Insight into User on E-commerce Search

dgliu

September 20, 2018

Part I
Data Preprocess and Statistics

We use the following procedure to generate the experimental data used in this paper. On the data generated in the first two steps, the distribution of session number per user is shown in Fig 1. We determine the threshold and filter out abnormal users in third steps. The statistical information of the data set can be found in Tab 1.

(i) Collect all query and action data from 20180726 to 20180804 and remove invalid data.

(ii) Divide the session by determining whether the time interval between adjacent queries is greater than 30 minutes.

(iii) Remove user with more than 16 sessions.

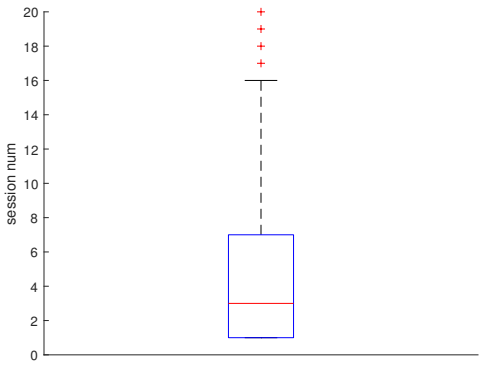

Figure 1: The Box plot of user session number

The question we are interested in is whether users follow certain commonalities in query. To answer this question, we first start with the length of the

Table 1: The statistics of the data set

Timestamp	20180726-20180804
Session	985,277,625
User	251,995,143
Item	214,904,724
First level category	161
Leaf category	14,361

query and the action, then analyze the modification pattern of the query, and finally give the relationship between the query and the action pattern.

Part II
The Length

Considering the characteristics of query in E-commerce, we give the distribution of session length, click length and the number of words and fit them in turn. We will also explain the reasons behind the phenomenon with the optimal distribution.

The length distribution plots for Session number are presented in Fig 2. The length follows distribution is one that is like a log-normal at low and intermediate values, with a characteristic peak and turnover, but transitions to a power-law distribution at high values. Similarly, we also observed that the length distribution for Click number follows a modified log-normal with a power-law (MLP) distribution in Fig 4. This shows that in e-commerce queries, users want to capture the target through fewer queries or clicks. But there are still small groups that are willing to spend more time inquiring and clicking to find satisfactory products.

From Fig 3, we can see that the distribution of word length follows a log-normal distribution. Because users usually adopt query pattern that (modifiers, target), and often not too long.

Part III
The Modification

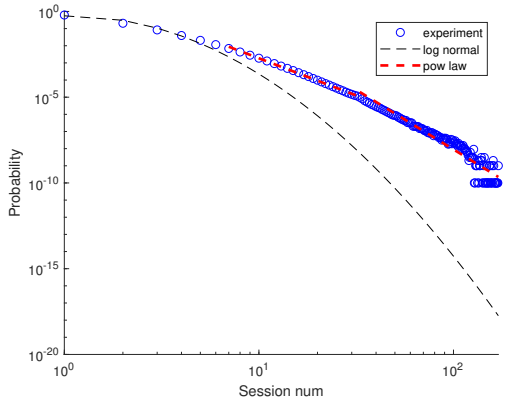

Figure 2: The distribution of session length

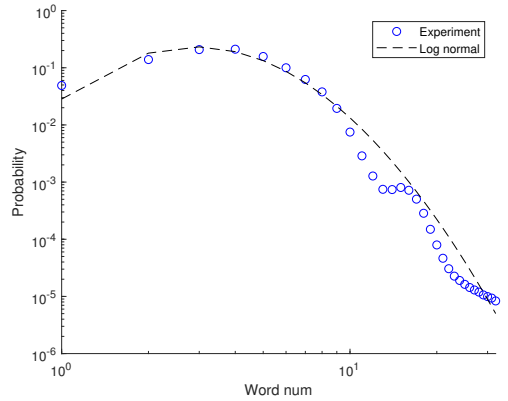

Figure 3: The distribution of word length

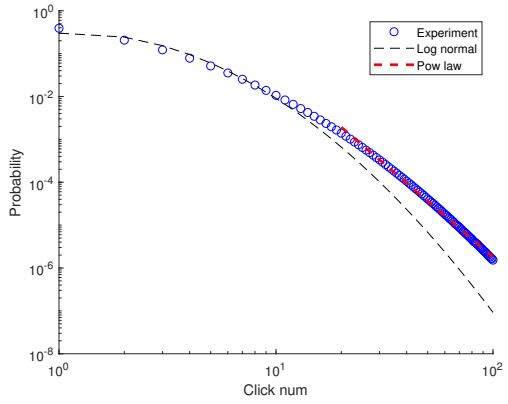

Figure 4: The distribution of click length