
Improving Question Retrieval in CQA with Label Ranking

Presented by WONG Wei

2010-10-5

Outline

- Problem & Motivation
 - Label Ranking
 - Improved QR Based on Label Ranking
 - Experiment Design
-

Question Retrieval

The image shows a search engine interface with a search bar containing the text "what can we do to encourage people to start recycling" and a "Search Y! Answers" button. Below the search bar, there are sorting options: "Sort by: Relevance | Newest | Most Answers". Three search results are listed, each with a smiley face icon, a title, a snippet of text, and a star icon indicating the category and date.

query

Search

what can we do to encourage people to start recycling **Search Y! Answers**

Sort by: Relevance | Newest | Most Answers

What can we do to encourage people to start recycling?
★ In Green Living - Asked by Dr. Madelyn - 5 answers - 1 week ago

I'm doing a project in school on recycling?
...down a bit and **did recycling** plastics. **We have to** make an advertisement...something about how an ad **can help encourage people to do what** you are trying...persuade them **to do to start** with. **We have to** have a picture...
★ In Green Living - Asked by NatWolfflover1417 - 3 answers - 1 year ago

As a parent, what's your opinion of these rules?
...fish still "speak", **we** just don't understand their language. This is **to** be respected. Take... reduce/reuse/**recycle** habits. 6. I'...never personally seen you **do** this, and I don't believe... inclusion... if she **starts to** stutter...dell tells her, and allows other **people to** tell her, "Come on..., repeat back **to what** she had so far. **Encourage** other **people to do** the same... that dragons **can** be killed. So, if...
★ In Parenting - Asked by What Dreams May Come - 18 answers - 3 months ago

Existed similar questions

Existing Methods of QR

- Vector Space Model
 - Okapi BM25 Model
 - Language Model
 - Translation Model
 - Translation-Based Language Model
-

An Example

The screenshot shows the Yahoo! Answers interface. At the top, there is a green header with the 'Ask' button and a search bar containing 'What would you like to ask?'. Below this is a white search bar with the text 'What are you looking for?' and a 'Search Y! Answers' button. The breadcrumb trail reads: 'Home > All Categories > Environment > Green Living > Resolved Question'. The main content area features a 'Resolved Question' section with a green background. The question is 'What can we do to encourage people to start recycling?' asked by 'Dr. Madelyn' 1 week ago. A 'Report Abuse' button is visible below the question. To the right of the question is a 'Show me another »' link. Below the question is the 'Best Answer' section, marked 'Chosen by Asker', by 'Ashley Johnson' 1 week ago. The answer text reads: 'Tell them that they can earn money by doing it. Tell them they could even save lives! The Ronald Mcdonald house saves pop tabs (tabs on top of a coke can) to pay bills for their place and then certain lids on top of yogurt lids can be recycled for a cure towards breast cancer. I hope this helps! Its not easy getting people to understand all this!'.

What's the asker really wants to get from others?

Opinions?

Instructions?

Recommendation?

An Example

The screenshot shows the top navigation bar of Yahoo! Answers with 'Ask' and 'Ans' buttons. Below is a search bar with the text 'What are you looking for?' and a 'Search Y! Answers' button. The breadcrumb trail reads 'Home > All Categories > Arts & Humanities > Dancing > Resolved Question'. The main content area features a 'Resolved Question' section with a green header. The question is 'What is a good solo song?!?!?!?' by Gabby Raw, posted 1 week ago. The 'Best Answer' is chosen by the asker, thelaraw..., and suggests 'The Sound Of Music' and 'Annie'.

Home > All Categories > Arts & Humanities > Dancing > Resolved Question

Resolved Question [Show me another »](#)

What is a good solo song?!?!?!?

Gabby Raw
Hi! I need a solo song for a 12 year old! It can be jazz or musical theater! THANKS!!!
Gabby <3
1 week ago
[Report Abuse](#)

Best Answer - Chosen by Asker

thelaraw...
The Sound Of Music:
16 going on 17 is nice, but you're 12...lol :)
Annie:
Tomorrow, Maybe, It's A Hard-Knock Life...basically any song from Annie

What's the asker really wants to get from others?

Recommendation?

How-to question?

Definition?

Motivation

- Existing methods does not consider users' different intent.
 - Multi-label classification.
 - Preferences among different labels.
-

Question Classification Based on User's Intention

Table 2: Types of questions

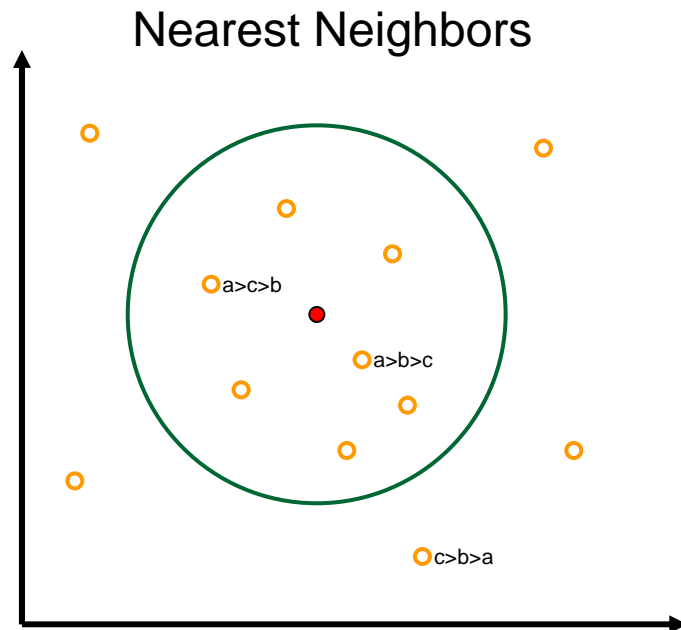
Type	Purpose	Expected answer
REC	Recommendation	Suggested item and the reason
FAC	Seeking facts	Data, location, or name
YNO	Yes/No decision	Yes/No and reason
HOW	How-to question	The instructions
WHY	Seeking reason	The explanation
TSL	Translation	Translated text in target language
DEF	Definition	The definition of given entity
OPN	Seeking opinion	Opinions of other users
TRA	Transportation	Navigation instructions
INT	Interactive	Discussion thread
MAT	Math problem	Solution with steps

- By Xiance Si et.al's work on VLDB 2010

Label Ranking (Formally)

- Given:
 - a set of training instances: $\{\mathbf{x}_k \mid k = 1 \dots m\} \subseteq \mathbf{X}$
 - a set of labels $L = \{l_i \mid i = 1 \dots n\}$
 - for each training instances \mathbf{x}_k : a set of pairwise preferences of the form $l_i \succ_{\mathbf{x}_k} l_j$ (for some of the labels)
 - Find:
 - a ranking function ($\mathcal{X} \rightarrow \Omega$ mapping) that maps each $\mathbf{x} \in \mathbf{X}$ to a ranking $\succ_{\mathbf{x}}$ of L (Permutation $\pi_{\mathbf{x}}$) and generalizes well in terms of a loss function on rankings (e.g. , kendall's tau coefficient)
-

Instanced Based Label Ranking



- Distribution of ranking is (approx.) constant in a local region.
- **Core part** is to estimate the locally constant model.

Instance Based Label Ranking

- Mallows model (Mallows, Biometrika, 1957)

$$\mathcal{P}(\sigma|\theta, \pi) = \frac{\exp(-\theta d(\pi, \sigma))}{\phi(\theta, \pi)}$$

- with
 - center ranking $\pi \in \Omega$
 - spread parameter $\theta > 0$
 - and $d(\cdot)$ is a metric on permutations
- Computational issues** arise when the training data contains **incomplete rankings**.
-

Instance Based Label Ranking

- Approximation via a variant of EM, viewing the non-observed labels as hidden variables.
 1. Start with an initial center ranking (via generalized Borda count)
 2. Replace an incomplete observation with its most **probable extension** (first M-step, can be done efficiently)
 3. Obtain MLE as in the complete ranking case (second M-step)
 4. Replace the initial center ranking with current estimation
 5. Repeat until convergence

Probable Extension:

Observation σ	Extensions $E(\sigma)$
$a _ b$	$a _ b _ c$ $a _ c _ b$ $c _ a _ b$

Improved QR Based on Label Ranking

- Linear combine label ranking with QR

$$S_{q,u} = \gamma CON_{q,u} + (1 - \gamma) INT_{q,u},$$

- With

$CON_{a,u}$ is the content similarity,
 $INT_{q,u}$, is the intent similarity.

Experiment Design

- **Data Analysis**
 1. Number of relative question for each query
 2. Demonstrate the classification system degree
 3. Average number of labels for each question
 - **Classification methods** for comparison:
 - SVM & Naive Bayes & LR
 - **LR** + Translation Model/Language Model
/Translation-Based Language Model
 - **SVM**+ Translation Model/Language Model
/Translation-Based Language Model
 - **Naive Bayes**+ Translation Model/Language Model
/Translation-Based Language Model
-

Thanks!
Question and Comments?
