Study on Agricultural Information Push Technology Based on User Interest Model

Xiaorong Yang, Qingtian Zeng, Nengfu Xie, Lihua Jiang

Study on Agricultural Information Push Technology Based on User Interest Model

Xiaorong Yang[1,*], Qingtian Zeng[2], Nengfu Xie[1,*], Lihua Jiang[1,*]

[1] Agriculture Information Institute, Chinese Academy of Agriculture sciences, Beijing, P. R. China

*Key Laboratory of Agricultural Information Service Technology（2006-2010）, Ministry of Agriculture, The People's Republic of China

[2] Shandong Science and Technology University, Shandong Province, P. R. China

Abstract.: Traditional agricultural information systems lack availability because of ignoring user preferences. This paper proposes a user interest modeling method which combines auto-modeling with artificial modeling. The value of user interest measure can be obtained by counting the frequency of information accesses. The statistics method based on the heat of information accesses is adopted to perform cluster agricultural information push. The result of experiment proved the user interest modeling method can express a user's interest in agricultural information better. The intelligently and personalization of the system are improved.

Keywords: User interest model, Information push, Personalized service

1 Introduction

Traditional agricultural information systems focus on the integration and query optimization of information resources. The systems lack availability because of ignoring user preferences. Personalized services are the inevitable tendency of Internet. The personalized system can push satisfactory information resources to users actively by analyzing their personality and habits. Pablo Castells from Spanish Madrid university applied semantic-based personalized techniques to develop an extensible retrieval system which could automatically estimate the value of user personalization. Alexander Pretschner from Germany and Susan Gauch from America cooperatively studied an ontology-based personalized retrieval system. The system

could reorder and filter search results by matching search results and user character so that the search performance of the system was improved. Automatic Department of Tsinghua University presented a maximum distance-based Ranking algorithm to match information and developed the Bookmark system. The system adopted expanding the Bookmark function of the browser to record a user's access behavior and obtained a user's demand by analyzing his commentary. Professor Qingtian Zeng from Shandong Science and Technology University studied the obtaining technique of users' explicit and implicit knowledge needs. And he applied the technique to develop the E-learning system and multi-user interactive Question-Answering system.

Usual user interest model can be based on neural network, evaluation matrix, vector space model or ontology. However the neural network-based model cannot be readily intelligible and is now rarely applied. The poor adaptability of evaluation matrix-based model results in the difficulty in renewing a user's interest. The ontology-based model adopts multi-level domain knowledge to denote users' interest. At present building ontology is only manual or semiautomatical so that it takes too much time. The vector space-based model often results in deviation. So this paper integrates auto-modeling and artificial modeling to build user interest model. The value of user interest measure can be obtained by counting the frequency of information accesses. The statistics method based on the heat of information accesses is adopted to perform cluster agricultural information push.

2 User Interest Model

User interest model is a kind of user description which is algorithms-oriented and has a specific data structure. It can store and manage a use's background information and history behavior of accesses. It can record a user‘s interest points and describe his relatively inflexible information needs at a certain time. This paper integrates auto-modeling and artificial modeling to build user interest model. On one hand

In this paper, a combination of automatic modeling and artificial modeling construct user interest model, on the one hand, to record the user's behavior in the data access and mining user access logs to complete the modeling of the user, while allowing users to manually customize their own user model, which can be as effective as

possible, complete access to users' needs and to meet agriculture personalized information retrieval service requirements（Barrett,R. et .al，2007）.

2.1 Automatic Modeling

Automatic modeling by mining the log information can analyze the behavior of different users, and dynamically learn and improve in order to obtain the explicit needs and implicit needs of the user. For example, whenever the user to enter the agricultural science and technology information sharing system, user interest in a variety of behaviors will be recorded, including his clicking on the information classification of agricultural information resources, entering your search keywords, and browsing information. The order of these information classification accessed in one operation will be recorded such as A→D→E→C→D→B is a resource access sequence after a user login system which means that the user has accessed the system resources in the class A, D, E, C, D, B. There are many ways which can automatically create a user model. This paper uses a word frequency mining algorithms accessed high-frequency access in the recent time（Zeze Wu et .al, 2009）to establish the user interest model. The algorithm filters out the retrieval keywords which were most frequently used recently, and represents the user model with a weighted directed graph or a weighted vector model. The user interest model is shown in Figure 1. In order to distinguish different resources junction point between a user's interest points set and interest vector set, a set of rectangular nodes with weights is used to represent a user's interest point set and a directed graph of circular nodes with weights is used to represent a user's interest vector set.

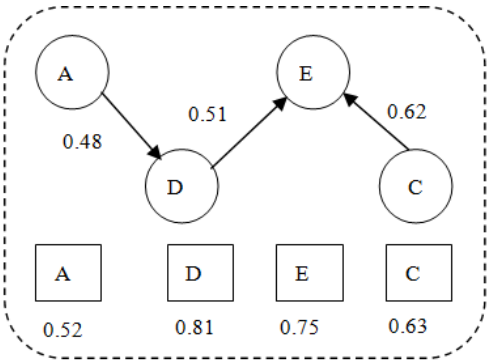

Figure 1 User Interest Model Map

2.2 Artificial Modeling

On the basis of the automatic modeling, by personalized mapping the fields of data sheet in an agriculture database, each user can customize his favorite search criteria of agricultural information resources so as to meet the requirements of personalization. Besides, a user can improve the user model by filling in his personal information such as gender, age, occupation, education, personal interests and so on. Thus the system can cluster users so as to improve the user model.

3 Push Technology of Agricultural Information Resources based on User Demand-driven

3.1 User Interest Statistics based on the Number of Access

As a traditional user interest statistical method, the principle of user interest statistics based on the number of access is that the more a user is interested in some information resource, the more frequently he access the resource. The resource tree is displayed to users at agriculture database granularity. When a user accesses the resource R of a database D in a time TS, Interest (D) of the database D increases. Combined with the user's access frequency Freq (d), the interest of the user on the database D can be gotten according to the following method（Xiaomin Ying,2003）:

$$I=Interest(D)*Freq(D) \quad (1)$$

Interest (D) represents a user's degree of interest for database D. And Freq (D) represents the frequency of a user access to the database D.

But this method simply focuses on the number and frequency of a user's access to resources without the attenuation of the user's interest with time. On this basis, this paper adds time decay for the user interest. Time-attenuation coefficient α can be gotten according to the following formula:

$$\alpha=10^{-\ passedTime/secsInMonth} \quad (2)$$

PassedTime represents elapsed time. And secsInMonth is a constant which is 30 days a month, that is 2592000 seconds.

User's actual interest can be derived according to the following formula（Yong Li,2002）:

$$I'=I*\alpha \quad (3)$$

User interest calculated according to the above algorithm corresponds to the resources tree which is displayed after a user logins system. Database accessed more frequently will be enhanced its sort. A user's interest in agricultural information resources can be counted step by step and information push can be achieved implicitly.

3.2 Agricultural Information Resources Push Based on the Heat of Access

For groups of users, sometimes a certain category of information resources is generally concerned. For example "the price of wheat today " is closely related to life. During individual users are searching, this kind of information is easily ignored due to different interest areas of users or different concern in current time. But the information is actually helpful for the user. By mapping multiple users' interest for an information resource, the weight of the information resources can be calculated:

$$I(r)=\sum I(i)* \alpha，\ i=1,2,3\ldots n \qquad (4)$$

According to the weight, the sort of information resources produces the hottest resources sequences. When he retrieves in the agricultural science and technology information database, a user can get not only matching results but also pushed information which is the most accessed by other users. Pushed information may be useful for him. Thus clustering information resources push is achieved.

In view of a larger amount of user group's access, a certain threshold should be set to ensure the effectiveness and timeliness of agricultural information resources recommended. When the access weight of certain type of information resources within a certain time limit T can not reach the specified experience value e, the information resources will be removed from the recommended resource queue to ensure the efficiency and timeliness of information resources recommended queue.

4 The Experimental Results

According to the above techniques and methods, the intelligent retrieval platform of agricultural science and technology information based on user interest model was designed and developed. The platform adopts the user interest model to obtain users' need preferences and pushes the interested information to them. Not only the information meeting the search keywords is retrieved, but also the information which

contains the synonyms and related information of the search keywords is also retrieved. Retrieval results are displayed to the user in accordance with his interest priority. By sorting resources dynamically based on user interest, customizing personalized fields and recording accessed information, personalized service can be achieved. As the log of user behavior, user history records are very important for building user model, speculating user behavior and mining user interest. Because user behavior is difficult to speculate and users ' history data are more complicated, a user can not get effectively feedback when he checks his history data. This platform uses a uniform approach- "unified recording, classify browsing" to integrated into users' access records to the user model. The platform collects a user interest in the classification of resources based on each user's behavior. Whenever a user accesses a resource, the system will record his behavior. When he login the system next time, the system will automatically display the frequently accessed database resources to the top row.Besides, he can define the fields in the database as own favorite language or vocabulary in order to make retrieval conditions more in line with his habits. He can also operate his own accessed records to determine whether they are useful. By analyzing the accessed information resources, the system can obtain user interest for a certain type of information resource and improve user interest model. The example showed that the proposed user interest model is better able to push useful resources for users.

5 Conclusion

This paper used an agricultural information push method based on user interest model and presents the combination of automatic modeling and artificial modeling to build user interest model. Further the study proposed the user interest statistics techniques based on the number of access and agricultural information push based on the heat of access. According to the above techniques and methods, the intelligent retrieval platform of agricultural science and technology information based on user interest model was designed and developed to verify the effectiveness of the proposed method.

Acknowledgements

The work is supported by the special fund project for Basic Science Research Business Fee “The demonstration of Tibet agricultural information personalized service system”, AII (No. 2012-J-08).

References

1. Xiaorong Yang. Study and Application of Key Technologies for Distributed Agricultural Science and Technology Information Sharing. [Doctoral Dissertation]. Bejing: Graduate School of Chinese Academy of Agricultural Sciences，2011.
2. Barrett,R.，Maglio,P.P.，Kellen,D.C. How to Personalize the Web. Proceedings of the Conference on Human Factors in Computing Systems (CHI’97)，ACM Press，2007.
3. Zeze Wu，Qingtian Zeng，Xiaowen Hu. Mining Personalized User Profile Based on Interesting Points and Interesting Vectors. Information Technology Journal，2009，6（8）: 830-838.
4. Xiaomin Ying . Study on User Modeling Techniques for Internet personalized service. [Doctoral Dissertation]. Changsha: National University of Defense Technology, 2003.
5. Yong Li . Study and Application of Ontology-based Personalized User Modeling Techniques in Intelligent Retrieval. [Master Dissertation]. Changsha: National University of Defense Technology, 2002.