A Method of Mining User’s Interest in Intelligent e-Learning

Xiyuan Wu*, Ping Wang, and Min Liu

Abstract—A key problem of personalized Web applications is how to recognize accurately a user’s interests. This research is to investigate approaches that implicitly build user interest models and propose a new method to mine and describe user’s stable interests. It generates the position and degree of interest of users to intelligently adjust the presentation and content of web-learning resources to users’ needs and realize the intelligence of web-learning environment.

Keywords—User Model, User Interest, e-Learning, Web Mining

I. INTRODUCTION

Due to the information explosion in nowadays, to provide the information which users have interest in is a high-priority issue involved in the development of Internet and WWW. The visiting data of users include the interest pattern of users[1], so users’ interest can be intelligently mined by analyzing the visiting data.

Interest is positive tendency in human recognition of objects. It is only when this tendency of recognition is stable that human interest can be formed. Learning interest is a kind of interest, namely the interest carrying positive emotion and promoting students to learn more. Learning interest plays an important role in the maintaining of students’ motivation, improving the proactivity in learning, and the enhancement of learning efficiency.

The discovery of users’ interest in intelligent network learning environment is to automatically recognize the current interest model of users and to find out the spontaneous interest of users. However, the spontaneous interest possible can not facilitate user in achieving certain learning objective. Therefore, it is important to find a way to provide guidance, adjustment and incentives to users’ interest and to generate the position of interest which facilitates users’ learning and should be visited by users but can not be gained by users simply relying on their own knowledge. This article aims at proposing a method to intelligently recognize users’ learning interest in intelligent network learning environment to solve this problem.

The modeling of users’ learning interest is basis of recognition of users’ learning interest.

There are two solutions to build user interest model: explicit and implicit. However, the explicit approach involves the user, takes time and effort, and user interest may change over time. This research focuses on implicit approaches that capture user interests without the user intervention and reflect changing user interests.

According to the sources of user data, implicit user modeling approaches can be based on a user’s behavior or the contents of web pages that a user has visited, or both.

A user behavior based approach observes a user’s action such as click, visiting retention time, times of visit, navigation path, action of saving, editing, revising, downloading, and the key words input in search engine, etc. This technique can obtain common user profiles based on association rule discovery and usage based clustering etc. However, as the association rule discovery is based on the web pages already visited, it can only recommend those web pages which have been visited by old users to new users whereas can not recommend those unvisited web pages.

A content based approach analyzes the contents of web pages that a user has visited. Kim et al. propose a divisive hierarchical clustering approach to build user interest hierarchy model that can be learned from the contents of web pages bookmarked by a user[2]. This approach can only recommend the resources similar to those resources which have been visited by users whereas some researchers have found that users consider the unexpected more valuable.

A hybrid approach observes user’s behavior and the contents of web pages visited by a user. Trajkova and Gauch build user profile based on concepts from a predefined ontology[2][3]. Tan Qiong etc. use the local autonomous agents to percept user’s action and adopt a learning algorithm to get user interest profile[4]. These techniques show that the usefulness and accuracy of the resulting recommendations are increased[5].

This article uses method of invisible tracking, which does not require users to provide information and recognize users’ interest by system automatically inspecting using actions.

II. METHOD OF RECOGNITION OF USERS’ INTEREST

A. Model of users’ interest

Users’ interest is expressed in the form of weighted key words vector, including the interest in single knowledge point and the ranking of interest in visiting several knowledge points.

The interest in single knowledge point is expressed as below:
User _ Interest = (UID, knlElemID, InterestGrade) In which:

UID is the user’s ID;
knlElemID is ID of knowledge point;
InterestGrade is UID’s degree of interest in the knlElemID.

If user’s visiting time exceeds a certain threshold value, then it is supposed that user has interest in the content visited. The threshold is determined by previous experiences.

The computing of degree of interest involves three parameters: the frequency of visiting, last time visited, and the retention time of visiting. The longer of retention time of visiting, more of frequency of visiting, the nearer of last time visited to current time, the higher of user’s interest in one knowledge point. Degree of interest is in proportion to retention time of visiting and frequency of visiting and in inverse proportion to the difference between last time visited and current time.

B. Interest pattern

Interest pattern is the interest of certain user, namely assigning a certain figure to user’s interest, including two parts: one is interest in single knowledge point, expressed as: User _ Interest = (UID, knlElemID, InterestGrade) the other is ranking of interest, expressed as:

User _ Interest1 → User _ Interest2

→ User _ Interest3

C. Processing flow

The overall processing flow can be expressed as Fig. 1.

Fig. 1 Overall processing flow

The updating of users’ model can be classified as updating in real-time and background updating. The updating in real-time of users’ model is performed during the log-in time of users; and the background updating of users’ model is performed after users have logged off.

The detailed processing flow and algorithm can be described as below:

1) The users’ model form is scanned when one user logs in. If this user can not be found in the users’ model form, then this user’s ID would be added to the form with all other attributes expressed in the value of zero; if this user is found in the form, than the step 2) would be performed.

2) Select the knowledge point with the highest degree of interest as the search criteria, which would be proposed to search model and the search result would be forwarded to user by search model.

3) After user logging off, the background updating would be started to process the data in user log in order to update the degree of interest in knowledge points in user’s model. The detailed processing flow is as below:

   Step 1: Rank the resources visited by user according to <duration of visit> and <the time when resource is visited>;
   Step 2: Select the top-N resources;
   Step 3: According to the results computed by the following formula, the degree of interest in all knowledge points is updated. In the formula, Kweight is the degree of interest in knowledge point, Ninter is the time of clicking on knowledge point, Linter is the retention time on knowledge point, Tinter is the time difference between the time when knowledge point is visited and current time.

\[ Kweight = Ninter \times Linter / Tinter \]

By processing described above, the user’s learning interest model can be set up to guide user’s learning. When log on next time, a user would enter the automatically showed resource page suitable for his or her interest instead of the default home page. If there is any new resource suitable for his or her interest, this resource can be directly recommended to him or her in order to suit his or her individual needs.

III. EXPERIMENT

We used the undergraduate students’ on-line English learning actions in Xi’an Jiaotong University as source of data and collected 10471 learning log records of undergraduate students in the School of Electronic and Information Engineering, the School of Energy and Power Engineering and the School of Human Settlement and Civil Engineering from on-line learning system.

On-line English learning includes seven parts: reading, listening, writing, translation, verbal and grammar. Each part involves a great amount of learning materials. It can be inferred from one user’s visiting frequency and retention time on web pages if the user is interested at the content of web pages or not.

The form of log record is as below:

<table>
<thead>
<tr>
<th>UserID</th>
<th>LogIP</th>
<th>Intime</th>
<th>OutTime</th>
<th>DataType</th>
<th>DataID</th>
</tr>
</thead>
</table>

UserID is learner’s ID name used when registering in the on-line learning website. LogIP is the ip address when user logs in. Intime and OutTime are respectively the time user starts and finishes reading one article. DataType is the code of article type, which represents one of the seven parts mentioned above. DataID is the serial number of the article read by user.

After analyzing 10471 learning log records, we have obtained the interest data of 318 users, some of which are listed as TABLE I.

The first column is the learner’s ID name used when registering. The data in column 2 to column 8 respectively
represent users’ degree of interest in each of the seven parts. We performed normalization on the computing results of the formula for comparison.

<table>
<thead>
<tr>
<th>UserID</th>
<th>reading</th>
<th>listening</th>
<th>speaking</th>
<th>writing</th>
<th>translation</th>
<th>verbal</th>
<th>grammar</th>
</tr>
</thead>
<tbody>
<tr>
<td>04032002</td>
<td>.74367625</td>
<td>.04005059</td>
<td>.037942664</td>
<td>.019392917</td>
<td>.06576729</td>
<td>.093591906</td>
<td></td>
</tr>
<tr>
<td>04041098</td>
<td>.26397887</td>
<td>.35205182</td>
<td>.0004799616</td>
<td>.38300937</td>
<td>.0037942664</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>04041200</td>
<td>.40756914</td>
<td>.32605532</td>
<td>.00189229</td>
<td>0</td>
<td>0</td>
<td>.2561863</td>
<td></td>
</tr>
<tr>
<td>04055050</td>
<td>.83224523</td>
<td>.00407375</td>
<td>.002820285</td>
<td>.006006163</td>
<td>.00208910</td>
<td>.15250431</td>
<td></td>
</tr>
<tr>
<td>04055079</td>
<td>0</td>
<td>0</td>
<td>.000527473</td>
<td>0</td>
<td>0</td>
<td>.93705493</td>
<td></td>
</tr>
</tbody>
</table>

We performed statistic analysis to find out the head count of users’ with highest degree of interest in each of the seven parts in those 318 users and the result is as below:

![Figure 2 Analyze result](image)

The figure above shows that most learners (125 head count) were interested in reading part and the next part with most people having highest degree of interest in it was listening. Grammar is the part with least people having highest degree of interest in it. We can infer from this result that the learners’ interest in different parts of English learning varies from learner to learner, which has proved the validity of our model.

IV. CONCLUSION

This article discusses the discovery of users’ interest in web-learning environment and proposes a method to automatically recognize users’ interest by analyzing users’ learning log, which generates the position and degree of interest of users to intelligently adjust the presentation and content of web-learning resources to users’ needs and realize the intelligence of web-learning environment.

In our research, the threshold value, parameters and weights in formula are all determined by previous experiences and our next step is to research on the selection strategy of self-adapting. Besides, users’ visiting actions are various, such as clicking, dragging scroll bar, searching, adding to Favorite and so on. So it is necessary to find out which actions are essential and the relation among these actions in order to increase the accuracy of interest recognition. For instance, the action of simple clicking does not reveal users’ interest effectively and it is necessary to take into account of the retention time, moving time and clicking times of mouse on web page, time of dragging scroll bar etc.

REFERENCES


http://dx.doi.org/10.1016/j.eswa.2006.08.034