Query Performance Prediction for Microblog Search

A Preliminary Study

Maram Hasanain, Rana Malhas, Tamer Elsayed

11 July 2014

SoMeRA’14 Workshop in conjunction with SIGIR’14
Why?

**Expectation**
High quality results

**Reality**
Some queries are difficult

*Poor results*
What’s QPP?

Query

Retrieval model

Result list ($R$)

sigir awards

Query Performance Prediction (QPP)

estimated performance
QPP in Microblog Search?

- QPP is not a new problem

**RQ1:** How well the existing state-of-the-art predictors perform in the context of microblog search?

- Microblog search is *different*

**RQ2:** Will the predictors' performance be consistent across different retrieval models, specifically *temporal ones*?
Setup of the Study ...
Overview

• Examine *frequently-used* predictors for *tweets search*

• 2 types of *predictors*:
  o *Content-based*: consider terms in tweets and queries
  o *Temporal*: also consider time factor

• 2 types of *retrieval models*:
  o *Content-based*
    e.g. Query Likelihood
  o *Temporal*
    e.g. Time-based Exponential Priors
QP Predictors

Content-based predictors

• Standard deviation ($\sigma$)
  o Normalized Standard Deviation (NSD)
  o Normalized Query Commitment (NQC)

• KL-divergence
  o Clarity (CLR)

• Information Gain
  o Weighted information gain (WIG)
QP Predictors

• Inverse document frequency (IDF)
  $\text{SumIDF}, \text{MaxIDF}, \text{AvgIDF}, \ldots$

• Collection-query similarity (SCQ)
  $\text{SumSCQ}, \text{MaxSCQ}, \text{AvgSCQ}, \ldots$

• Simplified clarity score (SCS)
QP Predictors

Temporal predictor

• KL-divergence
  Temporal Clarity ($t$-CLR)

Post-retrieval
Retrieval Models

Content-based

• Query Likelihood (QL)

\[ P(D|Q) \propto P(Q|D) \cdot P(D) \]

Temporal

• QL with temporal prior (t-EXP)

\[ P(D|Q) \propto P(Q|D) \cdot r \cdot e^{-r \cdot t_d} \]

• Temporal relevance modeling (t-QRM)

\[ P(w|Q) = \sum_{t \in T} P(w|t, Q)P(t|Q) \]

\[ P(w|t, Q) = \sum_{D \in t} P(w|D)P(D|t, Q) \]
Evaluation
Setup

Datasets

<table>
<thead>
<tr>
<th>Source</th>
<th>Tweets2011</th>
<th>Tweets2013</th>
</tr>
</thead>
<tbody>
<tr>
<td>TREC’11-12</td>
<td>~16M</td>
<td>~243M</td>
</tr>
<tr>
<td>Tweets</td>
<td>108</td>
<td>60</td>
</tr>
<tr>
<td>Queries</td>
<td>~2 weeks</td>
<td>~2 months</td>
</tr>
</tbody>
</table>

Evaluating retrieval

Evaluation measure: Average precision (AP)
Setup

Evaluating prediction

• **Correlation** between *predicted AP* & *actual AP*.

• Linear correlation: **Pearson’s $r$**

• Rank correlation: **Kendall’s $\tau$**

Training/Testing

• 75% of queries for parameter tuning

• Repeat and average with 120 trials
Results (Tweets2011)

- **t-CLR is best**
- **NQC: Increase in performance**
- **Not significant**
- **SumIdf: Comparable quality**
- **CLR: Decline in quality**
- **WIG: Decline in quality**

**Pearson’s correlation**

**Retrieval model**

- QL
- t-EXP
- t-QRM
Results (Tweets2013)

- CLR is best when using Pearson's correlation.
- NQC: Increase in performance
- t-CLR has good performance
- Not significant
Combining Predictors

• Using linear regression
• Feature selection to find best predictors combination
• Only over Tweet2011
• 40% of queries for parameter tuning
• Train & test combined model by cross-validation with 60% of queries.
Combining Predictors (Tweets2011)

- QL: Pearson's correlation
  - Combined: 0.42
  - Best: 0.21

- t-EXP: Pearson's correlation
  - Combined: 0.56
  - Best: 0.27

- t-QRM: Pearson's correlation
  - Combined: 0.60
  - Best: 0.46

- t-CLR in best combinations: 21.6%
- Pre-retrieval predictors in best combinations:
  - Combined: 46.5%

Predictors:
- t-CLR, CLR, WIG, SCS
- t-CLR, WIG, SCS
- t-CLR, NQC, NSD, SumIDF
Summary

• First comprehensive study focusing on testing QPP in microblog search with different retrieval models.

• Temporal predictors might be more suitable for microblog search.

• Combining predictors improved prediction quality.

• Some pre-retrieval predictors are showing promising results.
Future Work

• *Experiment* with more *temporal predictors & retrieval models*

• *Develop* new...
  o *Temporal* predictors
  o Predictors considering *tweet-specific* features

• *Use* QPP in ...
  o *Selective* query expansion
  o *Dynamic* query expansion
Thank You 😊