Improving Search Relevance for Short Queries in Community Question Answering

A joint research by Microsoft and USTC

Presented by: Huang Zonghao
Community Question
Answering
Community Question Answering

stackoverflow
YAHOO! ANSWERS
Quora
MATHEMATICS
Chegg
Baidu
Cwiki
Topic

• Short question query on CQA
  • Users tend to type in short questions while searching on CQA
  • The website is supposed to get users intention behind the short questions
    • To promote the questions
    • To give the suggestions
the new macbook

New macbook........
I am buying a new macbook even though i just bought mine about 4-5 months ago. But when i buy it i want to transfer the old to the new, and then i want to clear off my old one. so when i sell it th...
1 Answer · Laptops & Notebooks

Is the new MacBook...?
really worth it? how does it compare to the old solid white one? what better features does it have? i'm not really a computer whizz when it comes to technical terms so try to explain it plainly.
1 Answer · Laptops & Notebooks

Should I regret buying a new Macbook?
I just bought a brand new macbook yesterday. first of all, i'm finding it VERY confusing since i'm so used to my pc. second, i was excited about being able to edit my home dvds. what they...
5 Answers · Laptops & Notebooks

If i get a new macbook?
If i get a new macbook, can i transfer all my itunes purchases to the new computer? thanks! thats great and all, but i want a macbook so can you answer my question?
5 Answers · Laptops & Notebooks

Buying a new macbook?
Hello. I want to buy a new macbook (the white one), because I don't want to spend more money on the Macbook Pro as I do not need what is much better and I prefer the new design. Can anyone help?
• Examine how to improve search relevance for short queries in CQA question search
Difficulties

- Short — lack of information
- Community — lack of universality
- Need to find both adequate samples and accurate models
Approach

- User intent mining — built up the database
- Model tuning
- Combine the result of both source
User Intent Mining

- Two expectations
  - The most interesting and the most important aspect of the query
  - The most popular subtopic of the query
Three Different Sources

- CQA archives
- Web Search Logs
- Top search results from a commercial search engine
CQA Archives

• Use the relation between the question and description
  • treat the questions as sources and the corresponding descriptions as targets
  • term by term translation model
  • \( P_{cqa}(t|q) = \varepsilon \sum_{w \in q} P_{tp}(t|w)P_{ml}(w|q) + (1 - \varepsilon)P_{ml}(t|C) \)
  • rank the terms and generate the intent word set

\[ W = \{(t, \varphi)\} \]
Query Log

• Users click reflects their intents
• Find the relation between the input keywords and the clicks
• generate the intent word set

\[ W = \{(t, \varphi)\} \]
Web Search Results

- Capture the searching results from the commercial searching engine
  - for time sensitive factors
  - use scores to rank each term
Input: Query $q$, top $M$ search results $R$, window size $l$, weight parameters $\eta, \sigma, \tau$
Output: Intent term set $W = \{(t, \varphi)\}$

1: $H$ $\leftarrow$ titles of documents in $R$
2: $S$ $\leftarrow$ snippets of documents in $R$
3: $U$ $\leftarrow$ URLs of documents in $R$
4: $A$ $\leftarrow$ concatenate $T, S$ and $U$ as a single string;
5: $L$ $\leftarrow$ length of $A$
6: $F$ $\leftarrow$ $\{(t, f)\} = \emptyset$
7: for $i$: 1 to $L$ do
8: if $A[i] \in q$ then
9: for $j$: $-l$ to $l$ do
10: if $F$ contains key $A[i + j]$ then
11: $F[A[i + j]]$ $\leftarrow$ $F[A[i + j]] + 1$
12: else
13: $F$ $\leftarrow$ $F \cup \{(A[i + j], 1)\}$
14: $W$ $\leftarrow$ $\{(t, \varphi)\} = \emptyset$
15: for each $(t, \varphi) \in F$ do
16: $\varphi$ $\leftarrow$ 0
17: for each $h \in H$ do
18: $\varphi$ $+$ $\leftarrow$ $\eta \cdot BM25(t, h, H)$
19: for each $s \in S$ do
20: $\varphi$ $+$ $\leftarrow$ $\sigma \cdot BM25(t, s, S)$
21: for each $u \in U$ do
22: $\varphi$ $+$ $\leftarrow$ $\tau \cdot BM25(t, u, U)$
23: $\varphi$ $\leftarrow$ $\varphi \cdot f$
24: $W$ $\leftarrow$ $W \cup \{(t, \varphi)\}$
25: rank $W$ in descending order of $\varphi$
26: return $W$
Comparison of three Sources

- CQA — more descriptions, comments and suggestions
- Query log — more information based
- Web search result — more time sensitive

<table>
<thead>
<tr>
<th>Query</th>
<th>Intent words from CQA</th>
<th>Intent words from query log</th>
<th>Intent words from search results</th>
</tr>
</thead>
<tbody>
<tr>
<td>usain bolt</td>
<td>fastest, world, record, olympics</td>
<td>biography, twitter, girlfriend</td>
<td>2013, gold, moscow, championship</td>
</tr>
<tr>
<td>superbowl</td>
<td>patriots, steelers, giants, nfl</td>
<td>story, history, 2012, ticket, xlvi</td>
<td>ticket, 2013, nfl, history, commercial</td>
</tr>
<tr>
<td>egypt</td>
<td>cairo, country, arabic, pyramids</td>
<td>morsi, election, ancient, history</td>
<td>revolution, brotherhood, police</td>
</tr>
</tbody>
</table>
Models

- Language model for information retrieval
- Translation based language model
- Intent based Language model
Language model for information retrieval

\[ P(q|Q) = \prod_{w \in q} [(1 - \lambda)P_{ml}(w|Q) + \lambda P_{ml}(w|C)] \]

- Works well when there is a great deal of overlap between a query and a candidate question
- but fail otherwise
Translation based language model

\[ P_{trb}(q|Q) = \prod_{w \in q} [(1 - \lambda)P_{mx}(w|Q) + \lambda P_{ml}(w|C)] \]

where

\[ P_{mx}(w|Q) = \alpha P_{ml}(w|Q) + \beta P_{tr}(w|Q) \]

\[ P_{tr}(w|Q) = \sum_{v \in Q} P_{tp}(w|v)P_{ml}(v|Q). \]
A little improvement

\[ P_{trba}(q|Q) = \prod_{w \in q} [(1 - \lambda)P_{mx}(w|Q, a) + \lambda P_{ml}(w|C)], \]

where

\[ P_{mx}(w|Q, a) = \alpha P_{ml}(w|Q) + \beta P_{tr}(w|Q) + \gamma P_{ml}(w|a) \]

\[ \gamma \text{ is an extra parameter satisfying} \quad \alpha + \beta + \gamma = 1. \]

- work well for the long questions
Intent Based Language Model

\[ P_{ib}(q|Q) = \pi_0 P_{trba}(q|Q) + \sum_{i=1}^{3} \pi_i \sum_{j=1}^{N} \varphi_{ij} P_{trba}(t_{ij}|Q) \]

\[ W_i = \{(t_{ij}, \varphi_{ij})\}, \ 1 \leq i \leq 3, \ 1 \leq j \leq N \]

- a combination of translation based language model and the preprocessed intent word set
Experiment

- Collect the CQA data from Yahoo Answers
- Recruit human judges to label the relevance of the candidate questions regarding the queries
### Table 2: Overview of two CQA data sets

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Yahoo</th>
<th>Quora</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question #</td>
<td>127,787,139</td>
<td>649,843</td>
</tr>
<tr>
<td>Description #</td>
<td>103,605,696</td>
<td>375,829</td>
</tr>
<tr>
<td>Answer #</td>
<td>894,855,746</td>
<td>1,743,259</td>
</tr>
</tbody>
</table>

### Table 3: Length distribution of the labeled queries

<table>
<thead>
<tr>
<th>Total</th>
<th>Query Length</th>
<th>Avg. Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total 1 2 3 ≥4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Length</td>
<td></td>
</tr>
<tr>
<td>1782</td>
<td>658 732 289 103</td>
<td>1.94</td>
</tr>
</tbody>
</table>

### Table 4: Overview of two labeled data sets

<table>
<thead>
<tr>
<th>Statistics</th>
<th>Yahoo</th>
<th>Quora</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queries #</td>
<td>1,782</td>
<td>1,782</td>
</tr>
<tr>
<td>Questions #</td>
<td>12,947</td>
<td>13,739</td>
</tr>
<tr>
<td>Questions #/query</td>
<td>7.27</td>
<td>7.71</td>
</tr>
</tbody>
</table>
## Result

### Table 6: Evaluation results on Yahoo data and Quora data

<table>
<thead>
<tr>
<th></th>
<th>Yahoo data</th>
<th></th>
<th>Quora data</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NDCG@1</td>
<td>NDCG@3</td>
<td>NDCG@5</td>
<td>NDCG@1</td>
</tr>
<tr>
<td>LDA[18]</td>
<td>63.35</td>
<td>69.91</td>
<td>72.50</td>
<td>55.72</td>
</tr>
<tr>
<td>LMIR[23]</td>
<td>67.02</td>
<td>73.83</td>
<td>76.07</td>
<td>60.72</td>
</tr>
<tr>
<td>TR[4]</td>
<td>68.27</td>
<td>73.84</td>
<td>75.98</td>
<td>60.52</td>
</tr>
<tr>
<td>TBL[15]</td>
<td>68.13</td>
<td>73.89</td>
<td>76.37</td>
<td>60.97</td>
</tr>
<tr>
<td>TAL[28]</td>
<td>69.44</td>
<td>74.95</td>
<td>76.53</td>
<td>61.96</td>
</tr>
<tr>
<td>TAL+LDA</td>
<td>70.03</td>
<td>74.02</td>
<td>75.92</td>
<td>62.71</td>
</tr>
<tr>
<td>IBLM</td>
<td><strong>71.33</strong></td>
<td><strong>77.12</strong></td>
<td><strong>77.70</strong></td>
<td><strong>64.04</strong></td>
</tr>
</tbody>
</table>
Conclusion
Q&A