A Research on Personalized Retrieval of Mobile Learning Resources

Lin Hu, Honghua Xu*, and Xinfang Wei
Changchun University of Finance and Economics, Changchun, Jilin Province, China
*Corresponding author, Email: huhu315@sina.com, honghuax@126.com

Abstract—With the widespread use of mobile devices and fast development of wireless network, m-learning has become another hot topic except e-learning. Considering the less utilization of traditional learning resources and the poor-directed users, this paper puts forward an improved method of m-learning. Due to oceans of m-learning resources, this paper suggests an integration platform of information resources based on granular in order to improve learning efficiency. Meanwhile, personalized concept is introduced into the system. This paper studies personalized problems of hierarchical retrieval model and the methods of how to dynamically obtain users’ interest. According to the hierarchical features of website structure, the system can actively obtain users’ interest as well as output retrieval results based on different users’ backgrounds under the help of ant colony algorithm. This method is easy to realize and can effectively trace users’ short-term and long-term interest changes. It is suitable for the complicated and changeable network environment.

Index Terms—Mobile Learning; Personalized Retrieval; Hierarchical Model

I. INTRODUCTION

Because the fast development of internet technology, there are more and more forms of resources expressions. From traditional plain text to multimedia, the internet environments where resources exist are different. Because the internet information resources boom; the sorts of resources greatly are enriched; retrieval systems of database and information resources are more and more; retrieval ways and means are various, data redundancy is resulted. The degree of interrelatedness is low. Abundant information isolated islands appear, meanwhile, the retrieval burden of users is gradually aggravating, whereas, there exist many problems in current information system such as module dispersion, resources isolation, inefficient integration. How to gather, order and correlate these information resources which are isomerous and distributed in a certain range in a logical or physical way in order to make it easy to manage, utilize and serve [1, 2]. Information resources integration can also be understood as gathering discreet resources, changing unordered resources into ordered ones to make users easy to search information and to make it more convenient to serve users. This is a general definition of information resources integration. It includes the processes of information picking up, organizing, processing and serving. Learning resources integration is mainly about m-learning. Integrating relative and discreet learning resources is convenient for retrieving, managing, using and realizing personalized study.

Learning information resources integration generally includes the following three layers [3]: The first is data layer. Information resources all exist in various high-capacity servers. The data’s forms of expression are various due to different database types. The second is operation layer, that is, relative learning information resources are utilized through software or platform. The third is application layer, that is, it is an all-round integration of learning data content, software systems and basic setup.

After integrating learning information resources, the integrated platform provides a single-stop-style visiting method for all the resources. Moreover, integrated accessible site categorizes all the learning information resources, from the originally addressable service to active service to make visiting more convenient and satisfy users’ demands.

The current era is the one with information explosion. Facing oceans of information, most people are at a loss, not clear about their actual demands. During the process of learning and retrieval, users should be more well-directed to visit the websites of learning information resources [4, 5]. Through adopting the integrated platform of learning information resources, learning information resources can be reasonably categorized; the needed resources can be classified and categorized more briefly and the users’ response methods are more personalized.

Personalized retrieval may quickly solve the problem of “one for all users”. It can offer corresponding advices
according to the differences in users. By means of hierarchical information of the websites, users’ visiting records can be classified in order to find out the changes of users’ interest. Thus, personalized retrieval method can be realized based on users’ interest.

On the basis of m-learning resources integration, users may quickly find out the information they need with the help of personalized retrieval technique. Therefore, the visiting efficiency can be improved and the time users need to visit information websites can be reduced.

II. GRANULAR COMPUTING AND QUOTIENT SPACE

Granular computing [6, 7] with other subjects’ philosophic thinking and methodology in return abstracts them into methods and strategies unrelated with concrete fields. The process of granular computing is the process of solving to complicated problems. The comprehensive understanding to complicated problems usually has multiple points of view [8, 9]. Each angle is multi-stratal. The result of granular computing presents a granular structure with multiple points of view and multiple levels. This structure is a systematic and approximate description and solving to complex problems.

The granulation of problem space refers to decompose concrete problem into several small ones and the understanding to the whole problem into analyzing in from different aspects [10-12]. The granulation of problem space may use the principles of approximate decomposition and loose coupling and it also may adopt the top-down method or bottom-up method. Granular computing can solve problems through visiting granular structure including upward and downward interactions in hierarchical structure and the moving way in the same hierarchy. There are two ways: One is the transformation and reasoning among granularities in the same granular layer. Two is the transformation and reasoning among particles in different granular layers. The connection among different granular layers is shown by mapping. The same problem on different granular layers is shown by different details and granularity.

Quotient space [13, 14] is a common method of granular computing. In the theory of quotient space, there are triples (X, F, T) to represent the objects to be studied. Among the three elements, x stands for universe [9]. T stands for topology structure on X and F stands for property function. Suppose an equivalence relation R on the triples (or an equivalent division). The question of projection to be discussed by the theory of quotient space is to consider quotient structure and quotient property function on quotient set X defined by equivalent relation R, that is, the construction problem of [X], [F], [T]. In fact, it is the transformation problem from different layers of granulation.

The acknowledged feature of human beings is that we can observe and analyze the same question from extremely different granulations. The so-called granulation is to categorize the elements with similar features to a new element, then to study it. If a subset is considered as an element, the elements in the subset are considered as equivalent ones, which is in accordance with the mathematic concept of equivalence relation which is also called equivalence relation or indistinguishable relation. According to the relation mentioned, all the equivalent elements can be called as new elements. All these new elements compose a set called quotient set. So granulation analysis is changed into the object analysis of certain quotient set.

In the study of different granulations, we find that the granulation of each question has one-to-one correspondence with its indistinguishable relation. Faced with a complicated question, people can’t consider all the details of the question; therefore, they need to analyze the question from multi-level granulation. In order to find out multi-level granulation solving, thinning the question gradually may greatly improve the computing efficiency.

To begin with, it is necessary to find out the most suitable division rules which include property division, structure division, constraint division and projection division [15, 16]. There are many kinds of methods to divide granulation. It is quite important to choose proper division method according to different granulation. In the solving process of actual granulation division, there needs to adopt quotient space reasoning model to reason and analyze on high-level granulation to get a certain property. Then we can thin the question gradually in order to solve the problem.

III. LEARNING INFORMATION RESOURCES RETRIEVAL ON MULTI-LAYER HIERARCHICAL MODEL

A. The Retrieval Model of Learning Resources

With the fast development of m-learning, the amount of m-learning resources is quite large. The types of learning resources are richer and richer. Efficient retrieval for learning information resources is the technique which can make learners realize study at any time, at any place. Brief m-learning process is shown in Fig. 1.

![Figure 1. M-learning Process](image)

However, there exist a lot of isomerous learning resources on net, which brings difficulty in retrieval and integrating learning information resources. Learning information resources which are from different places (online class, learning video, forum etc.) and shown by different ways (testing, class learning, posing questions etc.) may decrease m-learning efficiency. Therefore, this
Learning resources retrieval is an important part in the process of resources integration. The purpose of integrating isomeric learning resources which are shown in various formats stored in different places and from different databases is to make users more convenient and more reasonable to make use of resources. How to provide users with high-efficiency resources retrieval service to make them more convenient and faster to find out the needed resources is a problem we must face in the process of resources. If a united retrieval platform is needed for information resources from different servers and different databases, an agent program is also needed to transform the inputs of the users into inquire statements of various databases and then search base by base. The agent program transforms the searching results into a united format (such as XML etc.) and then sends them to browser terminal in order to show the users.

To better find out the learning resources and information the users need and simultaneously to avoid the unnecessary repeated labor and resources waste, resources retrieval platform may provide the function of synonym extension, that is, similar meanings or similar language expressions are considered as synonyms. For example, while retrieving word “language”, various expressions containing “language” can be found. We may take the current hot learning keys as initial granularity to perform equivalent divisions on different layers in order to form equivalence classes. Because there is no problem for human beings’ thoughts to switch among different granularity layers, we can construct quotient space based on universe to go back and forth freely on various granularity in the actual application.

Fig. 2 takes key “language” as an example, illustrating how to divide equivalently and how to fulfill resources retrieval on different granularity layers according to the choices of users.

![Figure 2. The retrieval process of learning information resources](image)

When users search key “language”, they may not be quite clear about their intentions. So related searching help is needed. On different layers (granularities), related searching obtains granularities through drawing granularities from universes. When users find out proper granularity, the key the user had chosen will be considered as retrieval condition on the granularity. The system is to construct different SQL statement to search for information and then return results to users.

The research/study applies common related methods of quotient space in granular computing to the retrieval process of learning information resources integration. It directs users to look for definite purposes. It helps find out the needed resources and improve the efficiency of learning resources retrieval to make m-learning process more efficient.

B. The Algorithm of Learning Resources Retrieval

There exist large amount of m-learning resources on net. How to design efficient retrieval algorithm of learning information resources is the key to realize study at any time, at any place. Therefore, this part puts forward a learning information resources retrieval algorithm based on quotient space after studying relative methods of granular computing and quotient space and also puts forward an integration technique based on granular computing. In the retrieval algorithm, because learning resources are obtained from web, a united URL class of returning results should be designed [17-19]. These returning results contain the struts of the retrieved linking address URL, retrieval keys, corresponding URL’s introduction and the retrieved number.

Firstly, according to retrieval keys to set up various layers’ granularities to be considered as initial granularities, the system performs equivalent division of various layers to form equivalence classes. Because human beings’ thoughts can freely switch among various granularity layers, we may construct quotient space according to granularities drawn from universes. But, sometimes users are not quite clear about what they intend to retrieve and what they express. Therefore, the related results in many databases through many SQL statements should all be displayed for users to choose from. The related searching should be carried out on various layers.

Secondly, while providing intelligent displays for clients, if there is a certain display that is suitable for a user’s demand, the key that the user has chosen will be an retrieval condition on that granularity. The results will be returned to users through constructing various SQL statements. If there are too many display records returned, users may choose one as child key. Then the procedures above can be repeated. Through the repeated retrieval, the granularity may become smaller and smaller with repeated retrieval.

At last, because of considering learning information resources integration, in the process of designing algorithm, the problem is that how to organize different types or various constructions of data should be considered, that is, the problem is how to convert among different formats such as doc, pdf, html, txt etc. in order to use a united and merged way to show users.

In Fig. 3, when the user inputs keys on the interface web, there will be corresponding intelligent clues provided for users to choose from. Here, there are two
situations. One is that the needed information will be found in the first retrieval. Then it will be transformed and merges to show the user. The second is that the needed information won’t be found in the first retrieval. Then the sub-retrieval will occur until the satisfying granularity can be retrieved.

![Diagram](image)

Figure 3. The algorithm of learning information resources retrieval

IV. PERSONALIZED RETRIEVAL BASED ON USERS’ INTEREST

When people search for information materials, they always hope to find out abundant and accurate information, that is, they hope that the result information can satisfy the following two demands: One is that the result information can include all the relative content concerning the retrieval theme. Two is that the result information should exclude the irrelevant content concerning the retrieval theme.

In order to more accurately express the retrieval theme of the user to satisfy the demands of different users, we put forward a personalized retrieval model, that is, when information is retrieved, the retrieval not only retrievals according to retrieval statements but also considers the user’s background and interest.

A. User’s Interest Model

The key technique of personalized retrieval is to obtain user’s interest. The obtaining methods include two sorts: explicit and implicit. Displaying obtainment needs users to actively provide interest modes. Generally there are two ways: One is to express their preferences and retrieval intentions through the method of selecting field [20, 21]. Two is to evaluate the retrieved results and feedback (including positive feedback, negative feedback) through certain method. Displaying obtainment can be easily realized and can get the actual interest of the user but it adds some extra operations to the user, so it is not easy to use. Implicit obtainment is transparent to users. Through tracking the users’ scanning actions, potential users’ interest mode can be mined out by means of web servers, users’ terminals and other personal devices [22-24]. These information containing many content of user’s scanning including the data such as the webpages users scan, the time spent on the webpage, collecting or not, file length, visiting frequency, URL address, turning the page etc. Through analyzing these information, feature data of the users can be mined out to be taken as users’ interest. Mining data mainly include web content mining, web structure mining and web usage mining.

According to different time span, user’s interest mode can be classified as long-term mode and short-term mode. Users’ long-term interest and background are the main elements of their personalized information and it is called long-term interest mode. During a short period of time or in a process of retrieval, there is a definite retrieval aim which is not changeable and it is called short-term interest mode. Because users scan less and record less in short period of time, it is quite hard to find out users’ short-term interest model.

Personalized retrieval can be conducted according to users’ different interest. Through modifying or filtering retrieval statements or retrieval results by means of users’ interest information, retrieval complexity can be effectively decreased and retrieval accuracy can be increased. As for the layer retrieval based on the theory of quotient space, retrieval complexity can be decreased and retrieval accuracy can be increased by means of a gradual thinning tree.

Users’ interest mode needs to be vectorizably handled in personalized layer retrieval. When the users’ interest mode is vectorizable, weight is users’ interests. Meanwhile, to make retrieval convenient, when user has several interest preferences, several vectors can be added and the result shows the user’s interest.

The core thinking of personalized layer retrieval based on user’s interest is that in layer retrieval, layered user’s interest is treated as equivalent class of corresponding layer. According to retrieval mode and user’s interest, weight value of each equivalent class can be calculated. The weight value of equivalent class is above the given threshold value, it is the ultimate retrieval field.

B. The Way of Obtaining User’s Interest Model

In order to quickly and accurately retrieval the needed information from oceans of data, people have made large amount of rewarding researches in the aspect of retrieval accuracy of retrieval system. In fact, personalized retrieval mode based on user’s interest mode is an effective method of improving precision ratio, whereas, it is the basis of personalized retrieval for the mining of user’s interest model.

Currently, internet websites all have certain layer structure. As is shown in Fig. 5, the highest layer is home (as opposed to home page). The child nodes are called inner nodes (as opposed to the first-level module/category). Inner node can also contain smaller inner node (as opposed to smaller module). The lowest layer is leaf node (as opposed to content page of specific information). We define root node as the layer 0, whose child node is layer 1. So, content page is the layer n.
As for the website of hierarchical structure, we consider each layer as a collection of several class nodes. For example, as is shown in Fig. 4, the nodes on the first layer are made up of n nodes (N_{i1}, N_{i2}, ..., N_{ia}), the nodes on the second layer are made up of b nodes (N_{i21}, N_{i22}, ..., N_{i2b}). When users visit leaf nodes (class nodes), each class node on the path is considered to be visited. Considering the nodes on higher layer, the visiting numbers are more because of their positions. Therefore, when users visit certain node, the feedback value of each node on the path shows exponential decrease layer by layer from bottom to top.

Essentially speaking, users’ interest mode reflects their scanning interest and habits. For the website of hierarchical structure, users’ interest mode also bears layer feature. And certain interest mode can only be expressed by the visiting sequence of the layer’s class node. Therefore users’ interest mode is different from the aspect of each layer. As is shown in the website in Fig. 1, for example, a visiting sequence of interest mode in leaf node layer is D_1D_{k+1}. The corresponding visiting sequence of layer 1 is N_{i1}N_{i2}, the visiting sequence of layer 2 is N_{i21}N_{i22}.

When users scan webpages to obtain information, they may first log on the home. Then they will scan N_{i1}, N_{i2} and the content page D_k. They also may open N_{i22} through other ways such as favorites. Then they scan content page D_k. No matter what is the visiting object, class node or content page, log records have the same format on the server.

V. ALGORITHM OF OBTAINING USER’S INTEREST

According to the website’s fixed hierarchical feature, we put forward a new obtaining method of multilayer users’ interest mode by means of ant colony algorithm [25-27]. This method can dynamically capture the latest users’ interest mode of various layers to satisfy the demand of layered user’s interest in hierarchical retrieval model.

The thought of ant colony algorithm roots from swarm intelligence when ants seek for food. Without the help of vision, ants can find the shortest path from food sources to ant nest through collaboration. In the process of foraging, ants may release as well as sense pheromone which is in proportion to the quality of food sources. When seeking foods, they tend to move toward the direction of higher concentration pheromone. Therefore, ants’ collective behavior shows us the phenomenon of positive feedback. The more ants some path is chosen, the higher probability the followers will choose it. So the food sources of higher quality and short distance will attract more and more ants. The pheromone density will increase more quickly. In order to apply ant colony algorithm to hierarchical pattern recognition of users’ interest, several parameters such as ant colony algorithm’s support, pheromone density and preference function etc. should first be defined:

Support degree η_l stands for the probability of users’ visiting on path l, which is represented by visiting frequency of this path. Suppose N_{ij} and N_{im} are two nodes in layer i. The path [(i, j, m)] represents some path from node N_{ij} to node N_{im}. The users’ visiting number on the path [(i, j, m)] is recorded as CI[(i, j, m)]. Then for the hierarchical interest pattern, the support degree η_l(i, j, m) is defined as:

\[
\eta_l(i,j,m) = \frac{C_l(i,j,m)}{\sum_{k=1}^{n} C_l(i,k)}
\]  

Pheromone density τ_l(t) stands for users’ visiting interest in some path l, which decreases as time passes by and increases as users visit to path l.

\[
\tau_l(t + 1) = (1 - p)\tau_l(t) + \Delta \tau_l(t)
\]  

In Eq.2, p is the mobilization factor of pheromone. δ is the constant which is related with pheromone density. Δ τ_l is pheromone value added.

\[
\Delta \tau_l(t) = \frac{F_{fb}}{t-1} - \tau_l(t)
\]  

where F_{fb} is the feedback value of the node which users visit at moment t+1.

Preference function P_l(t) represents the users’ preference to visiting path l, which includes pheromone and support level. Supposek(t) and η_k stand for corresponding pheromone and support level opposite to path K. Then preference function P_l(t) is defined as:

\[
p_l(t) = \frac{\tau_l(t)^{\eta_k} t_n^{\eta_k}}{\sum_{l=1}^{n} \tau_l(t)^{\eta_k} t_n^{\eta_k}}
\]

In Eq. 4, n is the total number of paths in this layer. If the threshold path node can be checked layer by layer according to Eq. 4. Then hierarchical interest model of group users can dynamically be obtained.

Considering that the interests of web users are gradually changing, the mining of users’ interest not only needs to find out their long-term interest but dynamically adjusts according to users’ interest. So the mining process is dynamic. Meanwhile, website bears hierarchical feature. Users’ interest bears hierarchical feature as well. Mining users’ interest from various layers can provide higher-level and more information of users’ interest and also can provide proper users’ information retrieval based on hierarchical retrieval method. Through comparing people’s information retrieval to evolving ant colony’s foraging action, dynamic obtainment of multi-layer users’ interest model can be better realized.

The basic thinking of the algorithm is as follows: comparing people’s information retrieval to evolving ant
colony’s foraging action, comparing once scanning process of various nodes on website to ant’s once foraging action period. All the actions of users’ scanning are recorded in log files, therefore we can get the value of preference function \( p(t) \) of all the path nodes layer by layer by means of ant colony algorithm. Then we can dynamically identify users’ interest mode quickly and accurately.

VI. SIMULATION AND ANALYSIS OF THE EXPERIMENT

Users’ interest can be divided into long-term and short-term. Long-term interest of users refers to the permanent and stable interest demonstrated in the process of long-term online process. Short-term interest of users refers to the interest in a relatively short period of time. Short-term interest changes quickly. Long-term interest is relatively stable. It can be separated based on web browsing activity of users.

In order to inspect the applicability of obtaining users’ interest, we conduct a simulation experiment on the accuracy of long-term interest and the sensibility of short-term interest change.

The process description of the algorithm is as follows:

Input: the threshold of the preference function is \( T_p \) after preprocessing m log records.

Output: internet pattern is \( \mathbf{H} \) at all levels on the website.

Step 1: Initialize the parameters of ant colony algorithm, \( t=0, \tau_i(t)=0, \eta_i=0, C_i=0 \). Initialize the structure tree of the website \( Tree=NULL \).

Step 2: Build up the website’s hierarchical structure tree “Tree” according to log records.

Step 3: Find out the path \( \mathbf{1} \) which is corresponding to \( N_i \) log record in Tree. Work out all the feedback values \( F_b \) of all the nodes on path \( \mathbf{1} \). Then update \( \tau_i(t) \) and \( C_i \). Repeat the operation above until processing all the records.

Step 4: Calculate all path nodes \( \eta_i \) according to Eq.1.

Step 5: Initialize \( \alpha \) and \( \beta \). Work out all the \( p(t) \) on \( N_i \) layer again. If \( p(t) \) is larger than the threshold value \( T_p \) of preference function, it is the interest pattern on each layer.

The experiment data adopt three groups. The training sample of the first group adopts visiting log of one month. The training sample of the second group adopts visiting log of one week. The training sample of the third group adopts visiting log of one day.

We conduct a simulation experiment on the data above. We only consider the first four layers of the website. The accuracy comparison of each layer’s users’ interest obtainment is shown in the following Fig. 6.

From Fig. 5, the obtainment of users’ long-term interest is more stable. Meanwhile, as the inspection layer is higher, the testing rate of accuracy is also increased. This situation is not at random but depends on the nature of users’ interest. The longer the time is, the higher the layer is. The hit rate of prediction is higher.

The other advantage of this method is that it may dynamically find out the interest changes of users. The following is a simulation experiment which shows the interest changes of users. The data are from the log files of certain company that record the users’ visiting traces of four products A, B, C and D. Over time, users continuously visit product A and demonstrate keen interest on it. But recently users don’t visit product A, which shows that users’ interest has slightly changed. Part of log records is as follows:

\[
../company/products/product_A/pageX/..
../company/products/product_A/pageX/..
../company/products/product_A/pageX/..
../company/products/product_A/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_B/pageX/..
../company/products/product_C/pageX/..
../company/products/product_C/pageX/..
../company/products/product_C/pageX/..
../company/products/product_C/pageX/..
../company/products/product_C/pageX/..
../company/products/product_C/pageX/..
../company/products/product_D/pageX/..
../company/products/product_D/pageX/..
../company/products/product_D/pageX/..
../company/products/product_D/pageX/..
../company/products/product_D/pageX/..
../company/products/product_D/pageX/..
\]

From Fig. 6, we may easily find that at the initial stage, the preference function value of product A node presents exponential growth because users continuously visit it. The preference function values of other manufacturers’ nodes are zero. In the latter half of the experiment, as the visiting number of product A decreases, the preference function values of other nodes increase because they are frequently visited by users especially product B. As a
result, this method may dynamically capture users’ interest pattern when their interest has changed.

VII. CONCLUSION

Because of a big amount of m-learning resources, the thesis puts forward the technique of learning resources integration based on granular computing. After studying relative methods of granular computing and quotient space, this thesis also puts forward an retrieval process of learning information based on quotient space. By means of retrieving keys to set up granularities of various layers, which are considered as initial granularities, equivalent divisions of various layers are carried out. Learning information resources integration is realized through organizing various types or different constructions of data in a certain way. Then introducing personalized concept into the system, the paper puts forward a personalized layer retrieval method based on users’ interest. Users’ information retrieval is opposed to ant colony’s foraging action. One-time scanning process of each node of the website is opposed to once ant colony’s foraging action period. According to users’ scanning log information, users’ interest mode can be dynamically identified. This method is easy to realize and can capture the short-term and long-term changes of users’ interest quickly and accurately.

ACKNOWLEDGMENT

The authors would like to acknowledge the financial support of the project of Educational Science the twelfth Five-year Plan of Jilin Province “The Theoretical Research and Practice of Educational Model based on Cloud Computing”.

REFERENCES