Estimating Credibility of User Clicks with Mouse Movement and Eye-tracking Information

Jiaxin Mao, Yiqun Liu, Min Zhang, Shaoping Ma

Department of Computer Science and Technology,

Tsinghua University

Background

- Search engine
  - One of the most popular Web applications
  - Search for relevant information on the Web
- Key issue:
  - Determine whether a document (webpage) is relevant to user's query or not
  - Rank the documents according to their relevance

Click Through Data

- An important source of implicit relevance feedback
  - Easy to collect at a large scale
  - Valuable for improving the Web search engine
    - (Joachims, T. 2002 and 2005)
  - Click models
    - Cascade model (Craswell, N. 2008), DBN model (Chapelle, O. 2009) etc.
- However, not every click is equally credible
  - Click spam
  - Large variation exists in users' personal characteristics and behavioral patterns
    - (Xing, Q. 2013), (White, R.W. 2007 and 2009)
  - Query difficulties and result qualities also affect click behaviors

Our Work

- Build an experimental search engine to collect a detailed user behavior log
- Characterize click credibility
- Estimate or predict click credibility
- Use the estimates of click credibility to improve the relevance feedback for search engine

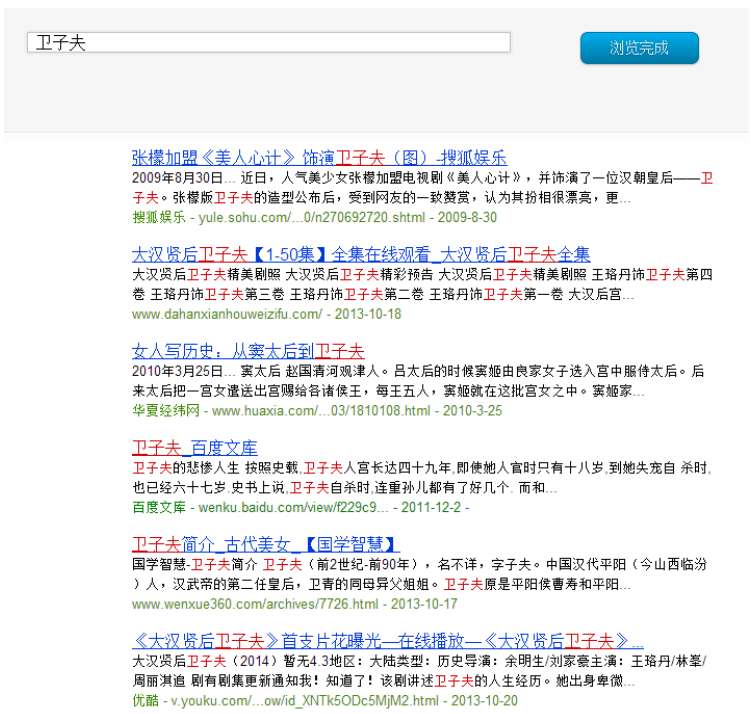
Collect User Behavior Log

- Build a experimental search engine
- Hire 31 subjects
  - 15 males and 16 females
  - first-year university students from two different majors
- Sample 25 queries
  - 10 informational, 10 transactional and 5 navigational
- Collect 774 valid sessions
  - <user, query> pairs
- Comprehensive interaction data including:
  - Mouse movement, click-through, eye-tracking data
  - relevance annotation

Characterize Click Credibility

- Treat a search engine user as a relevant document classifier
- Metrics:
  - Accuracy: $a_s = \frac{\#\{TP\}+\#\{TN\}}{\#\{TP\}+\#\{TN\}+\#\{FP\}+\#\{FN\}}$
  - True Positive Rate: $TP_S = \frac{\#\{TP\}}{\#\{TP\}+\#\{FN\}}$
  - True Negative Rate: $TN_S = \frac{\#\{TN\}}{\#\{TN\}+\#\{FP\}}$
- Only consider examined results
  - Use eye-tracking data as ground truth of examination

An Example



$$a_s = \frac{3}{5} = 0.6, TP_S = \frac{2}{3} = 0.67, TN_S = \frac{1}{2} = 0.5$$

Characterize Click Credibility

- Statistics of the Metrics
  - Users' click credibility is not so high
  - Observe a obvious variation
  - Query is a stronger influence and user

Table 2. Statistics for metrics of click credibility.

Metrics		Accuracy	True Positive Rate	True Negative Rate
Grouped by user	M	0.613	0.557	0.813
	SD	0.236	0.274	0.239
Grouped by query	M	0.613	0.557	0.811
	SD	0.200	0.257	0.247
Single session	M	0.613	0.557	0.812
	SD	0.249	0.306	0.269

Predict Click Credibility

- Regression
- 26 features:
  - 3 Click features
  - 4 session features
  - 13 mouse movement features
  - 2 query features
  - 4 user features
- Mouse movement data
  - An abundant behavioral data source
  - Can be collected in a large scale

Predict Click Credibility

- Regression model:
  - Support Vector Regression
  - Logit transformation: $\alpha = \text{logit}(p) = \log(\frac{p}{1-p})$
- Evaluation metric:
  - Mean Squared Error(MSE), Mean Absolute Error (MAE)
- Evaluation method:
  - 2-way leave-one-out cross validation
  - With disjoint user and query
- Baseline:
  - arithmetic mean of the metrics of the training set sessions

Predict Click Credibility

- Results

Table 4. Results of predicting click credibility, * indicates the improvement over baseline is significant with $p < 10^{-3}$.

	Baseline		SVR	
	MSE	MAE	MSE	MAE
Accuracy	0.071075	0.224086	0.061015(-14.1%*)	0.206813(-7.7%*)
True Positive Rate	0.109941	0.290940	0.082651(-24.8%*)	0.219068(-24.7%*)
True Negative Rate	0.086662	0.192934	0.069311(-20.0%*)	0.165896(-16.6%*)

Estimate Relevance

- Baseline: examination hypothesis:

$$r_i = P(C_i = 1 | E_i = 1)$$

- Accuracy model:

$$P(C_i = 1 | E_i = 1) = r_i \times a_s + (1 - r_i) \times (1 - a_s)$$
$$P(C_i = 0 | E_i = 1) = (1 - r_i) \times a_s + r_i \times (1 - a_s)$$

- Confusion matrix model:

$$P(C_i = 1 | E_i = 1) = r_i \times TP_s + (1 - r_i) \times (1 - TN_s)$$
$$P(C_i = 0 | E_i = 1) = (1 - r_i) \times TN_s + r_i \times (1 - TP_s)$$

Estimate Relevance

- Use predicted metrics (computed by LOO-CV) and maximum likelihood estimation to estimate r_i
- Evaluation method:
  - Rank the results according to r_i
  - Use MAP(Mean Average Precision) to evaluation the ranking

Estimate Relevance

- Results:
  - Click through data as feedback can improve ranking
  - Taking click credibility into consideration is useful

Table 5. Results of relevance estimation, * indicates the improvement is significant with $p < 0.05$ and ** indicates $p < 0.01$.

	Original Ranking	Baseline (Examination hypothesis without credibility estimation)	Accuracy Model	Confusion Matrix Model
MAP	0.805124	0.843075	0.884604	0.877654
Improvement over original ranking	-	+4.7%	+9.9%**	+9.0%**
Improvement over baseline	-	-	+4.9%**	+4.1%*

Conclusions

- Not every user click is credible and the click credibility varies across sessions
- User behavior data can be used to estimate the credibility of click.
- The predicted click crediblity can improve relevance estimation

Thank you!

- Q&A

References

- White, R.W., Dumais, S.T., Teevan, J.: Characterizing the influence of domain expertise on web search behavior. In: Proceedings of the Second ACM International Conference on Web Search and Data Mining. pp. 132–141. ACM (2009)
- White, R.W., Morris, D.: Investigating the querying and browsing behavior of advanced search engine users. In: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 255–262. ACM (2007)
- Xing, Q., Liu, Y., Zhang, M., Ma, S., Zhang, K.: Characterizing expertise of search engine users. In: Information Retrieval Technology, pp. 380–391. Springer (2013)
- Chapelle, O., Zhang, Y.: A dynamic bayesian network click model for web search ranking. In: Proceedings of the 18th international conference on World Wide Web. pp. 1–10. ACM (2009)
- Craswell, N., Zoeter, O., Taylor, M., Ramsey, B.: An experimental comparison of click position-bias models. In: Proceedings of the 2008 International Conference on Web Search and Data Mining. pp. 87–94. ACM (2008)
- Joachims, T.: Optimizing search engines using clickthrough data. In: Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 133–142. ACM (2002)
- Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately interpreting clickthrough data as implicit feedback. In: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval. pp. 154–161. ACM (2005)

Features

No.	Description	Accuracy	TP	TN
Click Features				
1	number of clicked results	-0.02	0.24*	-0.55*
2	lowest rank of clicks	-0.11	0.05	-0.36*
3	average difference in ranks between two clicks	-0.06	0.03	-0.16*
Session Features				
4	average time spent on the SERP for each click	-0.04	-0.05	0.09
5	total time spent on the search task	-0.07	0.03	-0.19*
6	total time spent on the SERP	-0.12*	-0.01	-0.22*
7	maximal continuous time spent on the SERP	-0.14*	-0.16*	0.06
Mouse Movement Features				
8	number of hovered results	-0.15*	-0.06	-0.26*
9	lowest rank of hovers	-0.18*	-0.10	-0.24*
10	number of results that are hovered over but not clicked	-0.18*	-0.23*	0.07
11	moving time of the mouse	-0.12	-0.02	-0.17*
12	idle time of the mouse	-0.06	0.10	-0.34*
13	dwell time of the mouse in the result region	-0.11	-0.03	-0.16*
14	length of the mouse trails	-0.08	0.08	-0.34*
15	velocity of mouse movement	0.12*	0.17*	-0.16*
16	horizontal moving distance of the mouse	-0.05	0.07	-0.19*
17	vertical moving distance of the mouse	-0.08	0.08	-0.37*
18	maximal y coordinate that the mouse has reached	-0.18*	-0.11	-0.22*
19	total distance of scrolling	-0.11	-0.09	-0.13
20	maximal displacement in y axis of scrolling	-0.15*	-0.13*	-0.14*
Query Features				
21	click entropy of the query	-0.17*	-0.13*	-0.13
22	N Results Satisfied Rate of the query	-0.08	-0.03	0.01
User Features				
23	click entropy of the user	-0.16*	-0.06	-0.21*
24	user's total number of clicks	-0.02	0.14*	-0.32*
25	average time that the user spends on each search task	-0.08	0.02	-0.20*
26	average time that the user spends on the SERP	-0.10	-0.05	-0.12