Context-Aware Query Recommendation by Learning High-Order Relation in query logs

Xiaohui Yan, Jiafeng Guo, Xueqi Cheng

Institute of Computing Technology, Chinese Academy of Sciences

{yanxiaohui, guojiafeng}@software.ict.ac.cn, cxq@ict.ac.cn

1. MOTIVATION

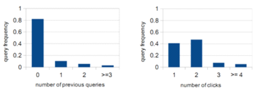
Query Recommendation(QR)

- QR is widely used in modern search engines to help users refine their queries and explore their nformation need.
- A good recommendation should match user's search intent.
- But queries are usually very short and ambiguous. How to catch user's search intent in this case is a big challenge.

Context-Aware QR

Methods	Query Context	Recommendation
query sequence pattern based	solar system → saturn	neptune, uranus, …
query-click pattern based	saturn → nineplanets.org	neptune, uranus, …

- **Tranditional ways: query sequence patterns based**
  - Do not address the ambiguity of queries explicitly
  - Query sequence patterns are scarce in query logs
- **Our way: query click-through patterns based**
  - Click-through also provides rich information, and can help to clarify user's search intent
  - Usually more than query sequence patters (see the bottom figure)



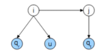
2. MODEL DETAIL

Problem Setting

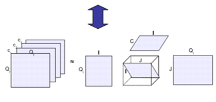
- Extracting all query click-through patterns from query sessions

- Training data
  - $\{<q_r,u,q_t>:n(q_r,u,q_t)\}$, $n(\cdot)$ is the frequency of $<q_r,u,q_t>$
- Objective
  - learning the joint probability $P(q_r, u, q_t)$, for all q_r, u, and q_t.

Our High-Order Model for QR



Graphical structure



An alternative view from tensor factorization

- **A genaerative view of the high-order model**
  - A user has search intent i with probability P(i).
  - He submits query q_r with probability $P(q_r|i)$
  - and then clicks a URL in result pages with probability $P(u|i)$
  - he derives another related search intent j with probability $P(j|i)$
  - and formulate the next query q_t
  - with probability $P(q_t|j)$

3. MODEL LEARNING

Parameters Estimation

- Rewrite the joint distribution

$$P(q_r,u,q_t)=\sum_{i,j}P(q_r|i)P(u|i)P(i,j)P(q_t|j).$$

- Likelihood function

$$l=\prod_{q_r,u,q_t}\left[\sum_{i,j}P(q_r|i)P(u|i)P(i,j)P(q_t|j)\right]^{n(q_r,u,q_t)}$$

with parameters $\Theta=\{P(q_r|i),P(u|i),P(i,j),P(q_t|j)\}$.

- Estimate Θ by EM

E-step:

$$P(i,j|q_r,u,q_t)=\frac{P(i,j)P(q_r|i)P(u|i)P(q_t|j)}{\sum_{i',j'}P(i,j)P(q_r|i')P(u|i')P(q_t|j')},$$

M-step:

$$P(i,j)=\frac{\sum_{q_r,u,q_t}n(q_r,u,q_t)P(i,j|q_r,u,q_t)}{\sum_{q_r',u',q_t'}n(q_r',u',q_t')},$$

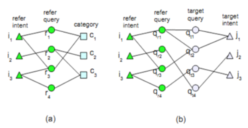
$$P(q_r|i)=\frac{\sum_{u,q_t,j}n(q_r,u,q_t)P(i,j|q_r,u,q_t)}{\sum_{q_r',u',q_t',j'}n(q_r',u',q_t')P(i,j'|q_r',u',q_t')},$$

$$P(u|i)=\frac{\sum_{q_r,q_t,j}n(q_r,u,q_t)P(i,j|q_r,u,q_t)}{\sum_{q_r',u',q_t',j'}n(q_r',u',q_t')P(i,j'|q_r',u',q_t')},$$

$$P(q_t|j)=\frac{\sum_{q_r,u,i}n(p_r,u,p_t)P(i,j|q_r,u,q_t)}{\sum_{q_r',u',q_t',i'}n(q_r',u',q_t')P(i',j|q_r',u',q_t')}.$$

Model Training

- Hidden Dimension Selection by max-diameter clustering methods
  - dimension of i: the number of clusters by clustering q_r's with feature $<q_t,u>$
  - dimension of j: the number of clusters by clustering q_t's with feature $<q_r,u>$
- By utilizing the clustering results to initialize the parameters, we can significantly reduce the number of parameters, save computation cost and alleviate overfitting
  - $P(q_r|i)$ and $P(q_t|j)$ are initialized by normalizing the weights of queries in each cluster.
  - $P(c|i)$ are initialized according nonzero $P(q_r|i)$ $P(q_t|j)$, and nonzero $n(q_r,u,q_t)$ (see below figure)



4. QUALITATIVE EXPERIMENTAL RESULTS

	i_{1935}	i_{19645}	j_{34564}	j_{29915}
queries	gmc vehicles **saturn** volvo suv honda chevrolet	**saturn** neptune saturn planet mercury venus	**saturn** mustang ford nissan hyundai	solar system planet mercury earth planets in order saturn
categories	Business/Automotive	Science/Astronomy		

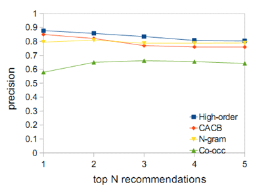
Examples for refer intents and target intents involve the query "saturn"

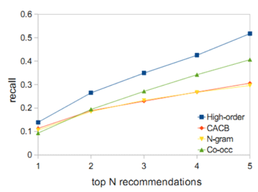
query context	Co-occurrence	N-gram	CACB	High-order
saturn ↓click nineplanets.org	solar system uranus diameter uranus saturn cars saturn roadster	pontiac solar system saturn cars scion kia	toyota mitsubish honda hyundai solar system	uranus uranus diameter solar system earth astrology
honda ↓submit saturn ↓click saturn.com	solar system uranus diameter uranus saturn cars saturn roadster	pontiac saturn cars toyota	toyota pontiac scion gm com	saturn sky saturn roadster saturn cars pontiac mazda

Query recommendation examples with different context

[1] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li. Context-aware query suggestion by mining click-through and session data. KDD2008,

5. QUANTITATIVE EXPERIMENTAL RESULTS





- Precision
  - the High-order outperforms all the three others.
  - context-aware methods (High-Order, CACB, N-gram) beat context-free method (Co-occurrence) signifiantly.
- Recall
  - High-order method significantly outperforms the other methods due to the sparsity of the query sequence patterns in query logs.
  - High-order method clearly benefit from the rich click-through information.

6. CONCLUSIONS

Main Contributions:

- We propose a novel context-aware query recommendation approach by modeling the high-order relation between queries and clicks in query log, which captures users' latent search intents in an unsupervised way.
- We show our model can be solved efficiently by EM algorithm, with a strategy for parameter initialization by clustering, which can reduce a large number of trival parameters.
- Empirical experiment results demonstrate that our approach outperforms the baseline methods in providing high quality recommendations for ambiguous queries.

Future Work:

- For the future work, it is interesting to combine the query sequence patterns and click-through information to conduct a more general context-aware query recommendation.
- Further regularization techniques are considered to be added in, since the number of parameters is still very huge in the high-order model.