Classifying User Search Intents for Query Auto-Completion

Jyun-Yu Jiang[1,2] and Pu-Jen Cheng[2]

[1]University of California, Los Angeles
[2]National Taiwan University

Sep. 14, 2016 (ICTIR)

Query Auto-Completion (QAC)

- A common feature in modern search engines
  - Help users formulate queries while typing in the search boxes
- Given a user-typed prefix, N ranked completions are shown

Why Query Auto Completion?

- Typing queries costs too much
  - Users can save their keystrokes
- Further benefits
  - Spelling errors, query expansion, ...

The goal of QAC

Rank the user's intended query in a high position
with as few keystrokes as possible

Query Auto-Completion (QAC)

- A common feature in modern search engines
  - Help users formulate queries while typing in the search boxes
- Given a user-typed prefix, N ranked completions are shown

Why Query Auto Completion?

- Typing queries costs too much
  - Users can save their keystrokes
- Further benefits
  - Spelling errors, query expansion, ...

The goal of QAC

Rank the user's intended query in a high position
with as few keystrokes as possible

Context-Aware Approach

- Context captures user's search intents.
  - submitted queries
  - click-through information

$$\underbrace{q_1 \rightarrow \cdots \rightarrow \cdot q_{T-1}}_{context} \rightarrow q_T$$

Query Session

- query dependencies [He2009]
- query similarity [Bar-Yossef2011]
- reformulation behavior [Jiang2014]

Click-through Data

- relevant queries [Mei2009]
- query clusters [Liao2011]
- click behavior [Ozertem2012]

More context may lead to more information.

Context-Aware Approach

- Context captures user's search intents.
  - submitted queries
  - click-through information

$$\underbrace{q_1 \to \cdots \to \cdot q_{T-1}}_{context} \to q_T$$

Query Session

- query dependencies [He2009]
- query similarity [Bar-Yossef2011]
- reformulation behavior [Jiang2014]

Click-through Data

- relevant queries [Mei2009]
- query clusters [Liao2011]
- click behavior [Ozertem2012]

More context may lead to more information.

Context-Aware Approach

- Context captures user's search intents.
  - submitted queries
  - click-through information

$$\underbrace{q_1 \rightarrow \cdots \rightarrow \cdot q_{T-1}}_{context} \rightarrow q_T$$

Query Session

- query dependencies [He2009]
- query similarity [Bar-Yossef2011]
- reformulation behavior [Jiang2014]

Click-through Data

- relevant queries [Mei2009]
- query clusters [Liao2011]
- click behavior [Ozertem2012]

More context may lead to more information.

Context-Aware Approach

- Context captures user's search intents.
  - submitted queries
  - click-through information

$$\underbrace{q_1 \to \cdots \to \cdot q_{T-1}}_{context} \to q_T$$

Query Session

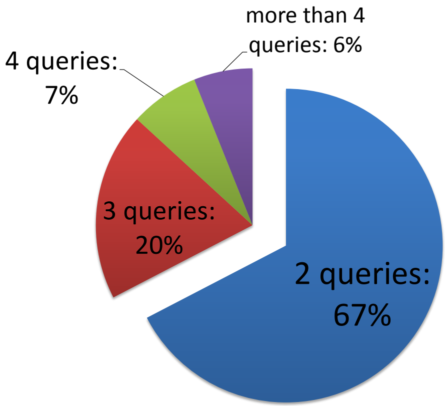
- query dependencies [He2009]
- query similarity [Bar-Yossef2011]
- reformulation behavior [Jiang2014]

Click-through Data

- relevant queries [Mei2009]
- query clusters [Liao2011]
- click behavior [Ozertem2012]

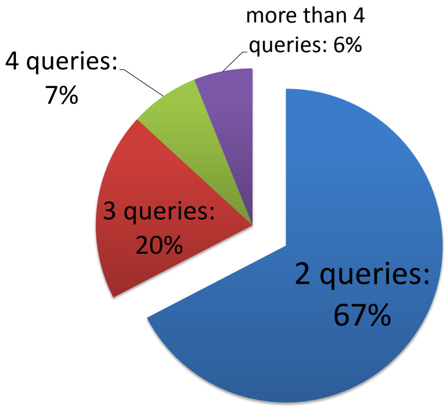
More context may lead to more information.

However, most of sessions are short and sparse!



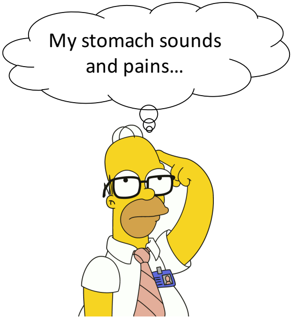
How to deal with the sparseness problem?

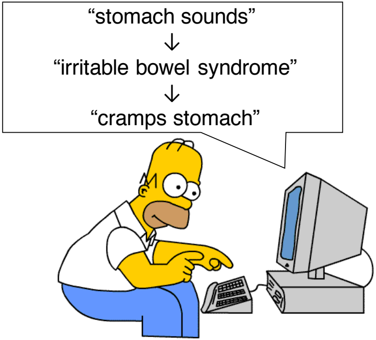
However, most of sessions are short and sparse!



How to deal with the sparseness problem?

Motivation: How are the queries decided?

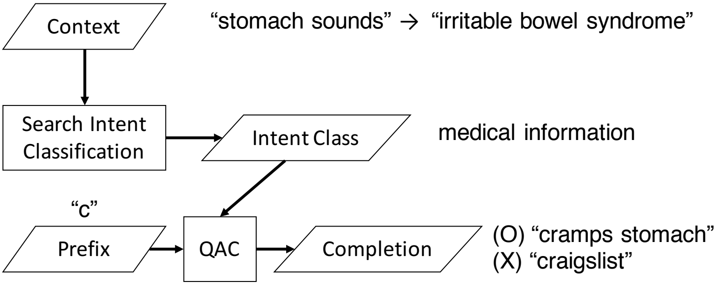




Context can be sparse, but search intents may be not!

Search Intent Classification and Query Auto-Completion

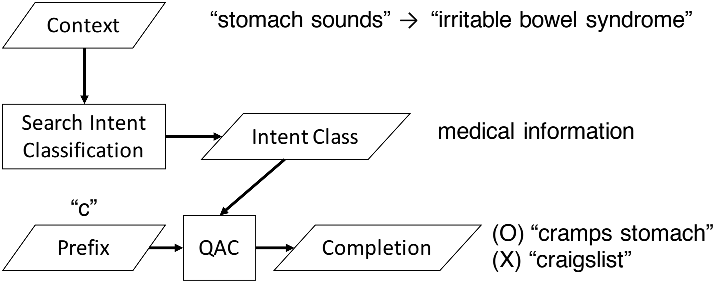
Search intents may not be predicted, but can be classified.



Existing classification structures can be helpful to enhance QAC

Search Intent Classification and Query Auto-Completion

Search intents may not be predicted, but can be classified.



Existing classification structures can be helpful to enhance QAC

Search Intent Classification for QAC

Problem Definition

- A session is a sequence of queries $\langle q_1, q_2, \cdots, q_T \rangle$
  - Each query q_i is issued in time t_i, and has clicked URLs $\boldsymbol{u_i}$.
  - Treat $\langle q_1, q_2, \cdots, q_{T-1} \rangle$ as the context and q_T as the intended query.
- Given the context, the prefix and a candidate set $Q_T = \{q'_j\}$
- The goal is to rank queries in Q_T and let q_T in a high position.

Our Approach

- Estimate the class distributions of the context and candidate queries
- Propose several features with three views of the context
- A supervised framework with *LambdaMART* learning-to-rank model.

Query and Session Classification

Estimate class distribution for the session and candidate queries

Distribution v.s. Single Class

- Smoothing techniques
- User intents are complicated
- More general representation

Classification Space

- Open directory project (ODP)
- Utilize 16 top-level categories
- Covered $53^{+}\%$ of clicks

More convenient to discover relations in the same classification space

Query and Session Classification

Estimate class distribution for the session and candidate queries

Distribution v.s. Single Class

- Smoothing techniques
- User intents are complicated
- More general representation

Classification Space

- Open directory project (ODP)
- Utilize 16 top-level categories
- Covered 53^{+}% of clicks

More convenient to discover relations in the same classification space

Query and Session Classification

Estimate class distribution for the session and candidate queries

Distribution v.s. Single Class

- Smoothing techniques
- User intents are complicated
- More general representation

Classification Space

- Open directory project (ODP)
- Utilize 16 top-level categories
- Covered $53^{+}\%$ of clicks

More convenient to discover relations in the same classification space

Query and Session Classification

Estimate class distribution for the session and candidate queries

Distribution v.s. Single Class

- Smoothing techniques
- User intents are complicated
- More general representation

Classification Space

- Open directory project (ODP)
- Utilize 16 top-level categories
- Covered $53^{+}\%$ of clicks

More convenient to discover relations in the same classification space

Query-class Distribution $P(c \mid q)$

Two Assumptions

- Query-class distribution is an aggregation over all relevant URLs.
- The distribution is only dependent to relevant URLs.

$$
\begin{aligned}
P(c \mid q) &= \sum_u P(c \mid u, q) \cdot P(u \mid q) && \text{(marginalization)} \\
&= \sum_u P(c \mid u) \cdot P(u \mid q) && \text{(by assumption),}
\end{aligned}
$$

We can compute $P(u \mid q)$ and $P(c \mid u)$ separately!

Query-class Distribution $P(c \mid q)$

Two Assumptions

- Query-class distribution is an aggregation over all relevant URLs.
- The distribution is only dependent to relevant URLs.

$$P(c \mid q) = \sum_u P(c \mid u, q) \cdot P(u \mid q) \qquad \text{(marginalization)}$$

$$= \sum_u P(c \mid u) \cdot P(u \mid q) \qquad \text{(by assumption)},$$

We can compute $P(u \mid q)$ and $P(c \mid u)$ separately!

URL-class Distribution $P(c \mid u)$

Smoothing with URLs in ODP data (i.e., “gold-standard” classification)

Assumption

URLs u with the same host $h(u)$ may have similar distributions.

$$P(c \mid u) = \frac{Occurs\left(h\left(u\right), c\right) + m \cdot P\left(c\right)}{m + \sum_{c_i} Occurs\left(h\left(u\right), c_i\right)}$$

Prior Distribution $P(c)$

Normalizing the number of websites in ODP for each category

Query-URL Relevance $P(u \mid q)$

Smoothing with clicked times in search logs

Assumption Again!

URLs u with the same host $h(u)$ may have similar distributions.

$$P(u \mid q) = \frac{C(h(u), q) + m \cdot P(h(u))}{m + \sum_{h(u)} C(h(u), q)}$$

Prior Distribution $P(h(u))$

Normalizing the number of times corresponding URLs are clicked in the log

Session-class Distribution $P(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle)$

Three views of the context

- All Preceding Queries (all)
  - Consider information of the whole search session

$$P_{all}\left(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle\right) = \frac{1}{\sum w_i} \sum w_i P\left(c \mid q_i\right).$$

  - w_i is a linear-decayed weight.
- Last Query (Last)
  - Too former queries may be noisy.
  - Only consider the last query as the context

$$P_{last}\left(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle\right) = P\left(c \mid q_{T-1}\right).$$

Session-class Distribution $P(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle)$

Three views of the context

- All Preceding Queries (all)
  - Consider information of the whole search session

$$P_{all}\left(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle\right) = \frac{1}{\sum w_i} \sum w_i P\left(c \mid q_i\right).$$

  - w_i is a linear-decayed weight.

- Last Query (Last)
  - Too former queries may be noisy.
  - Only consider the last query as the context

$$P_{last}\left(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle\right) = P\left(c \mid q_{T-1}\right).$$

Session-class Distribution $P(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle)$

Three views of the context

- All Preceding Queries (all)
  - Consider information of the whole search session

$$P_{all}\left(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle\right) = \frac{1}{\sum w_i} \sum w_i P\left(c \mid q_i\right).$$

  - w_i is a linear-decayed weight.
- Last Query (Last)
  - Too former queries may be noisy.
  - Only consider the last query as the context

$$P_{last}\left(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle\right) = P\left(c \mid q_{T-1}\right).$$

Session-class Distribution (Cont'd)

- Local-clicked URLs (Local)
  - Re-compute URL relevance with local click-through data

$$P_{local}\left(u \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle\right) = \frac{C_{local}\left(h\left(u\right)\right) + m \cdot P\left(h\left(u\right)\right)}{m + \sum_{h(u)} C_{local}\left(h\left(u\right)\right)}$$

  - Aggregate distributions of URLs with new relevance

$$P_{local}(c \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle) = \sum_{u_i \in \boldsymbol{u}} P(c \mid u_i) P_{local}(u_i \mid \langle q_1, q_2, \cdots, q_{T-1} \rangle)$$

Distribution-based Features

Find relations between the context and candidate queries by distributions

Feature	Query	Session	# in Model
Query Class Entropy (QCE)	√		1
Session Class Entropy (SCE)		√	3
Class Match (CM)	√	√	3
ArgMaxOdds (AMO)	√	√	3
MaxOdds (MO)	√	√	3
KL Divergence (KL)	√	√	3
Cross Entropy (CE)	√	√	3
Distribution Similarity (DS)	√	√	3

Apply *LambdaMART* to rank candidate queries

Experimental Settings

- 3-month AOL search engine log from 1 March, 2006 to 31 May, 2006

Data Pre-processing

- 30-minute threshold as the session boundary
- Firth 2-month data for training, the remaining for testing
- Drop queries appear less than 10 times
- Predict every query in sessions except the first one without context
- Test with different prefix length $\#p$

Experimental Settings (2/2)

Testing Datasets

- Divide testing cases into four datasets with different lengths of context
  - Overall (all tasks)
  - Short Context (1 query)
  - Medium Context (2 to 3 queries)
  - Long Context (4 or more queries)
- Evaluate performance on tasks with different context lengths

Evaluation Metric

- Mean Reciprocal Rank (MRR)

- Fine-tune our *LambdaMART* ranking model with parameters of 1,000 decision tress across all experiments.

Six Competitive Baselines

- Most Popular Completion (MPC)
  - Maximum Likelihood Estimation (MLE) approach
- Hybrid Completion (Hyb.C) [Bar-Yossef et al., 2011]
  - Consider both context information and the popularity
- Personalized Completion (Per.C) [Shokouhi, 2013]
  - Considers users personal information (only submitted history in AOL)
- Query-based VMM (QVMM) [He et al., 2009]
  - Context-aware query suggestion method
- Concept-based VMM (CACB) [Liao et al., 2011]
  - Concept-based context-aware query suggestion method
- Reformulation-based Completion (RC) [Jiang et al., 2014]
  - Model users' reformulation behavior for QAC

Six Competitive Baselines

- Most Popular Completion (MPC)
  - Maximum Likelihood Estimation (MLE) approach
- Hybrid Completion (Hyb.C) [Bar-Yossef et al., 2011]
  - Consider both context information and the popularity
- Personalized Completion (Per.C) [Shokouhi, 2013]
  - Considers users personal information (only submitted history in AOL)
- Query-based VMM (QVMM) [He et al., 2009]
  - Context-aware query suggestion method
- Concept-based VMM (CACB) [Liao et al., 2011]
  - Concept-based context-aware query suggestion method
- Reformulation-based Completion (RC) [Jiang et al., 2014]
  - Model users' reformulation behavior for QAC

Six Competitive Baselines

- Most Popular Completion (MPC)
  - Maximum Likelihood Estimation (MLE) approach
- Hybrid Completion (Hyb.C) [Bar-Yossef et al., 2011]
  - Consider both context information and the popularity
- Personalized Completion (Per.C) [Shokouhi, 2013]
  - Considers users personal information (only submitted history in AOL)
- Query-based VMM (QVMM) [He et al., 2009]
  - Context-aware query suggestion method
- Concept-based VMM (CACB) [Liao et al., 2011]
  - Concept-based context-aware query suggestion method
- Reformulation-based Completion (RC) [Jiang et al., 2014]
  - Model users' reformulation behavior for QAC

Six Competitive Baselines

- Most Popular Completion (MPC)
  - Maximum Likelihood Estimation (MLE) approach
- Hybrid Completion (Hyb.C) [Bar-Yossef et al., 2011]
  - Consider both context information and the popularity
- Personalized Completion (Per.C) [Shokouhi, 2013]
  - Considers users personal information (only submitted history in AOL)
- Query-based VMM (QVMM) [He et al., 2009]
  - Context-aware query suggestion method
- Concept-based VMM (CACB) [Liao et al., 2011]
  - Concept-based context-aware query suggestion method
- Reformulation-based Completion (RC) [Jiang et al., 2014]
  - Model users' reformulation behavior for QAC

Six Competitive Baselines

- Most Popular Completion (MPC)
  - Maximum Likelihood Estimation (MLE) approach
- Hybrid Completion (Hyb.C) [Bar-Yossef et al., 2011]
  - Consider both context information and the popularity
- Personalized Completion (Per.C) [Shokouhi, 2013]
  - Considers users personal information (only submitted history in AOL)
- Query-based VMM (QVMM) [He et al., 2009]
  - Context-aware query suggestion method
- Concept-based VMM (CACB) [Liao et al., 2011]
  - Concept-based context-aware query suggestion method
- Reformulation-based Completion (RC) [Jiang et al., 2014]
  - Model users' reformulation behavior for QAC

Six Competitive Baselines

- Most Popular Completion (MPC)
  - Maximum Likelihood Estimation (MLE) approach
- Hybrid Completion (Hyb.C) [Bar-Yossef et al., 2011]
  - Consider both context information and the popularity
- Personalized Completion (Per.C) [Shokouhi, 2013]
  - Considers users personal information (only submitted history in AOL)
- Query-based VMM (QVMM) [He et al., 2009]
  - Context-aware query suggestion method
- Concept-based VMM (CACB) [Liao et al., 2011]
  - Concept-based context-aware query suggestion method
- Reformulation-based Completion (RC) [Jiang et al., 2014]
  - Model users' reformulation behavior for QAC

Six Competitive Baselines

- Most Popular Completion (MPC)
  - Maximum Likelihood Estimation (MLE) approach
- Hybrid Completion (Hyb.C) [Bar-Yossef et al., 2011]
  - Consider both context information and the popularity
- Personalized Completion (Per.C) [Shokouhi, 2013]
  - Considers users personal information (only submitted history in AOL)
- Query-based VMM (QVMM) [He et al., 2009]
  - Context-aware query suggestion method
- Concept-based VMM (CACB) [Liao et al., 2011]
  - Concept-based context-aware query suggestion method
- Reformulation-based Completion (RC) [Jiang et al., 2014]
  - Model users' reformulation behavior for QAC

Overall Performance

$\#p$	MPC	Hyb.C	Per.C	QVMM	CACB	RC	Ours
1	0.1724	0.1796	0.1935	0.2028	0.1987	0.2049	**0.2140**
2	0.2703	0.2733	0.2770	0.2868	0.2828	0.2841	**0.2939**
3	0.4004	0.4025	0.4026	0.4066	0.4014	0.4122	**0.4193**
4	0.5114	0.5137	0.5129	0.5179	0.5126	0.5244	**0.5358**

Our approach outperforms all baselines with all prefix lengths

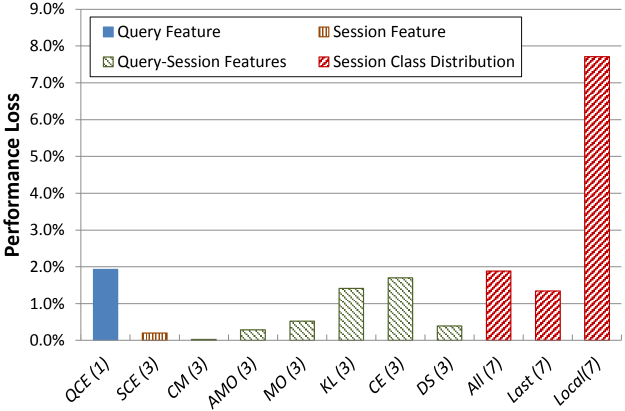
Performance and Context Lengths

	#p	Short	Medium	Long	Overall
RC	1	0.1842	0.2399	**0.2284**	0.2049
	2	0.2635	0.3196	**0.3076**	0.2841
Ours	1	**0.1966**	**0.2438**	0.2247	**0.2140**
	2	**0.2792**	**0.3226**	0.3036	**0.2939**
RC+Ours	1	**0.2055**	**0.2556**	**0.2439**	**0.2245**
	2	**0.2864**	**0.3356**	**0.3182**	**0.3024**

- Traditional context-aware baselines are stronger with longer contexts
- Our approach do better with shorter contexts
- Ensemble model can reach higher performance.

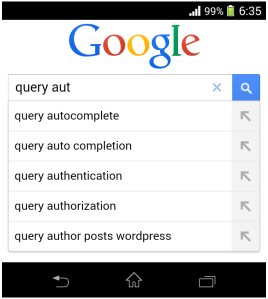
Feature Effective Analysis

Leave-one-out feature selection for analyzing feature effectiveness

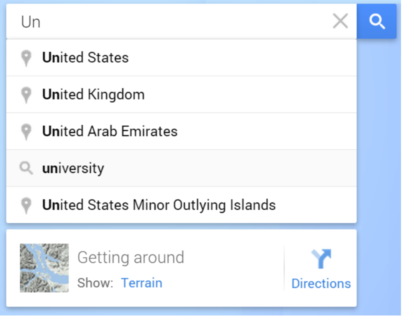


MRR is NOT intuitive for QAC

The key is to reduce users' keystrokes!



(a) Android Smartphone



(b) Google Maps

New Metric for Query Auto-Completion

Keystroke at top-k (KS@k)

- The average keystrokes users spend so that the actual queries can be found in the top-k queries

Measure	No Comp.	MPC	Hyb.C	Per.C
KS@1	11.0034	8.4294	6.8694	6.5761
KS@2	-	6.8625	5.6452	5.5078
KS@3	-	5.9830	4.9616	4.6965
KS@4	-	5.3038	4.5353	4.1793

Measure	QVMM	CACB	RC	Ours
KS@1	5.8704	6.1135	5.0129	**4.7479**
KS@2	4.1562	4.7813	3.9295	**3.6660**
KS@3	3.7044	4.0173	3.6523	**3.5880**
KS@4	3.6076	3.9138	3.5928	**3.5818**

Conclusions

- Propose a novel approach for query auto-completion
- Classify users' search intents in contexts by deriving class distributions
- Extensive experiments with six competitive baselines
- Propose a new metric for evaluating query auto-completion
- Our approach can reach good performance with only few contexts.
- Our approach can actually reduce users' keystrokes.

Q & A

Thanks for your attention.

Thank to SIGIR for the generous travel grant.