[15] CiteULike Data Sets.Who -post -what data [OL].[2010-12-13].http://static.citeulike.org/data/Current.bz2.