PREDICTING USER INTEREST FROM MOUSE DATA

Akshay Bhargava  
Jaypee Institute of Information Technology  
Sector 128, Noida  
bhargavaakshay07@gmail.com

Anuradha Gupta  
Jaypee Institute of Information Technology  
Sector 128, Noida  
anuradha.gupta@jiit.ac.in

Vartika Srivastava  
Jaypee Institute of Information Technology  
Sector 128, Noida  
vartika.srivastava@jiit.ac.in

ABSTRACT

Several studies show the correlation between user satisfaction and webpage interface, with the help of mouse behavior on the webpage because mouse data is very useful for identifying user’s interests [1]. There are different mouse behavior patterns like clicking, reading, hesitating, scrolling, etc. and each of these have a different meaning. In this work, we have created a ranking model for identifying the user interests on a webpage. Study has been conducted on 17 users and analyzes their mouse activity’s data. Using these data, we identified their top two interests on the page. In this work 65% user’s results matched correctly with their respective top 2 interests. This new technique can help in making use of mouse activity as a better approach to identify user interests on a webpage.

General Terms

Human Factors

Keywords

Mouse Activity; User Activity; User Interest; Web Interaction; Ranking Model; Weighted Model; Normalization.

1. INTRODUCTION

Based on a recent survey 40% of total world population has access to internet1. In this internet world, the challenge for e-commerce website is to satisfy its users. To make the webpage which fulfill the user’s expectations and make the user contented, we need to take the user feedback, but it is difficult to take user’s feedback manually. Many researchers are working to collect user’s preferences and expectations automatically without bothering the user [2]. Previous studies shows that with the help of eye tracking [3], we can predict the user’s interest. Another method to know the user’s perception is the mouse activity and path of cursor movement. Researches show that use of mouse hovering and scrolling technique play a very crucial role in user interest tracking [4, 5].

In our work, we are recording the user mouse activity on a webpage and identified user’s interest. Many researchers have been done which receives the text query from the user to know their interest and show the interested data on top of the webpage on next visit [6]. But in our work, the primary focus is to interpret user’s mouse cursor behaviors on the webpage and infer their interest. To make such a reliable system, our webpage capture user interaction [7] and consequently propose recommendations to the user, based on the analysis of previous activities of users. We are considering mouse hovering, scrolling, clicking and hovering time to capture the cursor movement of user on webpage.

2. METHODOLOGY

We have designed an E-commerce website with categories: Electronics, Mobiles, Fashion, decor & furnishing, kitchen appliances, movies etc. Each of these categories has been divided into 6 sub-categories. In all we have 24 categories for the users from which they can choose in what they are interested.

2.1 Feature Sets

Following are the features that have been used to predict the user interests

- Number of clicks for each category where user within 2-6 seconds after page is loaded completely. We assume that 6 seconds are enough for the user to scroll through the page and select a category in which they are most interested if they are sure of buying it.
- Number of clicks for each category where user clicks on it after 6 seconds from loading of the page
- Cursor hovering time for every category and the total average hovering time for particular category calculated by the formula

\[
\text{Average Hovering Time} = \frac{\text{Total time hovered}}{\text{No. of times hovered}}
\]

- Clicks within first two seconds after loading of web page are considered fake as earlier studies have shown that such small time is not enough to predict interests of the user [3].
- If the mouse is idle for more than 5 seconds, we assumed that user is not active and this should not be considered as mouse activity.
- When user is just scrolling up and down, we assume user is not pointing to any particular category, so movements related to scrolling up and down have been removed from our feature set.

2.2 Normalization and Ranking

We have normalized each feature weight to obtain the values for each category ranging from 0 to 1. Normalization is done by dividing values of each category feature by the sum of values of all categories feature for that particular feature. Once normalization is done, total weight of each category for the user is calculated. This is done by assigning weights to each feature considering two factors:

- Importance of that feature in inferring user interest
- Possibility of happening of that feature

1 http://www.internettlivestats.com/internet-users/
The weights assigned are: ‘0.3’ to clicks within 2-6 seconds after loading of the page, ‘0.56’ to clicks after 6 seconds and ‘0.14’ to mouse hovering. The reason for ‘0.3’ weight to clicks in 2-6 seconds was that these clicks may show maximum interest towards a category but chances of clicks within 6 seconds are also very rare until the users are totally confident of what they have to select. After multiplying these respective weights with the normalized values that we calculated for every category and adding them, we get a total weight for every category considering all the features.

The categories with the highest weights are considered as most interesting and whenever the user will be logging in again, these categories will be displayed to them at the top of the webpage. This process of knowing interests will be continued for every time user logs in.

3. EXPERIMENT AND RESULT

A web page model is created for a statistical analysis of the proposed solution. For experimental setup shown in Figure 1, we asked 17 users, who were familiar to computers and shopping websites to use the website as per their interest. All of these users were college students of age group between 19-22 years and the reason behind considering them was that in today’s lifestyle, frequency of college student using shopping websites is much higher than any other group and they have a much better knowledge of how to use a shopping website. We recorded all the cursor activity of the users and saved them in the database.

To confirm the findings, participants were asked about the top two categories in which they were interested while using the website and these were manually noted in a separate logbook. When the data in the logbook was compared with the top two results we got as shown in Figure 2, we found that out of 17 users we were able to identified 11(65%) user interests correctly, 3(17%) users first interests were calculated correctly and 2(12%) user’s interests were interchanged with first and second priority interest.

4. CONCLUSION

In this paper, we have shown that cursor movement can also play an important role on determining the user interests web page. We conducted a user study to analyze cursor movements and testify the user interests. To find out the user interest we captured user’s mouse clicks and mouse hovering data. After that, we normalized the data value and combined with ranking model to get the final result. Experimental result shows that 65% user interests were identified correctly. In future work, we will simulate the proposed solution.

REFERENCES


