Click models cheat sheet

Terminology

Variables:

- E: a user examines an object on a SERP;
- A: a user is attracted by the object's representation;
- C: an object is clicked; and
- S: a user's information need is satisfied.

Expression	Meaning
u	A document (documents are identified by their URLs, hence the notation).
q	A user's query.
r	The rank of a document.
c	A placeholder for any concept associated with a SERP (e.g., query-document pair, rank, etc.).
s	A user search session.
$\mathcal{S}$	A set of user search sessions.
$\mathcal{S}_c$	A set of user search sessions containing a concept c.
u_r	A document at rank r.
r_u	The rank of a document u.
N	Maximum rank (SERP size); usually equals 10.
X	An event/random variable.
x	The value of a random variable X.
X_c	An event X applied to a concept c.
x_c	The value that a random variable X takes, when applied to a concept c.
$x_c^{(s)}$	The value that a random variable X takes, when applied to concept c in a particular session s.
$\mathcal{I}(\cdot)$	An indicator function.

Evaluation

Log-Likelihood

$$\mathcal{LL}(M) = \sum_{s \in \mathcal{S}} \sum_{r=1}^{n} \log P_M \left(C_r = c_r^{(s)} \mid \mathbf{C}_{<r} = \mathbf{c}_{<r}^{(s)} \right), \tag{1}$$

where P_M is the probability measure induced by the click model M.

Perplexity

$$p_r(M) = 2^{-\frac{1}{|\mathcal{S}|} \sum_{s \in \mathcal{S}} \left(c_r^{(s)} \log_2 q_r^{(s)} + \left(1 - c_r^{(s)}\right) \log_2 \left(1 - q_r^{(s)}\right) \right)}, \tag{2}$$

where $q_r^{(s)}$ is the probability of a user clicking the document at rank r in the session s as predicted by the model M, i.e., $q_r^{(s)} = P_M(C_r = 1 \mid q, \mathbf{u})$.

Click Models

Random Click Model (RCM)

$$P(C_u = 1) = \rho. \tag{3}$$

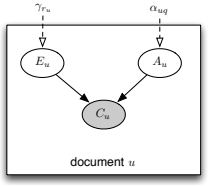
Rank-based CTR Model (RCTR)

$$P(C_r = 1) = \rho_r. \tag{4}$$

Document-based CTR Model (DCTR)

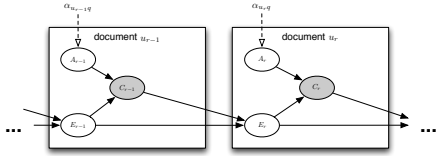
$$P(C_u = 1) = \rho_{uq}. \tag{5}$$

Position-based Model (PBM)



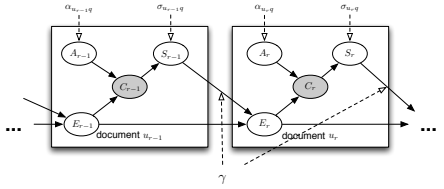
$$P(C_u = 1) = P(E_u = 1) \cdot P(A_u = 1) \tag{6}$$
$$P(A_u = 1) = \alpha_{uq} \tag{7}$$
$$P(E_u = 1) = \gamma_{r_u}. \tag{8}$$

Cascade Model (CM)



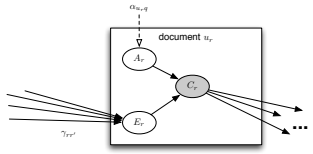
$$E_r = 1 \text{ and } A_r = 1 \Leftrightarrow C_r = 1 \tag{9}$$
$$P(A_r = 1) = \alpha_{u_r q} \tag{10}$$
$$P(E_1 = 1) = 1 \tag{11}$$
$$P(E_r = 1 \mid E_{r-1} = 0) = 0 \tag{12}$$
$$P(E_r = 1 \mid C_{r-1} = 1) = 0 \tag{13}$$
$$P(E_r = 1 \mid E_{r-1} = 1, C_{r-1} = 0) = 1. \tag{14}$$

Dynamic Bayesian Network Model (DBN)



$$E_r = 1 \text{ and } A_r = 1 \Leftrightarrow C_r = 1 \tag{15}$$
$$P(A_r = 1) = \alpha_{u_r q} \tag{16}$$
$$P(E_1 = 1) = 1 \tag{17}$$
$$P(E_r = 1 \mid E_{r-1} = 0) = 0 \tag{18}$$
$$P(S_r = 1 \mid C_r = 1) = \sigma_{u_r q} \tag{19}$$
$$P(E_r = 1 \mid S_{r-1} = 1) = 0 \tag{20}$$
$$P(E_r = 1 \mid E_{r-1} = 1, S_{r-1} = 0) = \gamma, \tag{21}$$

User Browsing Model (UBM)



$$P(E_r = 1 \mid C_1 = c_1, \ldots, C_{r-1} = c_{r-1}) = \gamma_{rr'}, \tag{22}$$
$$r' = \max \{k \in \{0, \ldots, r-1\} : c_k = 1\}, \tag{23}$$