Markov Modeling for User Interaction in Retrieval

Vu T. Tran
Information Engineering
University of Duisburg-Essen
Duisburg, Germany
vtran@is.inf.uni-due.de

Norbert Fuhr
Information Engineering
University of Duisburg-Essen
Duisburg, Germany
norbert.fuhr@uni-due.de

ABSTRACT

For applying the interactive probability ranking principle (IPRP), we derive Markov models (MM) from observing interactive retrieval via system logging and eyetracking. Then we discuss various applications of these models: 1) For time-based retrieval measures, we can derive the expected performance based on the MM. 2) By varying single parameters of the model, it is possible to simulate the effect of specific system improvements and its consequences on retrieval quality. 3) Based on the interactive PRP, the system can order the choices offered to the user in an optimum way, and guide the user to more successful searches. 4) While current approaches for simulating interactive IR assume a simplistic, deterministic user behavior, the MMs derived empirically can form the basis for more realistic simulations.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Search Process*

Keywords

Evaluation, Interactive Retrieval, Simulation, User Behavior

1. INTRODUCTION

The Interactive Probability Ranking Principle (IPRP) [1] is a probabilistic framework model for interactive retrieval. This model assumes that a user moves between situations s_i, in each of which the system presents a list of choices, about which the user has to decide, and the first accepted choice moves the user to a new situation. Each choice c_{ij} is associated with three parameters: the effort e_{ij} for considering this choice, the acceptance probability p_{ij}, and the benefit a_{ij} resulting from the acceptance.

Based on the IPRP we developed a new methodology for analyzing interactive IR [5] using log and eyetracking data from the INEX 2010 interactive track [3] (12 retrieval sessions, 84 queries). Based on this data, we represent the user's interaction as a Markov model (MM, see (Figure 1). After formulating a *query*, the user looks at one *result item* after the other, possibly regards its *details* and puts items found relevant into the *basket* (for further explanation on our interface, see [5]). The timings correspond to the effort e_{ij} for evaluating a choice c_{ij}, while the transition probabilities give the chances p_{ij} of accepting it. As a possible approach

SIGIR 2013 Workshop on Modeling User Behavior for Information Retrieval Evaluation (MUBE 2013), August 1, 2013, Dublin, Ireland.

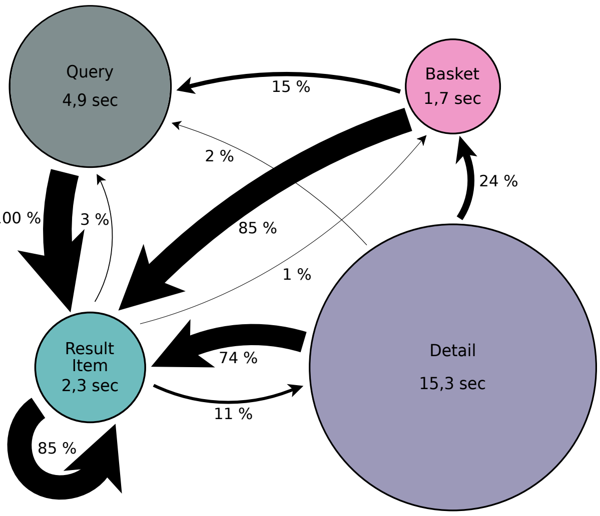

Figure 1: Transition probabilities and user efforts

for quantifying the benefit a_{ij} of a decision, we regard the time needed for finding the first (next) relevant document (see below). With some limitations, the same methodology can also be applied when only log data is available, e.g. for the 2011 TREC session track[1] (see the results in the next Section).

2. EVALUATION

Given the MM, we are able to estimate the time for finding the (next) relevant document. In turn, these values can be used for estimating retrieval quality in terms of time-based measures [4].

Let us denote the four MM states by q, r, d and b, the effort in these states by t_q, t_r, t_d and t_b, and the transition probability from state x to state y as p_{xy}.Then the expected times T_q, T_r and T_d for reaching the basket state from the query, results or details stage, respectively, can be computed via the following linear equation system:

$$\begin{aligned} T_q &= t_q + p_{qr}T_r \\ T_r &= t_r + p_{rq}T_q + p_{rr}T_r + p_{rd}T_d \\ T_d &= t_d + p_{dq}T_q + p_{dr}T_r \end{aligned}$$

As results, we get the time T_q for the query, T_r for the

[1] http://trec.nist.gov/data/session.html

result list and T_d for the details stage, which are shown in Table 1.

	INEX	TREC
T_q	122.7	285.4
T_r	117.8	255.2
T_d	104.9	204.2

Table 1: Times (in seconds) to basket/relevance for INEX and TREC

Here we have assumed that all result items are of equal quality, i.e. p_{rd} and p_{rr} are constant. In order to consider the decreasing quality of result items, we have to refine the MM by introducing a state r_i for each result item, along with a state d_i for the corresponding details. From our observation data, we derived statistics about rank-dependent click-through rates $p_{r_i d}$ (and the changing values of $p_{r_i r_{i+1}}$). This leads to the times displayed in Table 2, based on the following equation system:

$$\begin{aligned}
T_r &= T_{r_1} + \ldots + T_{r_n} \\
T_{r_1} &= t_r + p_{r_1 q} T_q + p_{r_1 r_2} T_{r_2} + p_{r_1 d} T_{d_1} \\
T_{d_1} &= t_d + p_{dq} T_q + p_{dr_2} T_{r_2} \\
T_{r_2} &= t_r + p_{r_2 q} T_q + p_{r_2 r_3} T_{r_3} + p_{r_2 d} T_{d_2} \\
T_{d_2} &= t_d + p_{dq} T_q + p_{dr_3} T_{r_3} \\
\ldots &= \ldots
\end{aligned}$$

T_q	T_{r_1}	T_{r_2}	T_{r_3}	...	T_{r_9}	T_{r_10}
122.7	117.8	117.9	119.1	...	122.5	123.5

Table 2: Times to basket/relevance for INEX with varying probabilities

3. SIMULATION

By varying the MM parameters, we can simulate the effect of system changes. As a first example, we consider the effect of ranking quality. As Figure 1 shows, only roughly 4% of the items in the result list are relevant ($p_{rb} + p_{rd} p_{db}$). Improving ranking quality would result in increasing the value of p_{rd} and p_{rb}, while p_{rr} would decrease by the same amount (assuming that p_{rq} remains unchanged). Then we can derive that retrieval improvements by 10%, 20% or 30% would reduce the time to basket T_q by 17%, 27% and 31% respectively. This shows that even small improvements in terms of ranking quality would have a big effect on the quality of interactive retrieval (while enhancements beyond 30% would have only minor effects). In a similar way, we can e.g. simulate the effect of improving the quality of the result summaries.

Besides studying the effects of small changes to an existing system, the MM derived from such a system could also be used for simulating interactive retrieval with other systems (see e.g., [6]). Currently, these simulations are based on a deterministic user behavior, and thus the results give only upper and lower bounds for retrieval quality, but hardly estimates of the average performance. Based on the MM, we now have an empirical basis for simulating the more realistic stochastic user behavior. Of course, this would require a large number of simulated runs, in order to implement the stochastic behavior represented by the MM. In addition, one would need full relevance information for the information needs studied. Also, we still need a model for query reformulation, which is still a research issue. Nevertheless, a stochastic user model is clearly the way for more realistic simulations of interactive retrieval.

4. GUIDANCE

The core idea of the IPRP is the optimum ranking of choices. In our simple MM studied here, the only choice the user has is the question when to reformulate a query instead of going to the next result item. Looking at the figures in Table 2, one can see the expected time for reaching the next relevant document increases as the user goes down the ranked list of result items. Comparing the values T_{r_i} with T_q it becomes apparent that it does not pay off to go beyond the ninth result item—the user should rather formulate a new query.

However, we should bear in mind that these figures only refer to the average case. Since there is a large variation in the ranking quality of queries, we need query-specific estimates. This requires methods for estimating the probabilities of relevance for a given query, along with the corresponding click-through rates—which is an issue of our current research.

5. OUTLOOK

In this paper, we have shown how the application of MMs to interactive retrieval can be used for evaluation, simulation and user guidance. However, the models regarded so far are very simple. For example, we assume that the quality of the queries is constant, i.e. reformulation does not lead to better answers. Moreover, the possible user interaction is rather limited (e.g., the system could suggest query expansion terms). Considering changes from query to query or new interaction possibilities increases the number of states in the MM significantly; as a consequence, estimating the increased number of parameters requires substantially more observation data, thus more extensive experimentation or access to logging data of operational systems.

6. REFERENCES

[1] N. Fuhr. A probability ranking principle for interactive information retrieval. *Information Retrieval*, 11(3):251–265, 2008.

[2] W. R. Hersh, J. Callan, Y. Maarek, and M. Sanderson, editors. *The 35th International ACM SIGIR conference on research and development in Information Retrieval, SIGIR '12, Portland, OR, USA, August 12-16, 2012.* ACM, 2012.

[3] N. Pharo, T. Beckers, R. Nordlie, and N. Fuhr. Overview of the inex 2010 interactive track. In *INEX*, Lecture Notes in Computer Science, pages 227–235. Springer, 2011.

[4] M. D. Smucker and C. L. A. Clarke. Time-based calibration of effectiveness measures. In Hersh et al. [2], pages 95–104.

[5] V. T. Tran and N. Fuhr. Using eye-tracking with dynamic areas of interest for analyzing interactive information retrieval. In Hersh et al. [2], pages 1165–1166.

[6] R. W. White, I. Ruthven, J. M. Jose, and C. J. V. Rijsbergen. Evaluating implicit feedback models using searcher simulations. *ACM Trans. Inf. Syst.*, 23(3):325–361, July 2005.