Click models evaluations
(IR2 project)

Luka Stout
10616713

Finde Xumara
10690832

1 Introduction

Modeling user behavior on a search engine result page is important for understanding the users and supporting simulation experiments. As result pages become more complex, click models evolve as well in order to capture additional aspects of user behavior in response to new forms of result presentation. In recent years many models have been proposed that are aimed at predicting clicks of web search users.

In this report, we will implement and evaluate different click models using evaluation metrics. These click models include dependent click model (DCM) [3], dynamic bayesian network click model (DBN) [1], user browsing model (UBM) [2], and task-centric click model (TCM) [6]. The evaluation metrics include likelihood, perplexity, click through rate prediction, relevance prediction, computation time and ranking performance. By doing this experiment we can know the performance of each different click model and this information can be used as a benchmark of the next click model proposal.

This report is organized as follows. Section 2 describes the methods and algorithms used in our implementation. The experiments and data source information will be covered in Section 3, followed by the Section 4 for the performance analysis.

2 Methodology

We compare UBM, DCM, DBN and TCM on the Yandex dataset [4]. In this section, we briefly describe their main characteristics and differences. We have implemented TCM ourselves, the other algorithms were taken from PyClick [5].

2.1 DCM

The dependent click model was first proposed by Guo et al. in [3]. In the paper they propose a new click model which can handle multiple clicks per query by introducing a position dependent parameter $lambda_j$ to reflect the chance that the user would like to see more results after a click at position j. A graphical representation of the model is presented in Figure1a

2.2 DBN

The dynamic bayesian network is an extension to the traditional cascade model proposed by Chapelle and Zhang in [6]. For a given position j, in addition to observed variable C_j indicating whether there was a click or not at this position, the following latent variable are defined to model examination, perceived relevance and actual relevance, respectedly:

- E_j: did the user $examine$ the document?
- A_j: was the user $attracted$ by the document?
- S_j: was the user $satisfied$ by the clicked document?

They introduce a variable s_u for each document u which describes the relevance of the document for this query. When the user clicks on this document, there is a certain chance that the user will be satisfied. If the user is not satisfied, he continues to examine the next document with a probability γ and stops otherwise. The parameter γ is known as the 'perseverance'. A graphical representation of the model is presented in Figure 1b.

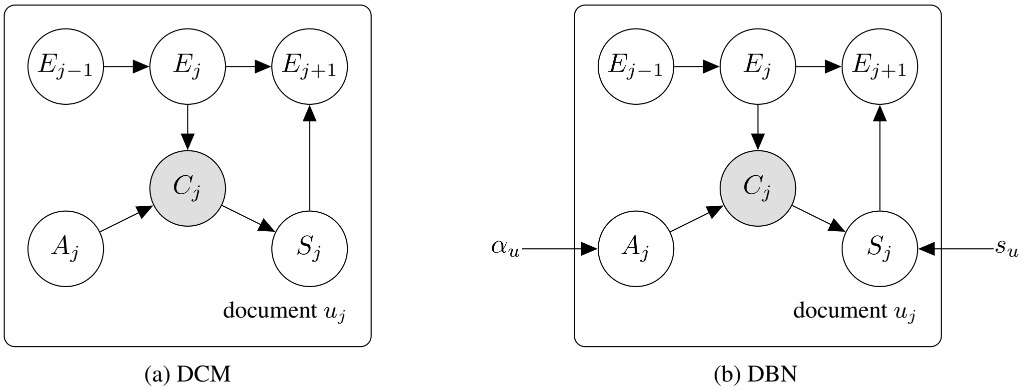


Figure 1: The graphical model of DCM and DBN.

2.3 UBM

In [2], Dupret and Piwowarski propose a new click model called the User Browsing Model (UBM). The main difference between UBM and other models is that UBM takes the distance into account from the current document u_j to the last clicked document $u_{j'}$ for determining the probability that the user continues browsing:

$$P(E_j = 1 \mid C_{j'} = 1, C_{j'+1} = 0, \ldots, C_{j-1} = 0) = \gamma_{jj'}$$

The probability that a document at rank j is examined E_j therefore depends on all possible paths the user could have taken to arrive at this document:

$$P(E_j = 1) = \sum_{j'=1}^{j-1} \gamma_{jj'}$$

A graphical representation of the model is presented in Figure 2.

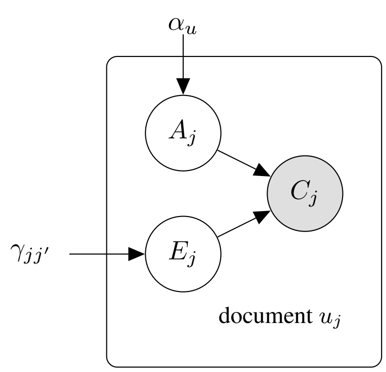


Figure 2: The graphical model of UBM.

2.4 TCM

The Task-centric Click Model (TCM) was first proposed by Zhang et al. in [6]. In the paper they propose a new click model which can handle multiple clicks of multiple queries in a task by introducing two new biases. The first bias indicates that users tend to express their information needs incrementally in a task, thus perform more clicks as their needs become clearer. The other bias indicates that users tend to click fresh documents that are not included in the results of previous queries. In their paper, they named the first assumption as `query bias`, and the second assumption as `duplicate bias`. A graphical representation of the state-of-the-art of the model is presented in Figure1a and the notations used in TCM are described in Table1.

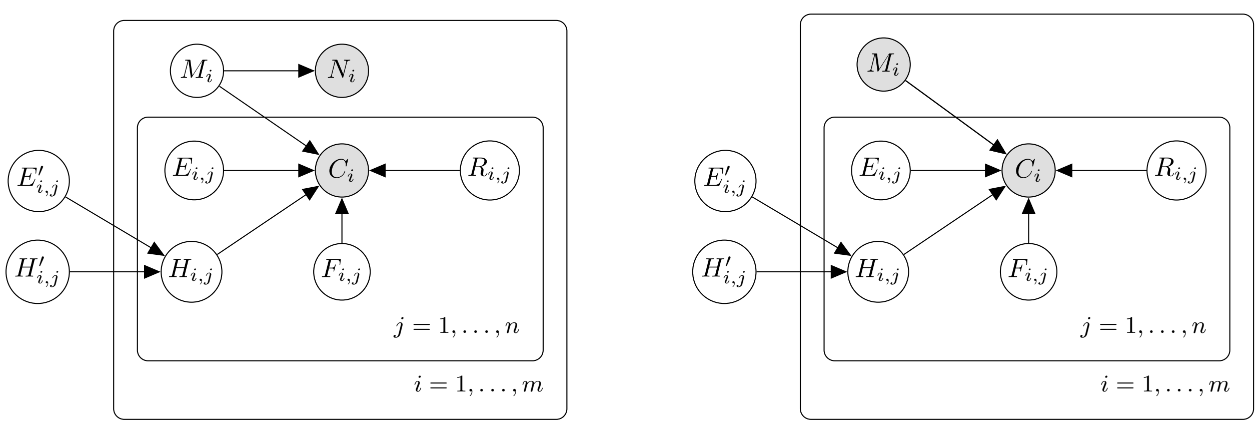
Symbol	Description
(i,j)	j-th ranking position in i-th query session.
M_i	Whether the i-th query matches the user's intent.
N_i	Whether the the user submits another query after i-th query session.
$E_{i,j}$	Examination of the document at (i,j).
$H_{i,j}$	Previous Examination of the document at (i,j).
$F_{i,j}$	Freshness of the document at (i,j).
$R_{i,j}$	Relevance of the document at (i,j).
$C_{i,j}$	Whether the the document at (i,j) is clicked.
(i',j')	Assume that d is the document at (i,j). i' is the latest query session where d has appeared in previous query sessions, and j' is the ranking position of this appearance.

Table 1: Notations used in TCM

This model can be formalized with the following conditional probabilities:

$$P(M_i = 1) = \alpha_1 \tag{1}$$
$$P(N_i | M_i = 1) = \alpha_2 \tag{2}$$
$$P(F_{i,j} = 1 | M_{i,j} = 1) = \alpha_3 \tag{3}$$
$$P(E_{i,j} = 1) = \beta_j \tag{4}$$
$$P(R_{i,j} = 1) = r_d \tag{5}$$
$$M_i = 0 \Rightarrow N_i = 1 \tag{6}$$
$$H_{i,j} = 0 \Rightarrow F_{i,j} = 1 \tag{7}$$
$$H_{i,j} = 0 \Leftrightarrow H_{i',j'} = 0, E_{i',j'} = 0 \tag{8}$$
$$C_{i,j} = 1 \Leftrightarrow M_i = 1, E_{i,j} = 1, R_{i,j} = 1, F_{i,j} = 1 \tag{9}$$

In our implementation, we simplified TCM model by assuming that M_i is observed from the click log data, thus eq.2 can be removed. Our second assumptions is that $M_i, E_{i,j}, R_{i,j}$ and $F_{i,j}$ are independent. The detail calculation for updating EM parameters of the simplified TCM can be found in the appendices section. The graphical model of our TCM implementation is presented in Fig 3.



(a) state-of-the-art TCM. (b) simplified TCM

Figure 3: The graphical model of state-of-the-art TCM and simplified TCM.

TODO:: add more click model

3 Evaluation

3.1 Evaluation Criteria

The equally evaluate each click models performance, we used evaluation metrics that was proposed along with the click model. The evaluation metrics used in this experiment is listed as follow:

3.1.1 Likelihood and Perplexity

Likelihood is already implied in each click model during E-step, where it should get better for every EM iteration. However, this evaluation can became interesting when we compare one click-model to the others. By comparing the likelihood evaluation per iteration, we can see the performance of each click-model based on its convergent speed.

Click perplexity is a widely used metric for evaluation click model accuracy. Perplexity can be seen as the log-likelihood powers which are computed independently at each position as also mentioned at [6] and [2]. For example, we assume that q_j^s is the probability of some click calculated from a click model, i.e. $P(C_j^s = 1)$ where C_j^s is a binary value indicating the click event at position j in query session s. Then the click perplexity at position j is computed as follows:

$$p_j = 2^{-\frac{1}{|S|}\sum_{s \in S}(C_j^s \log_2 q_j^s + (1-C_j^s)\log_2(1-q_j^s))}$$

The perplexity of a data set is defined as the average of perplexities in all positions. Thus, a smaller perplexity value indicates a better consistency between the click model and the actual click data.

3.1.2 Click-Trough-Rate prediction (CTR)

The purpose of click-through rates is to measure the ratio of clicks to impressions of an document. Generally the higher the CTR the higher change of that document being clicked. CTR can be formularized as following equation:

$$CTR = \frac{clicks}{impression} \times 100$$

3.1.3 Relevance prediction

Relevance prediction was used to evaluate performance of DBN model [1, p. 6]. The accuracy of CTR prediction may not directly translate to relevance, especially when we were to evaluate TCM. On TCM, the CTR of a particular document can be very low even if the document is relevant. This can happen because on TCM, we look over the whole task instead of single query, and the second time this document reappear, the user tend to ignore it by the freshness assumption. Therefore, relevance prediction can be used as additional inference of relevancy.

TODO:: add more info about general relevance algorithm

3.1.4 others

..

3.2 Evaluation setup

The experiment was run ...

3.3 Results

In Table 2 one can see the results of the experiments. It can be seen that ...

	Log-likelihood	Perplexity	Computation Time
Click Model A	0	0	
Click Model B	0	0	

Table 2: Results

4 Analysis

After running the experiments we were able to evaluate the different algorithms based on the ...

5 Conclusions

In this paper we showed that ...

In our implementation, we did not ...

References

[1] Olivier Chapelle and Ya Zhang. A Dynamic Bayesian Network Click Model for Web Search Ranking Categories and Subject Descriptors. *Www*, pages 1–10, 2009.

[2] Georges E Dupret and Benjamin Piwowarski. A user browsing model to predict search engine click data from past observations. *Sigir '08*, pages 331–338, 2008.

[3] Fan Guo, Chao Liu, and Yi Min Ym Wang. Efficient multiple-click models in web search. *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 124–131, 2009.

[4] Yandex LLC. Task and datasets, n.d.

[5] Ilya Markov. Click models for web and aggregated search, 2015.

[6] Yuchen Zhang, Weizhu Chen, Dong Wang, and Qiang Yang. User-click modeling for understanding and predicting search-behavior. *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '11*, page 1388, 2011.

APPENDICES

We give here some details about the inference in our TCM implementation outlined in section 2.4.

A Click probability

For $P(F_{i,j} = 1)$ we introduce a variable $f_{i,j}$, which will be derived later.
By assumption that $M_i, E_{i,j}, R_{i,j}$and$F_{i,j}$ are independent, click probability can be formularize as:

$$\begin{aligned} P(C_{i,j} = 1) &= P(M_i = 1) * P(E_{i,j} = 1) * P(R_{i,j} = 1) * P(F_{i,j} = 1) && (10) \\ &= \alpha_1 * \beta_j * r_{i,j} * f_{i,j} && (11) \end{aligned}$$

B Probability of the query match user intention

Because we remove equation that depends on α_2, we can now set α_1 as MLE.

$$P(M_i = 1) = \alpha_1$$

C Probability of user submit next query

User submit next query if the query does not match user intention (α_1) or user want to search more.

$$\begin{aligned} P(N_i = 1) &= \frac{1}{|S|} \sum_{i \in S} \mathcal{I}(N_i = 1) \\ &= \frac{q_i}{|S|} \\ &= n_i \end{aligned}$$

q_i is the number of submitted-queries where user submit another query after i-th query session.

$$\begin{aligned} P(N_i = 1 | M_i = 1) = \alpha_2 & \\ &= \frac{P(N_i = 1) - P(N_i = 1 | M_i = 0) P(M_i = 0)}{P(M_i = 1)} \\ &= \frac{n_i + \alpha_1 - 1}{\alpha_1} \end{aligned}$$

D Relevance probability

$$\begin{aligned} P(R_{i,j} = 1) = r_{i,j} & \\ &= \frac{\sum_{q_{i,j} \in S_{i,j}} P(R_{i,j} = 1 | C)}{|S_{i,j}|} \end{aligned}$$

Where $S_{i,j}$ are all sessions (queries) containing the document corresponding with the query i at rank j - document $P(R_{i,j} = 1 | C)$ will be derive on eq.14

$$\begin{aligned} P(R_{i,j} = 1 | C) &= \mathcal{I}(C_{i,j} = 1) P(R_{i,j} | C_{i,j} = 1) + \mathcal{I}(C_{i,j} = 0) P(R_{i,j} | C_{i,j} = 0) && (12) \\ &= c_{i,j} + (1 - c_{i,j}) \frac{P(C_{i,j} = 0 | R_{i,j} = 1) P(R_{i,j} = 1)}{P(C_{i,j} = 0)} && (13) \\ &= c_{i,j} + (1 - c_{i,j}) \frac{P(C_{i,j} = 0 | R_{i,j} = 1) r_{i,j}}{1 - P(C_{i,j} = 1)} && (14) \end{aligned}$$

Where $c_{i,j} = 1$ if (i,j) was clicked in the current session. $P(C_{i,j} = 0|R_{i,j} = 1)$ is the chance of no click given that it is relevant.

$$
\begin{aligned}
P(C_{i,j} = 0|R_{i,j} = 1) &= P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1)P(M_i = 1) + P(C_{i,j} = 0|R_{i,j} = 1, M_i = 0)P(M_i = 0) \\
&= \alpha_1 P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1, E_{i,j} = 1)P(E_{i,j} = 1) \\
&+ \alpha_1 P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1, E_{i,j} = 0)P(E_{i,j} = 0) \\
&+ (1 - \alpha_1)P(C_{i,j} = 0|R_{i,j} = 1, M_i = 0, E_{i,j} = 1)P(E_{i,j} = 1) \\
&+ (1 - \alpha_1)P(C_{i,j} = 0|R_{i,j} = 1, M_i = 0, E_{i,j} = 0)P(E_{i,j} = 0) \\
\\
&= \alpha_1\beta_j P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1, E_{i,j} = 1, F_{i,j} = 1)P(F_{i,j} = 1) \\
&+ \alpha_1\beta_j P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1, E_{i,j} = 1, F_{i,j} = 0)P(F_{i,j} = 0) \\
&+ \alpha_1(1 - \beta_j)P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1, E_{i,j} = 0, F_{i,j} = 1)P(F_{i,j} = 1) \\
&+ \alpha_1(1 - \beta_j)P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1, E_{i,j} = 0, F_{i,j} = 0)P(F_{i,j} = 0) \\
&+ (1 - \alpha_1)\beta_j P(C_{i,j} = 0|R_{i,j} = 1, M_i = 0, E_{i,j} = 1, F_{i,j} = 1)P(F_{i,j} = 1) \\
&+ (1 - \alpha_1)\beta_j P(C_{i,j} = 0|R_{i,j} = 1, M_i = 0, E_{i,j} = 1, F_{i,j} = 0)P(F_{i,j} = 0) \\
&+ (1 - \alpha_1)(1 - \beta_j)P(C_{i,j} = 0|R_{i,j} = 1, M_i = 0, E_{i,j} = 0, F_{i,j} = 1)P(F_{i,j} = 1) \\
&+ (1 - \alpha_1)(1 - \beta_j)P(C_{i,j} = 0|R_{i,j} = 1, M_i = 0, E_{i,j} = 0, F_{i,j} = 0)P(F_{i,j} = 0)
\end{aligned}
$$

We note that $P(C_{i,j} = 0|R_{i,j} = 1, M_i = 1, E_{i,j} = 1, F_{i,j} = 1) = 0$. Otherwise it is 1. From eq. 24 from TCM paper. Together with inserting our parameters this gives us the following:

$$
\begin{aligned}
P(C_{i,j} = 0|R_{i,j} = 1) &= \alpha_1\beta_j f_{i,j} + \alpha_1\beta_j(1 - f_{i,j}) + \alpha_1(1 - \beta_j)f_{i,j} + \alpha_1(1 - \beta_j)(1 - f_{i,j}) && (15) \\
&+ (1 - \alpha_1)\beta_j f_{i,j} + (1 - \alpha_1)\beta_j(1 - f_{i,j}) + (1 - \alpha_1)(1 - \beta_j)(f_{i,j} && (16) \\
&+ (1 - \alpha_1)(1 - \beta_j)(1 - f_{i,j}) && (17)
\end{aligned}
$$

expanding this we are only left with

$$
P(C_{i,j} = 0|R_{i,j} = 1) = 1 - (\alpha_1\beta_j f_{i,j}) \tag{18}
$$

Which seems intuitive as we assumed that all $M_i, R_{i,j}, E_{i,j}$ and $F_{i,j}$ are independent to get $P(C_{i,j} = 1)$. With this information we can calculate

$$
\begin{aligned}
P(R_{i,j} = 1|C) &= c_{i,j} + (1 - c_{i,j})\frac{(1 - (\alpha_1\beta_j f_{i,j}))r_{i,j}}{1 - \alpha_1\beta_j f_{i,j}r_{i,j}} && (19) \\
&= c_{i,j} + (1 - c_{i,j})\frac{r_{i,j} - \alpha_1\beta_j f_{i,j}r_{i,j}}{1 - \alpha_1\beta_j f_{i,j}r_{i,j}} && (20)
\end{aligned}
$$

E Examination probability

$$
\begin{aligned}
P(E_{i,j} = 1) &= \beta_j && (21) \\
&= \frac{1}{|S|}\sum_{i \in S} P(E_{i,j} = 1|C) && (22)
\end{aligned}
$$

Where S is all sessions and i is a query within that session. $P(E_{i,j} = 1|C)$ will be derive on eq.23

$$
\begin{aligned}
P(E_{i,j} = 1|C) &= \mathcal{I}(C_{i,j} = 1)P(E_{i,j}|C_{i,j} = 1) + \mathcal{I}(C_{i,j} = 0)P(E_{i,j}|C_{i,j} = 0) && (23) \\
&= c_{i,j} + (1 - c_{i,j})\frac{P(C_{i,j} = 0|E_{i,j} = 1)P(E_{i,j} = 1)}{P(C_{i,j} = 0)} && (24) \\
&= c_{i,j} + (1 - c_{i,j})\frac{P(C_{i,j} = 0|E_{i,j} = 1)\beta_j}{1 - P(C_{i,j} = 1)} && (25)
\end{aligned}
$$

Where $c_{i,j}$ indicates whether document i, j was clicked. Analog to eq 18 we can show that

$$P(C_{i,j} = 0|E_{i,j} = 1) = 1 - (\alpha_1 f_{i,j} r_{i,j}) \tag{26}$$

This gives us

$$P(E_{i,j} = 1|C) = c_{i,j} + (1 - c_{i,j})\frac{(1 - (\alpha_1 f_{i,j} r_{i,j}))\beta_j}{1 - \alpha_1 \beta_j f_{i,j} r_{i,j}} \tag{27}$$

$$= c_{i,j} + (1 - c_{i,j})\frac{\beta_j - \alpha_1 \beta_j f_{i,j} r_{i,j}}{1 - \alpha_1 \beta_j f_{i,j} r_{i,j}} \tag{28}$$

F Freshness probability

$$P(F_{i,j} = 1|H_{i,j} = 1) = \alpha_3 \tag{29}$$

$$\alpha_3 = \frac{1}{|S_{i,j}|}\sum_{q\in S}\sum_{(i,j)\in q} P(Fi, j = 1|H_{i,j} = 1, C) \tag{30}$$

Where (i,j) is a query, rank pair identifying a certain document. $P(F_{i,j} = 1|C)$ will be derive on eq.31
$P(F_{i,j} = 1)$ will be derive on eq.37

$$P(F_{i,j} = 1|H_{i,j} = 1, C) = \mathcal{I}(C_{i,j} = 1)P(F_{i,j} = 1|H_{i,j} = 1, C_{i,j} = 1) \tag{31}$$

$$+ \mathcal{I}(C_{i,j} = 0)P(F_{i,j} = 1|H_{i,j} = 1, C_{i,j} = 0) \tag{32}$$

$$= c_{i,j} + (1 - c_{i,j})\frac{P(C_{i,j} = 0|F_{i,j} = 1, H_{i,j} = 1)P(F_{i,j} = 1|H_{i,j} = 1)}{P(C_{i,j} = 0|H_{i,j} = 1)} \tag{33}$$

Analog to eq 18 we can show that

$$P(C_{i,j} = 0|F_{i,j} = 1, H_{i,j} = 1) = 1 - (\alpha_1 \beta_j r_{i,j}) \tag{34}$$

We can also show

$$P(C_{i,j} = 0|H_{i,j} = 1) = 1 - P(C_{i,j} = 1|H_{i,j} = 1) \tag{35}$$

$$= 1 - (\alpha_1 \alpha_3 \beta_j r_{i,j}) \tag{36}$$

The only difference between this and eq. 11 is that it is given that $H_{i,j} = 1$ and because $H_{i,j} = 1$ only has an influence on $P(F_{i,j} = 1)$, namely that $P(F_{i,j} = 1|H_{i,j} = 1) = 1$, we can substitute $f_{i,j}$ with $\alpha 3$ in eq. 11

Now we only need to calculate $f_{i,j} = P(F_{i,j}) = 1$

$$P(F_{i,j} = 1) = \mathcal{I}(H_{i,j} = 1)P(F_{i,j} = 1|H_{i,j} = 1) + \mathcal{I}(H_{i,j} = 0)P(F_{i,j} = 1|H_{i,j} = 0) \tag{37}$$

$$= \mathcal{I}(H_{i,j} = 1)\alpha_3 + \mathcal{I}(H_{i,j} = 0) \tag{38}$$

Where $\mathcal{I}(H_{i,j} = 1)$ is a binary indicator function from the data specifying whether document (i, j) was shown before in the current (q from eq. 30) session.

We could replace this indicator function with the probability that the document was examined the last time it was shown. This probability, called $H_{i,j}$ would depend on the probability that it was examined and $H_{i',j'}$ where i', j' is the last time this document was shown in the current session. It would look like this

$$P(H_{i,j} = 1) = P(E_{i',j'} = 1)P(H_{i',j'} = 1) \tag{39}$$

then eq. 37 becomes:

$$P(F_{i,j} = 1) = P(H_{i,j} = 1)\alpha_3 + P(H_{i,j} = 0) \tag{40}$$

$$= P(H_{i,j} = 1)\alpha_3 + (1 - P(H_{i,j} = 1)) \tag{41}$$

$$= \alpha_3 P(E_{i',j'} = 1)P(H_{i',j'} = 1) + (1 - P(E_{i',j'} = 1)P(H_{i',j'} = 1)) \tag{42}$$

Note that this discards the information that if (i', j') was clicked it surely was examined.

With eq 34 we can calculate $P(F_{i,j} = 1|C)$

$$P(F_{i,j} = 1|H = 1, C) = c_{i,j} + (1 - c_{i,j})\frac{(1 - (\alpha_1 \beta_j r_{i,j}))\alpha_3}{1 - \alpha_1 \alpha_3 \beta_j r_{i,j}} \quad (43)$$

$$= c_{i,j} + (1 - c_{i,j})\frac{\alpha_3 - \alpha_1 \alpha_3 \beta_j r_{i,j}}{1 - \alpha_1 \alpha_3 \beta_j r_{i,j}} \quad (44)$$