

CSCI5070 Advanced Topics in Social Computing

Community Question Answering

Irwin King

The Chinese University of Hong Kong

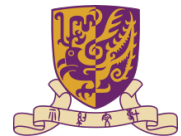
king@cse.cuhk.edu.hk

©2012 All Rights Reserved.

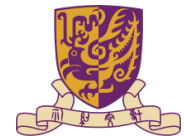


Outline

- Introduction
- Question Retrieval
- Question Recommendation
- Question Subjectivity Analysis
- Content Quality Evaluation



Introduction



Yahoo! Answers

[Home](#) > [All Categories](#) > [Consumer Electronics](#) > [Land Phones](#) > Resolved Question



Anna!!!

Resolved Question

[Show me another »](#)

Why do peoples' voices sound different when they're talking on the phone?

Some people say I sound like my mom when I'm talking to them on the phone, which I think is sort of weird... because I was adopted... Today I was talking to my boyfriend on the phone... This was the first time I've talked to him on the phone... (we've only been going out for like a week). His voice sounded a little deeper or something. Or could that just be because he was nervous?

4 years ago

[Report Abuse](#)



Paul_196...

Best Answer - Chosen by Asker

One major reason for voices sounding different is that the frequency response of the telephone system is limited. The range of the human ear can extend right up 20kHz or more, especially in younger people. A connection over the telephone has a much narrower bandwidth, typically restricting the highest frequencies transmitted to a little over 3kHz in many cases.

That's adequate to convey intelligible speech, but naturally it changes the sound of the voice subtly by filtering out the highest-pitched components. It's the same sort of effect as you would get by listening to your favorite record on A.M. radio versus listening to it on F.M. or from a CD.

The telephone also reduces frequencies at the very lowest end of the audible range as well.

4 years ago

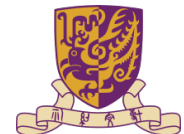
[Report Abuse](#)

259 people rated this as **good**

Asker's Rating: *****

Thanks =)

Action Bar: 265 Interesting! Email Comment (9) Save



Stack Overflow

[Artificial Neural Networks](#) [Machine Learning](#) [Edit](#)

How do convolutional neural networks work? [Edit](#)

Especially, what kind of benefits does convolution give you? [Edit](#)

[Comment](#) · [Post \(1\)](#) · [Wiki](#) · [Options](#) · [Redirect](#) · [Question](#)

2 Answers · [Create Answer Wiki](#)

 **Mikio L. Braun, Ph.D. in machine learning, 10+ years ...** 

 3 votes by [Kat Li](#), [Barak Cohen](#), and [Lucian Sasu](#)

Convolutional neural networks work like learnable local filters.

The best example is probably their application to computer vision. The first step in image analysis is often to perform some local filtering of the image, for example, to enhance edges in the image.

You do this by taking the neighborhood of each pixel and convolve it with a certain mask (set of weights). Basically you compute a linear combination of those pixels. For example, if you have a positive weight on the center pixel and negative weights on the surrounding pixels you compute the difference between the center pixel and the surrounding, giving you a crude kind of edge detector.

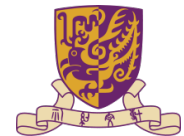
Now you can either put that filter in there by hand or learn the right filter through a convolutional neural network. If we consider the simplest case, you have an input layer representing all pixels in your image while the output layer representing the filter responses. Each node in the output layer is connected to a pixel and its neighborhood in the input layer. So far, so good. What makes convolutional neural networks special is that the weights are shared, that is, they are the same for different pixels in the image (but different with respect to the position relative to the center pixel). That way you effectively learn a filter, which also turns out to be suited to the problem you are trying to learn.

[Comment](#) · [Post](#) · [Thank](#) · Sep 29, 2011

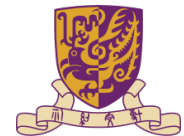
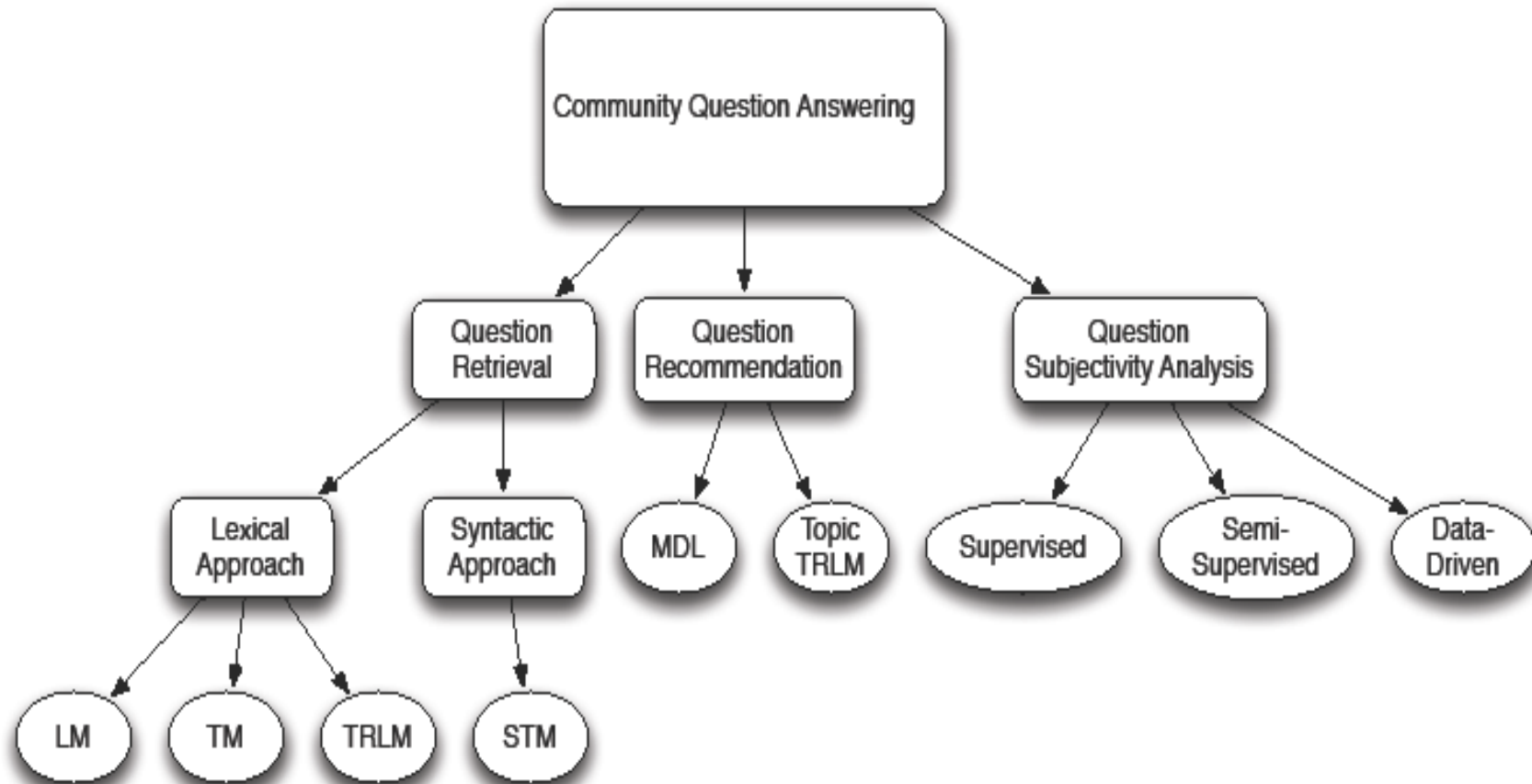


Advantages of CQA

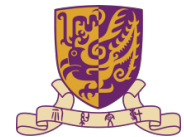
- Could solve information needs that are **personal, heterogeneous, specific, open-ended**, and **cannot** be expressed as a **short query**
- **No single Web page** will directly answer these complex and heterogeneous needs, **CQA users** should understand and answer better than a machine
- Have accumulated rich knowledge
 - More than one billion posted answers in Yahoo! Answers
<http://yanswersblog.com/index.php/archives/2010/05/03/1-billion-answers-served/>
 - More than **190 million resolved questions** in Baidu Zhidao
 - In China, **25%** of Google's top-research-results page contain at least one link to some Q&A site, Si et al., VLDB, 2010



Covered Topics



QUESTION RETRIEVAL



Ask A Question

[Home](#) > Ask Question

1 What's Your Question

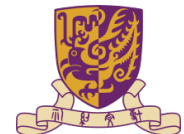
What should i do if my laptop got blue screen?

You have 64 characters left.

Now add a little more detail (optional)

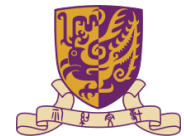
Make sure your question follows the [community guidelines](#).

Continue



Problem and Opportunity

- Problem
 - Askers need to wait some time to get an answer, **time lag**
 - **15%** of the questions **do not receive any answer** in Yahoo! Answers, which is one of the first CQA sites on the Web
- Opportunity
 - **25%** questions in certain categories are recurrent, **Anna, Gideon and Yoelle, WWW, 2012**
- Answer **new questions** by reusing past resolved questions
- **Question Retrieval**: find **semantically similar** past questions for a new question



Question Retrieval Example

Search

What should i do if my laptop got blue screen?

Search Y! Answers

Sort by: Relevance | Newest | Most Answers



What should I do if I keep getting the "blue screen of death" for my Windows7 laptop?

...I keep **getting** the **blue screen** of death telling **me** that the pc is **getting** prepared for a... scary to imagine **what** would happen **if I** wasn't. I just bought this Windows7 Toshiba **laptop** from office depot in the summer...to crash (so early)? **What should I do?**

☆ In Laptops & Notebooks - Asked by nelson316@verizon.net - 4 answers - 4 months ago



I just got a random blue screen of death, should I be worried?

...just suddenly **got** a random **blue screen** of death. I've never **got** one on this **laptop** before, and I've had no problems with **my laptop** at all until this bsod.... It said that **if** it was **my** first time...free. I don't even remember **what** sites I was on...with no problems. I **do** remember that the programs...off bsod like **my old laptop?** Or **should I** be worried? ...

1 ☆ In Other - Hardware - Asked by Kaylee - 6 answers - 2 weeks ago



why is my laptop showing the blue screen?

...to it, so I'm not sure **if** they could have **done** anything, but now when **i** turn **my laptop** on **i** would **get** a **blue screen** saying all this jumble...a boot disk, so **i** dont know **what** else **should i do**. Any help/advice?

☆ In Laptops & Notebooks - Asked by doodiec - 6 answers - 5 years ago



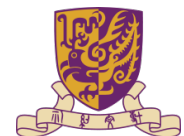
Laptop blue screen problem!!!?

...malicious URL block and then this **blue screen** comes up and **my laptop** turns off and asks **me** **if I** want to go into safe mode. **What should I do?** Is there any way...for a new **laptop** cause **I got** low practice SAT scores...

☆ In Laptops & Notebooks - Asked by Mathew Colman - 5 answers - 10 months ago

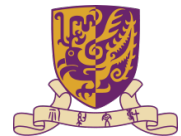


sony vaio **blue screen** problem, **what should i do?** please help?



Benefit of Question Retrieval

- Provide an alternative to **automatic question answering**
- Help askers get an answer in a **timely manner**
- Guide answerers to answer **unique questions**, better utilize users' answering passion








Notations

| Symbol | Description |
|-------------|--------------------------------|
| Q | A new question |
| D | A candidate question |
| $ \cdot $ | Length of the text |
| C | Background collection |
| w | A term in the new question |
| t | A term in a candidate question |

Search

Sort by: [Relevance](#) | [Newest](#) | [Most Answers](#)

-  **What should I do if I keep getting the "blue screen of death" for my Windows7 laptop?**
 ...I keep **getting** the **blue screen** of death telling **me** that the pc is **getting** prepared for a... scary to imagine **what** would happen **if I** wasn't. I just bought this Windows7 Toshiba **laptop** from office depot in the summer...to crash (so early)? **What should I do?**
 ☆ In Laptops & Notebooks - Asked by nelson316@verizon.net - 4 answers - 4 months ago
-  **I just got a random blue screen of death, should I be worried?**
 ...just suddenly **got** a random **blue screen** of death. I've never **got** one on this **laptop** before, and I've had no problems with **my laptop** at all until this bsod.... It said that **if** it was **my** first time...free. I don't even remember **what** sites **I** was on...with no problems. **I do** remember that the programs...off bsod like **my old laptop**? Or **should I** be worried? ...
 1 ☆ In Other - Hardware - Asked by Kaylee - 6 answers - 2 weeks ago
-  **why is my laptop showing the blue screen?**
 ...to it, so I'm not sure **if** they could have **done** anything, but now when **i** turn **my laptop** on **i** would **get a blue screen** saying all this jumble...a boot disk, so **i** dont know **what else should i do**. Any help/advice?
 ☆ In Laptops & Notebooks - Asked by doodiec - 6 answers - 5 years ago
-  **Laptop blue screen problem!!!?**
 ...malicious URL block and then this **blue screen** comes up and **my laptop** turns off and asks **me if I** want to go into safe mode. **What should I do?** Is there any way...for a new **laptop** cause **I got** low practice SAT scores...
 ☆ In Laptops & Notebooks - Asked by Mathew Colman - 5 answers - 10 months ago
-  **sony vaio blue screen problem, what should i do? please help?**

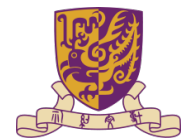


Lexical-based Approach: Language Model

- In language modeling, similarity between a **query** and a **document** is given by the **probability** of generating the query from the document language model
- Unigram language model, i. i. d. sampling

$$P(Q|D) = \prod_{w \in Q} P(w|D)$$

- In **question retrieval** syntax, query is the **new question**, document is a **candidate question**



Lexical-based Approach: Language Model

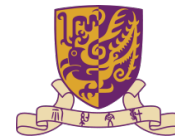
- To avoid zero probabilities and estimate more accurate language models, documents are smoothed using a background collection

$$P(w|D) = (1 - \lambda)P_{ml}(w|D) + \lambda P_{ml}(w|C)$$

$$P_{ml}(w|D) = \frac{\text{termfrequency}(w, D)}{\sum_{w' \in D} \text{termfrequency}(w', D)}$$

$1 \geq \lambda \geq 0$ is a smoothing parameter, $0 \lambda -$

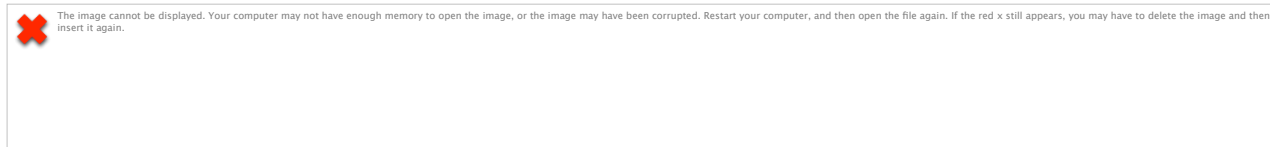
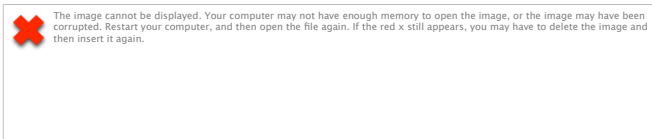
– Maximum likelihood estimator to calculate $P_{ml}(\cdot)$



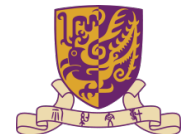
Language Model Example

- Query (q): revenue down
- Document 1 (d1): xyzy reports a profit but revenue is down
- Document 2 (d2): quorus narrows quarter loss but revenue decreases further

= 0.5 λ •



- Ranking: $d1 > d2$



Lexical-based Approach: Translation Model

- Language Model
 - Advantage: Simple
 - Disadvantage: Lexical Gap
- **Lexical Gap**, two questions that have the same meaning use very different wording
 - Is downloading movies illegal?
 - Can I share a copy of a DVD online?
- Jiwoon Jeon, W. Bruce Croft and Joon Ho Lee, Finding Similar Questions in Large Question and Answer Archives, CIKM, 2005



Lexical-based Approach: Translation Model

| |
|---|
| language Model |
| $P(w D) = (1 - \lambda)P_{ml}(w D) + \lambda P_{ml}(w C)$ |
| Translation Model |
| $P(w D) = (1 - \lambda) \sum_{t \in D} (T(w t)P_{ml}(t D)) + \lambda P_{ml}(w C)$ |

- $T(w | t)$ is the **probability** that word w is the **translation** of word t , denotes **semantic similarities** between words



Example

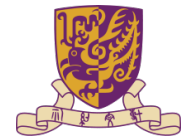
Table: Questions share few common words, but may have high semantic relatedness according to translation model

| |
|---|
| Id like to insert music into PowerPoint. How can I link sounds in PowerPoint? |
| How can I shut down my system in Dos-mode. How to turn off computers in Dos-mode. |
| Photo transfer from cell phones to computers. How to move photos taken by cell phones. |
| Which application can run bin files ? I download a game. How can I execute bin files ? |



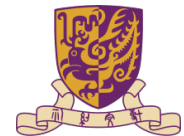
| Rank | bmp | format | music | intel | excel | font | watch | memory |
|------|----------|-----------|------------|-------------|-----------|---------------|------------|---------------|
| 1 | bmp | format | music | pentium | excel | font | watch | memory |
| 2 | jpg | format* | file | 4 | korean | korean | time | virtual |
| 3 | gif | xp | tag | celeron | function | 97 | background | shortage |
| 4 | save | windows | sound | amd | novice | add | start | ram |
| 5 | file | hard | background | intel | cell | download | date | message |
| 6 | picture | 98 | song | performance | disappear | control-panel | display | configuration |
| 7 | change | partition | play | support | convert | register | tray | 256 |
| 8 | ms-paint | drive | mp3 | question | if | install | power | extend |
| 9 | convert | disk | cd | buy | xls | default | screen | system |
| 10 | photo | C | source | cpu | record | photoshop | wrong | windows |

Figure: The **first** row shows the source words and each column shows top 10 words that are most semantically similar to source word. A higher rank means a larger $T(w | t)$ value



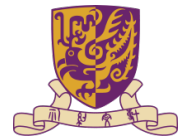
Lexical-based Approach: Translation Model

- How to learn $T(w | t)$?
 - Prepare a **monolingual parallel corpus** of pairs of text, each pair should be **semantically similar**
 - Employ machine translation model **IBM model I** on the parallel corpus to learn $T(w | t)$
 - **IBM model I**: Brown et al., Computational Linguistics, 1990
- How this paper prepares monolingual parallel corpus
 - Each pair contains **two questions** whose **answers** are very similar



Lexical-based Approach: Translation Model

- Delphine Bernhard and Iryna Gurevych, Combining Lexical Semantic Resources with Question & Answer Archives for Translation-Based Answer Finding, ACL, 2009
- Propose several methods to prepare parallel monolingual corpora
 - Question answer pairs: question – answer
 - Question reformulation pairs: question -- question reformulation by user



Lexical-based Approach: Translation Model

RUClimate (supervisor) [332] merged the question **Why iare clouds white** into **Why are clouds white** 9 Feb 2012 17:03

RUClimate (supervisor) [332] merged the question **What makes the clouds appeared to be white** into **Why are clouds white** 9 Feb 2012 16:44

RUClimate (supervisor) [332] merged the question **Why does Clouds appear white** into **Why are clouds white** 9 Feb 2012 16:44

RUClimate (supervisor) [332] merged the question **Why do clouds appear white** into **Why are clouds white** 9 Feb 2012 16:43

RUClimate (supervisor) [332] merged the question **Why do clouds look white** into **Why are clouds white** 9 Feb 2012 16:43

RUClimate (supervisor) [332] merged the question **Why do clouds in the sky appear white** into **Why are clouds white** 9 Feb 2012 16:43

RUClimate (supervisor) [332] merged the question **How does the cloud is white** into **Why are clouds white** 9 Feb 2012 16:43

RUClimate (supervisor) [332] merged the question **Why is it that the cloud is white** into **Why are clouds white** 9 Feb 2012 16:43

RUClimate (supervisor) [332] merged the question **Why we have a white clouds white clouds** into **Why are clouds white** 9 Feb 2012 16:42

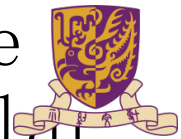
RUClimate (supervisor) [332] merged the question **Why is it the cloud is white** into **Why are clouds white** 9 Feb 2012 16:42

RUClimate (supervisor) [332] merged the question **Why do you have a white cloud** into **Why are clouds white** 9 Feb 2012 16:42



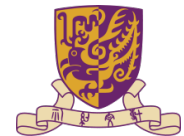
Lexical-based Approach: Translation Model

- Lexical Semantic Resources: **glosses** and **definitions** for the same lexeme in **different lexical semantic and encyclopedic resources** can be considered as **near-paraphrases**, since they **define** the same terms and hence have the same meaning
 - **Moon**
 - **Wordnet**: the natural satellite of the Earth
 - **English Wiktionary**: the Moon, the satellite of planet Earth
- **English Wikipedia**: the Moon (Latin: Luna) is Earth's only natural satellite and the **fifth largest natural satellite in the Solar**



Lexical-based Approach: Translation-based Language Model

- Translation Model
 - Advantage: Tackle lexical gap to some extent
 - Disadvantage: $T(w | w) = 1$ for all w while maintaining other word translation probabilities unchanged, produce inconsistent probability estimates and make the model unstable
- Xiaobing Xue, Jiwoon Jeon and W. Bruce Croft, Retrieval Models for Question and Answer Archives, SiGIR, 2008
- Translation-based Language Model



Lexical-based Approach: Translation-based Language

Model

Translation Model

$$P(w|D) = (1 - \lambda) \sum_{t \in D} (T(w|t) P_{ml}(t|D)) + \lambda P_{ml}(w|C)$$

Translation-based Language Model

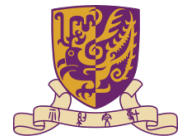
$$P(w|D) = \frac{|D|}{|D| + \lambda} P_{mx}(w|D) + \frac{\lambda}{|D| + \lambda} P_{ml}(w|C)$$

$$P_{mx}(w|D) = (1 - \beta) P_{ml}(w|D) + \beta \sum_{t \in D} T(w|t) P_{ml}(t|D)$$

- Linear combination of language model and translation model
- Answer part should provide additional evidence about relevance, incorporating the answer part

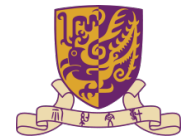
$$P_{mx}(w|(D, A)) = \alpha P_{ml}(w|D) + \beta \sum_{t \in D} T(w|t) P_{ml}(t|D) + \gamma P_{ml}(w|A)$$

$$\alpha + \beta + \gamma = 1$$



Syntactic-based Approach: Syntactic Tree Matching

- Some similar questions neither share many common words, nor follow identical syntactic structure
 - **How can I lose weight in a few months?**
 - Are there any ways of losing pound in a short period?
- Kai Wang, Zhaoyan Ming and Tat-Seng Chua, A Syntactic Tree Matching Approach to Finding Similar Questions in Community-based QA Services, SIGIR, 2009
- Syntactic tree matching



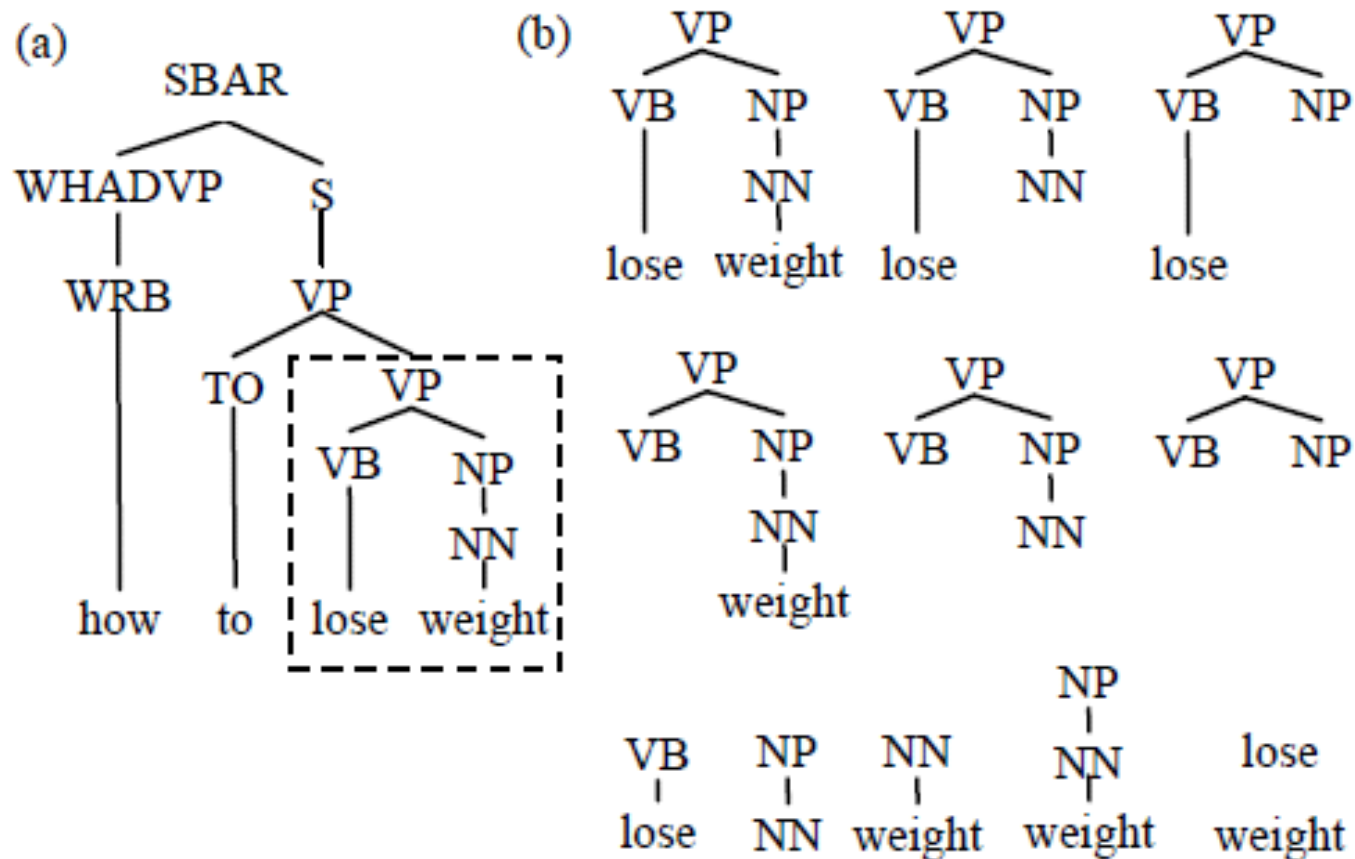


Figure: (a) The Syntactic Tree of the Question "How to lose weight?". (b)

Tree Fragments of the Sub-tree covering "lose weight".



Syntactic-based Approach: Syntactic Tree Matching

- Tree kernel: utilize structural or syntactic information to capture higher order dependencies between grammar rules

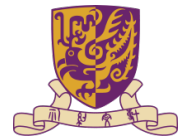
$$k(T_1, T_2) = \sum_{n_1 \in N_1} \sum_{n_2 \in N_2} C(n_1, n_2)$$

- N_1, N_2 are sets of nodes in two syntactic trees T_1 and T_2 , and $C(n_1; n_2)$ equals to the number of common fragments rooted in n_1 and n_2

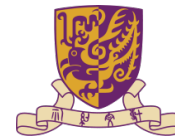


Syntactic-based Approach: Syntactic Tree Matching

- Limitation of tree kernel
 - Tree kernel function merely relies on the intuition of **counting the common number of sub-trees**, whereas the number **might not be a good indicator** of the similarity between two questions
 - Two evaluated sub-trees have to be identical to allow further parent matching, for which **semantic representations cannot fit in well**
- Syntactic tree matching
 - A new weighting scheme for tree fragments that are **robust against** some grammatical errors
 - Incorporate semantic features



QUESTION RECOMMENDATION



Motivation

- Question Recommendation
 - Retrieve and rank other questions according to their likelihood of being **good recommendations** of the **queried question**
 - A good recommendation provides **alternative aspects around users' interest**



Example

Queried question:

Any cool clubs in Berlin or Hamburg?

Question search:

What are the best/most fun clubs in Berlin?

Question recommendation

How far is it from Berlin to Hamburg?

Where to see between Hamburg and Berlin?

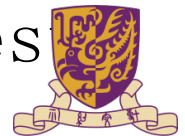
Hong long does it take to get to Hamburg from Berlin on the train?

Cheap hotel in Hamburg?



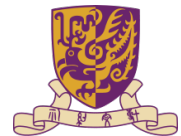
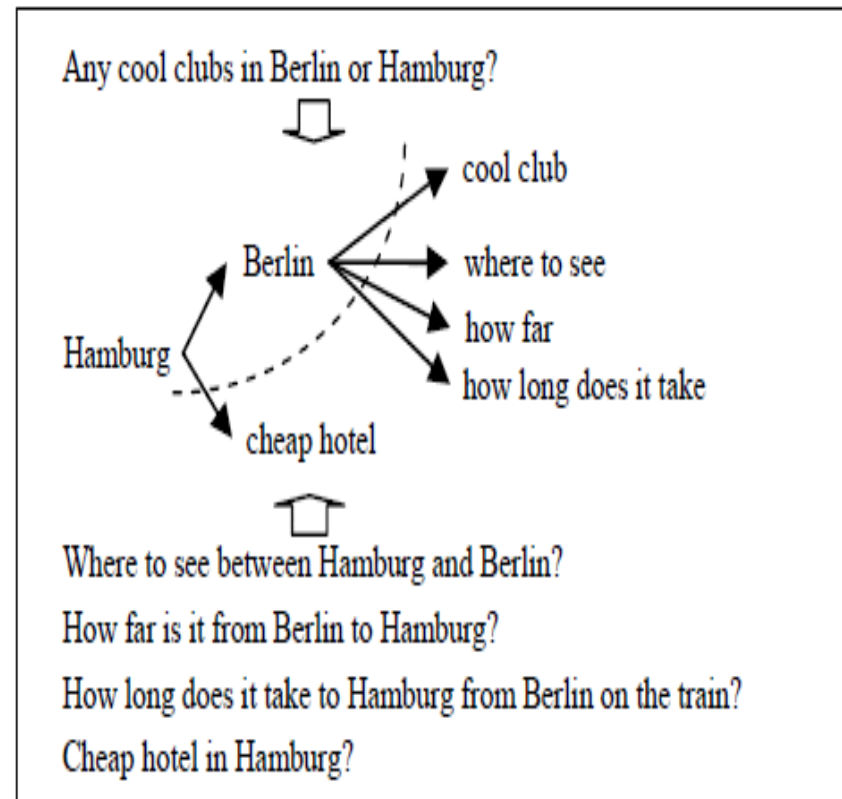
Question Recommendation: MDL-based Tree Cut Model

- Yunbo Cao, Huizhong Duan, Chin-Yew Lin, Yong Yu and Hsiao-Wuen Hon, Recommending Questions Using the MDL-based Tree Cut Model, WWW, 2008
 - Step 1: Represent questions as graphs of topic terms
 - Step 2: Rank recommendations on the basis of the graphs
- Formalize both steps as the tree-cutting problems and employ the MDL (Minimum Description Length) for selecting the best cuts



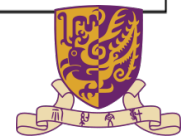
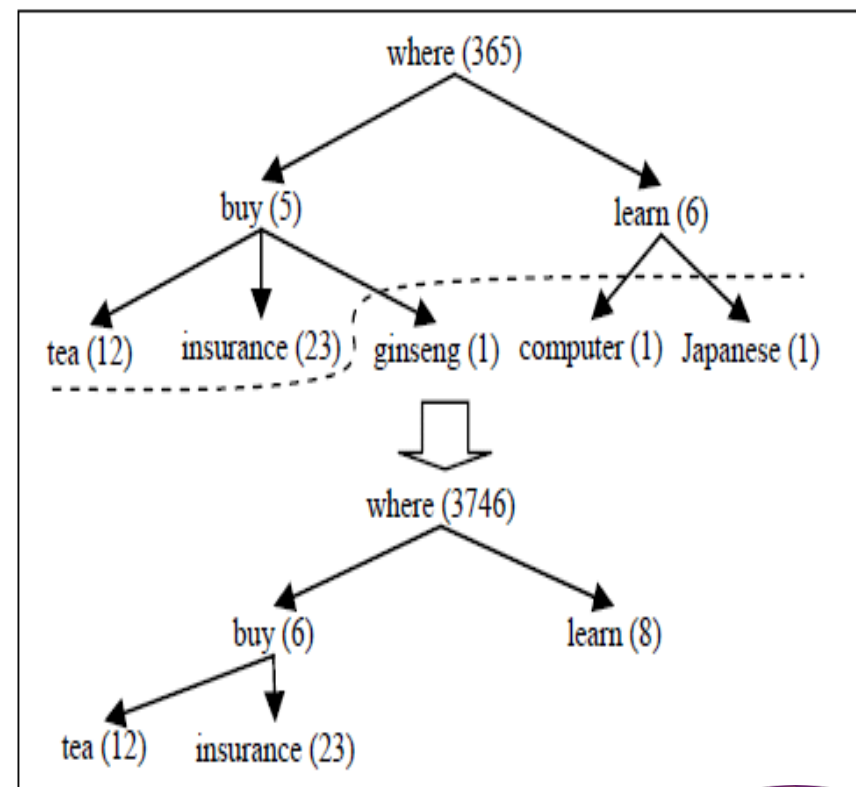
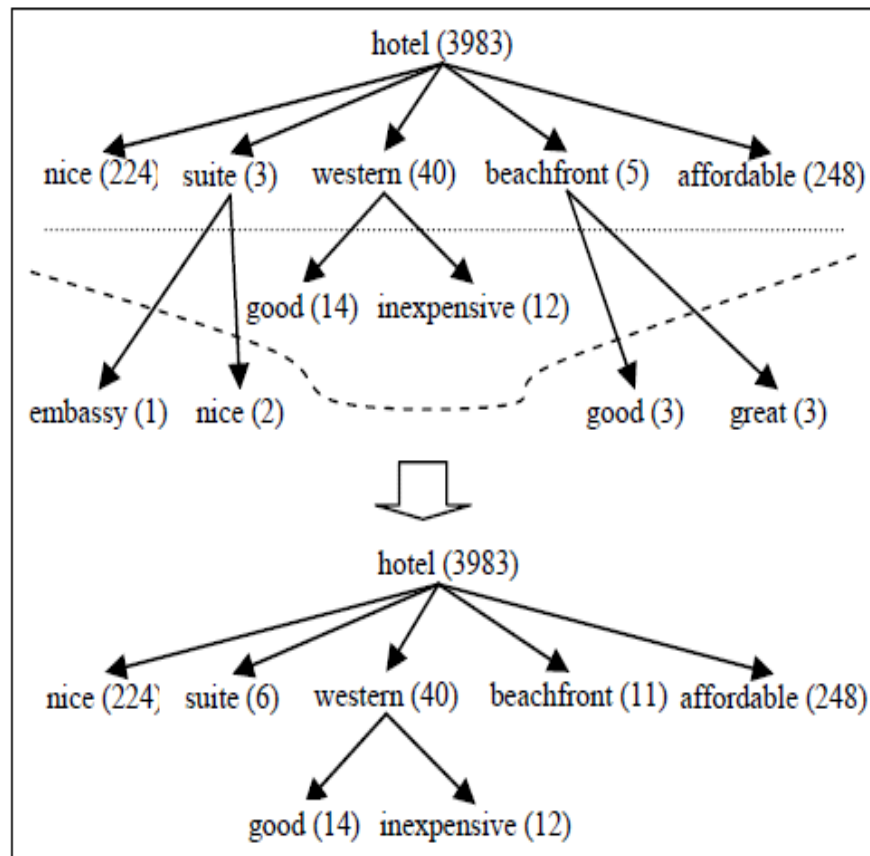
Question Recommendation: MDL-based Tree Cut Model

- Question
 - Any cool clubs in Berlin or Hamburg?
- Question topic
 - Major **context/constraint** of a question, characterize users' interests
 - Berlin, Hamburg
- Question focus
 - Certain **aspect** of the **question topic**
 - cool club
- Suggest alternative **aspects** of the queries question topic



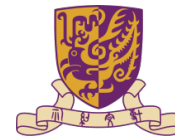
Question Recommendation: MDL-based Tree Cut Model

- Extraction of topic terms: **base noun phrase, WH-ngram**
- Reduction of topic terms: **MDL-based tree cut model**



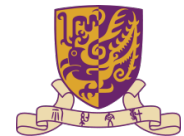
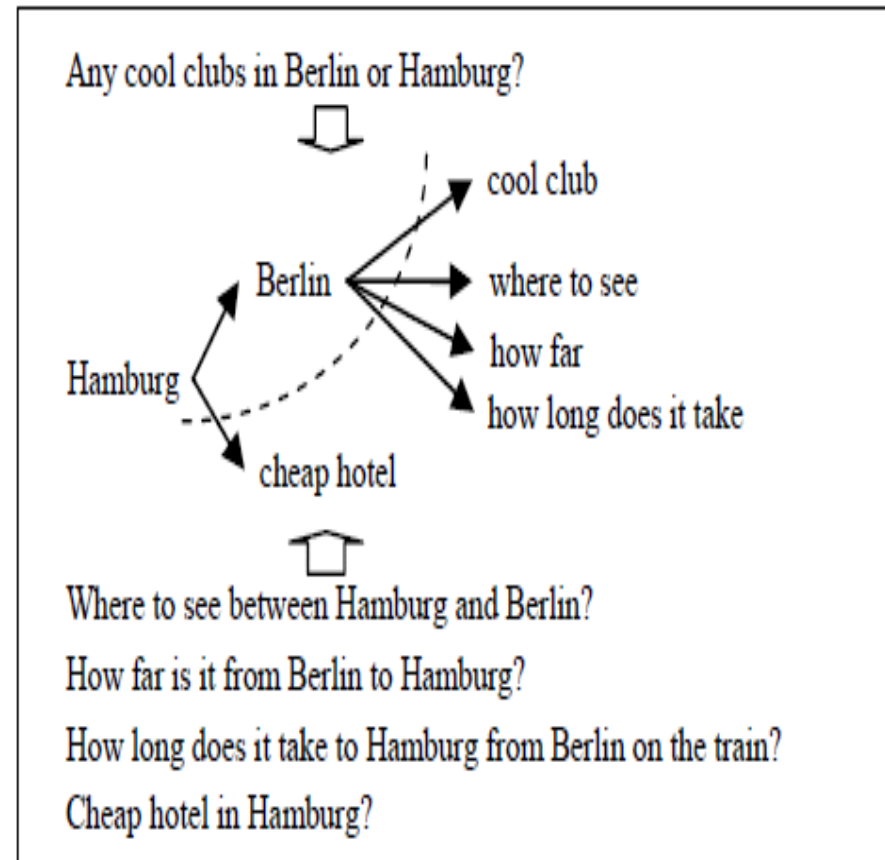
Question Recommendation: MDL-based Tree Cut Model

- Topic profile
 - Probability distribution of categories $\{p(c | t)\}_{c \in C}$
 - $$p(c|t) = \frac{\text{count}(c,t)}{\sum_{c \in C} \text{count}(c,t)}$$
 - $\text{count}(c,t)$ is the frequency of the topic term t within the category c
- Specificity
 - Inverse of the entropy of the topic profile
 - Topic term of **high specificity** usually specifies **question topic**
 - Topic term of low specificity is usually used to represent **question focus**
- Topic chain
 - Topic chain is a sequence of ordered topic terms sorted from big to small according to specificity
- Question tree
 - Prefix tree built over topic chains of the question set Q



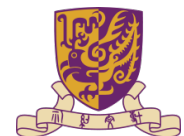
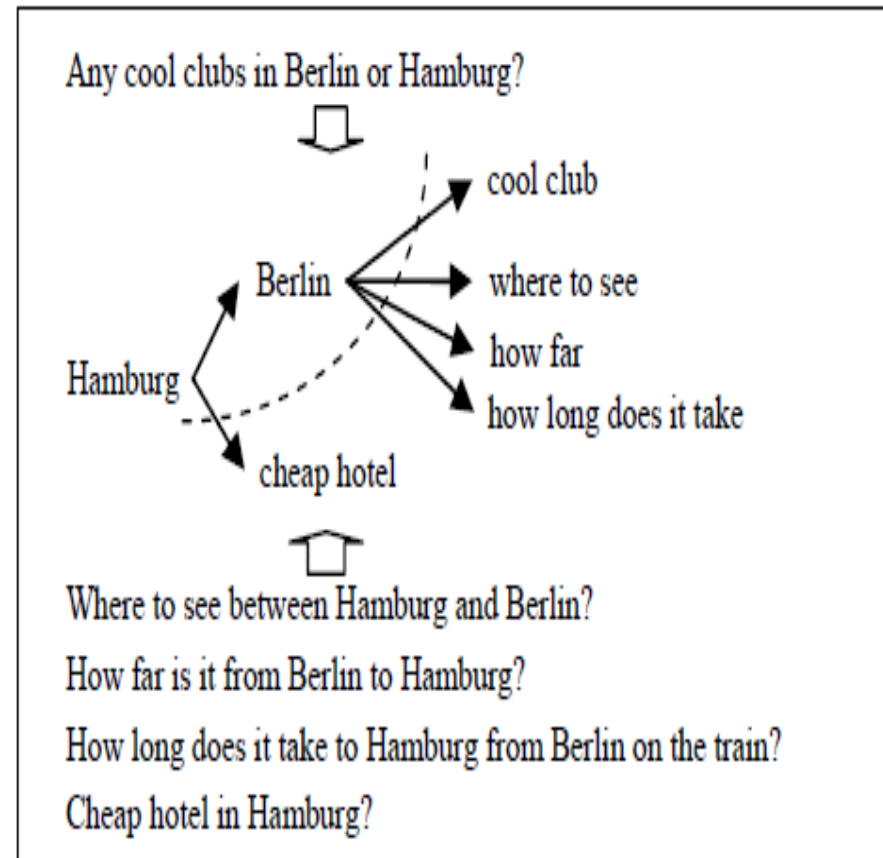
Question Recommendation: MDL-based Tree Cut Model

- Ranking recommendation candidates
 - Determine what **topic terms (question focus)** should be substituted
 - Collect a set of **topic chain** $Q^c = \{q_i^c\}_{i=1}^N$ such that at least one topic term occurs in both q^c and q_i^c
 - Construct a **question tree** from the set of topic chain $Q^c \cup q^c$
 - Employ MDL to separate topic chains into Head, H and Tail, T



Question Recommendation: MDL-based Tree Cut Model

- Ranking recommendation candidates
 - Score recommendation candidates rendered by various substitutions
 - Specificity: the more similar are $H(q^c)$ and $H(\hat{q}^c)$, the **higher** score
 - Generality: the more similar are $T(q^c)$ and $T(\hat{q}^c)$, the **lower** score



Question Recommendation: TopicTRLM

- Tom Chao Zhou, Chin-Yew Lin, Irwin King, Michael R. Lyu, Young-In Song and Yunbo Cao, Learning to Suggest Questions in Online Forums, AAAI, 2011
- Suggest semantically related questions in online forums
 - How is Orange Beach in Alabama?
 - Is the water pretty clear this time of year on Orange Beach?
 - Do they have chair and umbrella rentals on Orange Beach?
 - Topic: **travel in Orange Beach**
- Fuse both **lexical** and **latent semantic information**

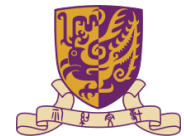
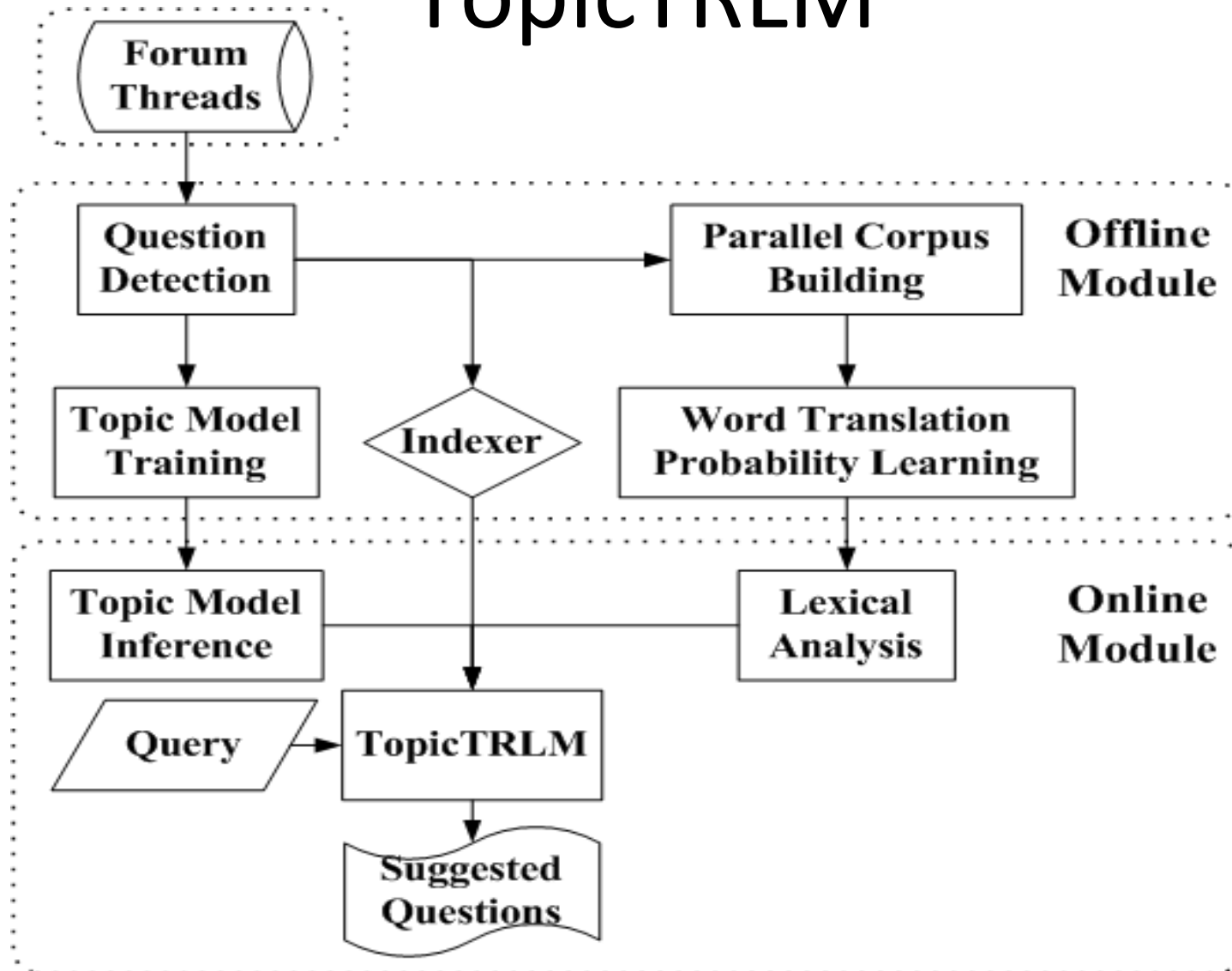


Question Recommendation: TopicTRLM

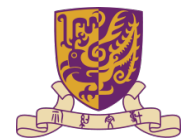
- Document representation
 - Bag-of-words
 - Independent
 - Fine-grained representation Lexically similar
 - Topic model
 - Assign a set of latent topic distributions to each word
 - Capturing important relationships between words
 - Coarse-grained representation
 - Semantically related



Question Recommendation: TopicTRLM

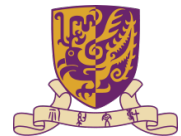


QUESTION SUBJECTIVITY ANALYSIS



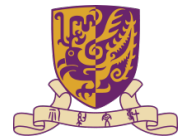
Question Subjectivity Analysis

- **Question Analysis** is to analyze characteristics of questions
- Understand **User Intent**
- Provide **rich information** to question search, question recommendation, answer quality prediction, etc.
- **Question Subjectivity Analysis** is an important aspect of **question analysis**



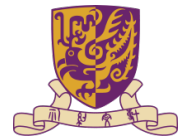
Definition

- Subjective question
 - Private statements
 - Personal opinion and experience
 - E.g. *What's the difference between chemotherapy and radiation treatments?*
- Objective question
 - Objective, verifiable information
 - Often with support from reliable sources
 - E.g. *Has anyone got one of those home blood pressure monitors? And if so what make is it and do you think they are worth getting?*



Motivation

- More accurately identify **similar questions**, improve **question search**
- Better **rank or filter** the answers based on whether an answer matches the question orientation
- Crucial component of inferring **user intent**, a long-standing problem in Web search
- **Route** subjective questions to **users** for answer, **trigger automatic factual question answering** system for objective questions



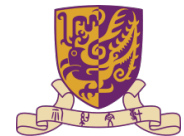
Challenge

- Ill-formatted, e.g., word capitalization may be incorrect or missing, consecutive words may be concatenated
- Ungrammatical, include common online idioms, e.g., using “u” means “you”, “2” means “to”

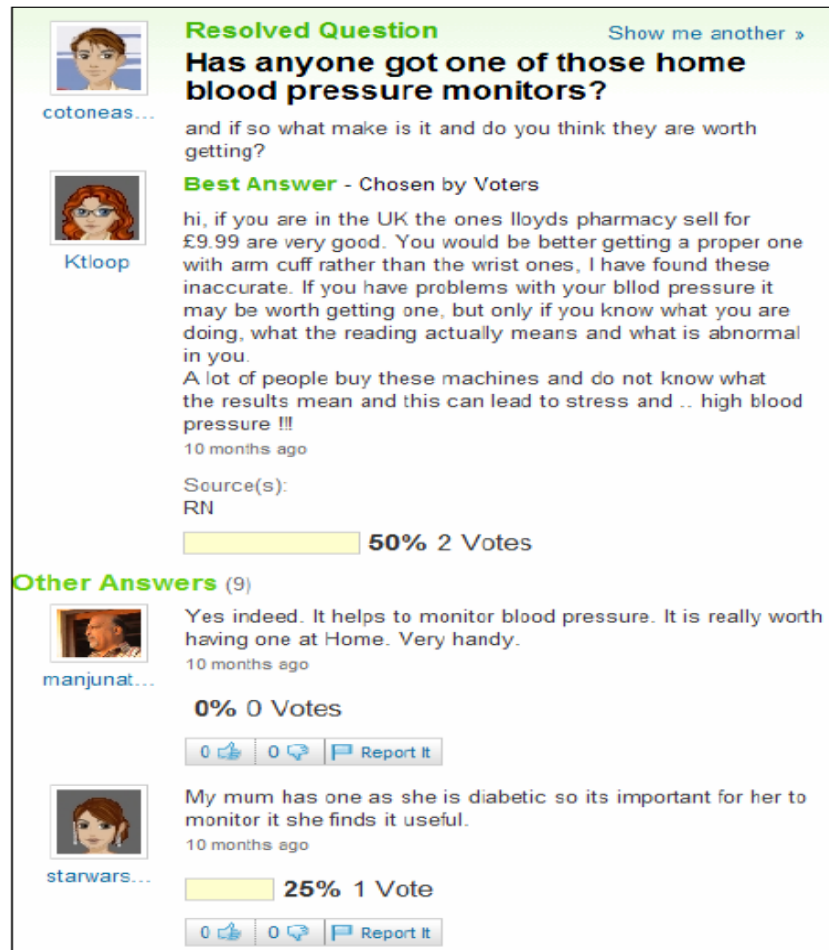


Question Subjectivity Analysis: Supervised Learning

- Baoli Li, Yandong Liu, Ashwin Ram, Ernest V. Garcia and Eugene Agichtein, Exploring Question Subjectivity Prediction in Community QA, SIGIR, 2008
- Support Vector Machine with linear kernel
- Features
 - Character 3-gram
 - Word
 - Word + character 3-gram
 - Word n-gram
 - Word POS n-gram, mix of word and POS tri-grams
- Term weighting schemes: binary, TF, TF*IDF



Question Subjectivity Analysis: Semi-Supervised Learning



Resolved Question [Show me another »](#)
Has anyone got one of those home blood pressure monitors?
cotoneas... and if so what make is it and do you think they are worth getting?

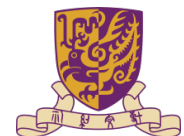
Best Answer - Chosen by Voters
Ktloop
hi, if you are in the UK the ones lloyds pharmacy sell for £9.99 are very good. You would be better getting a proper one with arm cuff rather than the wrist ones, I have found these inaccurate. If you have problems with your blod pressure it may be worth getting one, but only if you know what you are doing, what the reading actually means and what is abnormal in you.
A lot of people buy these machines and do not know what the results mean and this can lead to stress and .. high blood pressure !!!
10 months ago
Source(s): RN
50% 2 Votes

Other Answers (9)
manjunat...
Yes indeed. It helps to monitor blood pressure. It is really worth having one at Home. Very handy.
10 months ago
0% 0 Votes
0 thumbs up 0 thumbs down Report it

starwars...
My mum has one as she is diabetic so its important for her to monitor it she finds it useful.
10 months ago
25% 1 Vote
0 thumbs up 0 thumbs down Report it

- Baoli Li, Yandong Liu and Eugene Agichtein, CoCQA: Co-Training Over Questions and Answers with an Application to Predicting Question Subjectivity Orientation, EMNLP, 2008
- Incorporate relationships between **questions** and corresponding **answers**
- Co-training, two views of the data, **question** and **answer**

Figure: Yahoo Answers Example



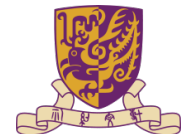
Input:

- F_Q and F_A are *Question* and *Answer* feature views
- C_Q and C_A are classifiers trained on F_Q and F_A respectively
- L is a set of labeled training examples
- U is a set of unlabeled examples
- K : Number of unlabeled examples to choose on each iteration
- X : the threshold for increment
- R : the maximal number of iterations

Algorithm CoCQA

1. Train $C_{Q,0}$ on $L: F_Q$, and record resulting $ACC_{Q,0}$
2. Train $C_{A,0}$ on $L: F_A$, and record resulting $ACC_{A,0}$
3. **for** $j=1$ to R **do**:
 - Use $C_{Q,j-1}$ to predict labels for U and choose top K items with highest confidence $\rightarrow E_{Q,j-1}$
 - Use $C_{A,j-1}$ to predict labels for U and choose top K items with highest confidence $\rightarrow E_{A,j-1}$
 - Move examples $E_{Q,j-1} \cup E_{A,j-1} \rightarrow L$
 - Train $C_{Q,j}$ on $L: F_Q$ and record training $ACC_{Q,j}$
 - Train $C_{A,j}$ on $L: F_A$ and record training $ACC_{A,j}$
 - if** $Max(\Delta ACC_{Q,j}, \Delta ACC_{A,j}) < X$ **break**
4. **return** final classifiers $C_{Q,j} \rightarrow C_Q$ and $C_{A,j} \rightarrow C_A$

- At step 1,2, each category has top K_j most confident examples chosen as additional “labeled” data
- Terminate when the increments of both classifiers are less than threshold X or maximum number of iterations are exceeded



Question Subjectivity Analysis: Data-driven Approach

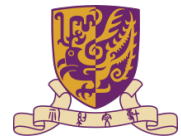
- Tom Chao Zhou, Xiance Si, Edward Y. Chang, Irwin King and Michael R. Lyu, A Data-Driven Approach to Question Subjectivity Identification in Community Question Answering, AACL, 2012
- Li et al. 2008 (supervised), Li et al. 2008 (CoCQA, semi-supervised) based on manual labeling data
- **Manual labeling** data is quite **expensive**



Question Subjectivity Analysis: Data-driven Approach

Web-scale learning is to use available large-scale data rather than hoping for annotated data that isn't available

- Halevy, Norvig and Pereira



Question Subjectivity Analysis: Data-driven Approach

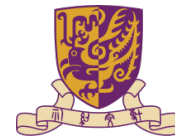
Whether we can utilize **social signals** to collect **training data** for question subjectivity identification with **NO** manual labeling?



Like Signal



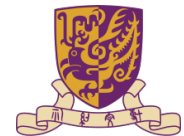
- Like an answer if they find the answer useful
- Intuition
 - **Subjective**: answers are **opinions, different tastes**; best answer receives **similar number of likes** with other answers
 - **Objective**: like an answer which explains **universal truth** in most detail; best answer receives **high likes** than other answers



Vote Signal



- Users could vote for **best answer**
- Intuition
 - **Subjective**: vote for different answers, support different opinions; **low percentage** of votes on best answer
 - **Objective**: easy to identify answer contain the most fact; percentage of votes of best answer is **high**



Source Signal


Who invented the computer mouse?
does anyone know who invented the first Computer mouse and when was it invented?
3 years ago [Report Abuse](#)

Best Answer - Chosen by Asker
A guy called Engelbart - here it is
<http://sloan.stanford.edu/MouseSite/Arch...>
...mmmmm, sweet!!
Source(s):
<http://inventors.about.com/library/weekl...>

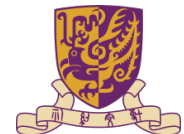
↓

Inventors of the Modern Computer
The History of the Computer Mouse and the Prototype for Windows - Douglas Engelbart
By [Mary Bellis](#)

"It would be wonderful if I can inspire others, who are struggling to realize their dreams, to say 'if this country kid could do it, let me keep slogging away'." - Douglas Engelbart



- Reference to authoritative resources
- Intuition
 - Only available for objective question that has fact answer



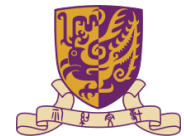
Poll and Survey Signal

- User intent is to seek **opinions**
- Very likely to be **subjective**
- What is something you learned in school that you think is useful to you today?
- If you could be a cartoon character, who would you want to be?



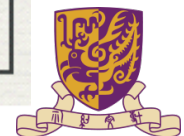
Answer Number Signal

- The **number of posted** answers to each question
- Intuition
 - Subjective: alert post opinions even they notice there are **other answers**
 - Objective: **may not post** answers to questions that has received other answers since an **expected** answer is usually fixed
 - A **large answer number** indicate **subjectivity**
 - A **small** answer number may be due to many reasons, such as **objectivity**, small **page views**




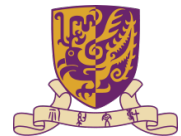
Question Subjectivity Analysis: Data-driven Approach

| Summary of Social Signals | | |
|---------------------------|---|----------------------|
| Name | Description | Training Data |
| Like | Capture users' tastes | Positive && Negative |
| Vote | Reflect users' judgments | Positive && Negative |
| Source | Measure confidence on authoritativeness | Negative |
| Poll and Survey | Indicate users' intent | Positive |
| Answer Number | Imply users' willingness to answer a question | Positive |



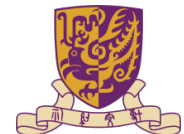
Question Subjectivity Analysis: Data-driven Approach

- Features
 - Word: term frequency
 - Word n-gram: term frequency
 - Word: term frequency
 - Question length: information needs of subjective questions are **complex**, users use **descriptions** to explain, **larger question length**  Request word: particular words to explicitly indicate their request for seeking **opinions**;
manual list of 9 words

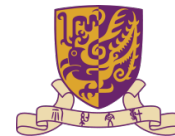


Question Subjectivity Analysis: Data-driven Approach

- Subjectivity clue: **external** lexicon, over 8000 clues, manually compiled word list from news to express opinions
- Punctuation density: density of **punctuation marks**
- Grammatical modifier: inspired by opinion mining research of using **grammatical modifiers** on judging users' opinions, **adjective** and **adverb**
- Entity: objective question expects fact answer, leading to **less relationships** among entities, subjective questions contains more descriptions, may involve relatively **complex relations**

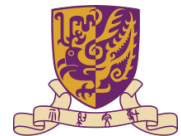


CONTENT QUALITY EVALUATION



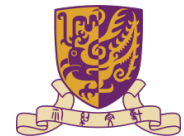
Content Quality Evaluation

- Motivation
 - High variance in the quality of answers & questions
 - Automatically find the best answer & spam
 - Significant impact on user satisfaction



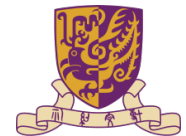
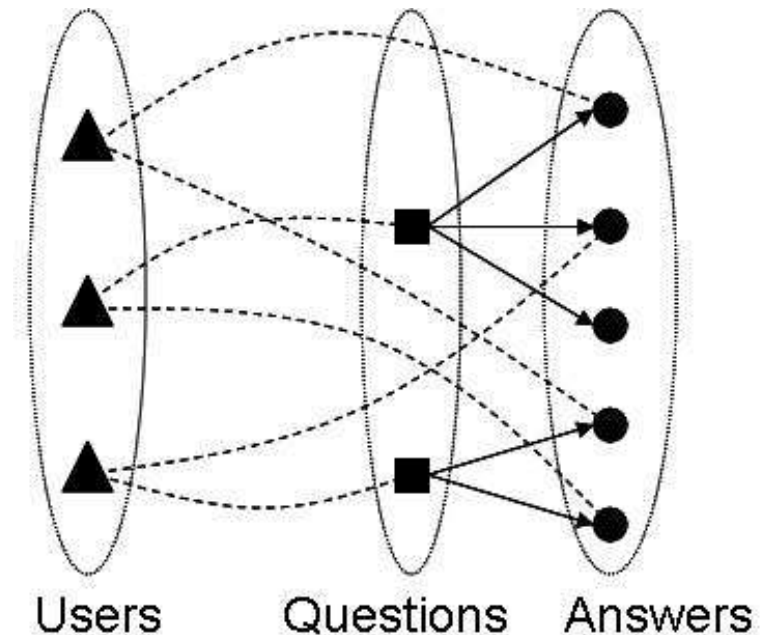
Approaches

- Maximum Entropy (Jeon et al. 2006)
- Learning to Rank (Surdeanu et al. 2008)
- Analogical Reasoning (Wang et al., 2009)
- Graph-based Models
 - Coupled Mutual Reinforcement (Bian et al., 2009)
 - EXHITS (Suryanto et al., 2009)
- Logistic Regression (Shah et al. 2010)



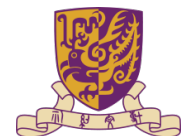
Recognizing Reliable Users and Content with Coupled Mutual Reinforcement

- Given a CQA archive
- Determine the quality of each **question** and **answer** and the **answer-reputation** and **question-reputation** of each user simultaneously

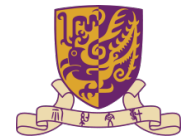
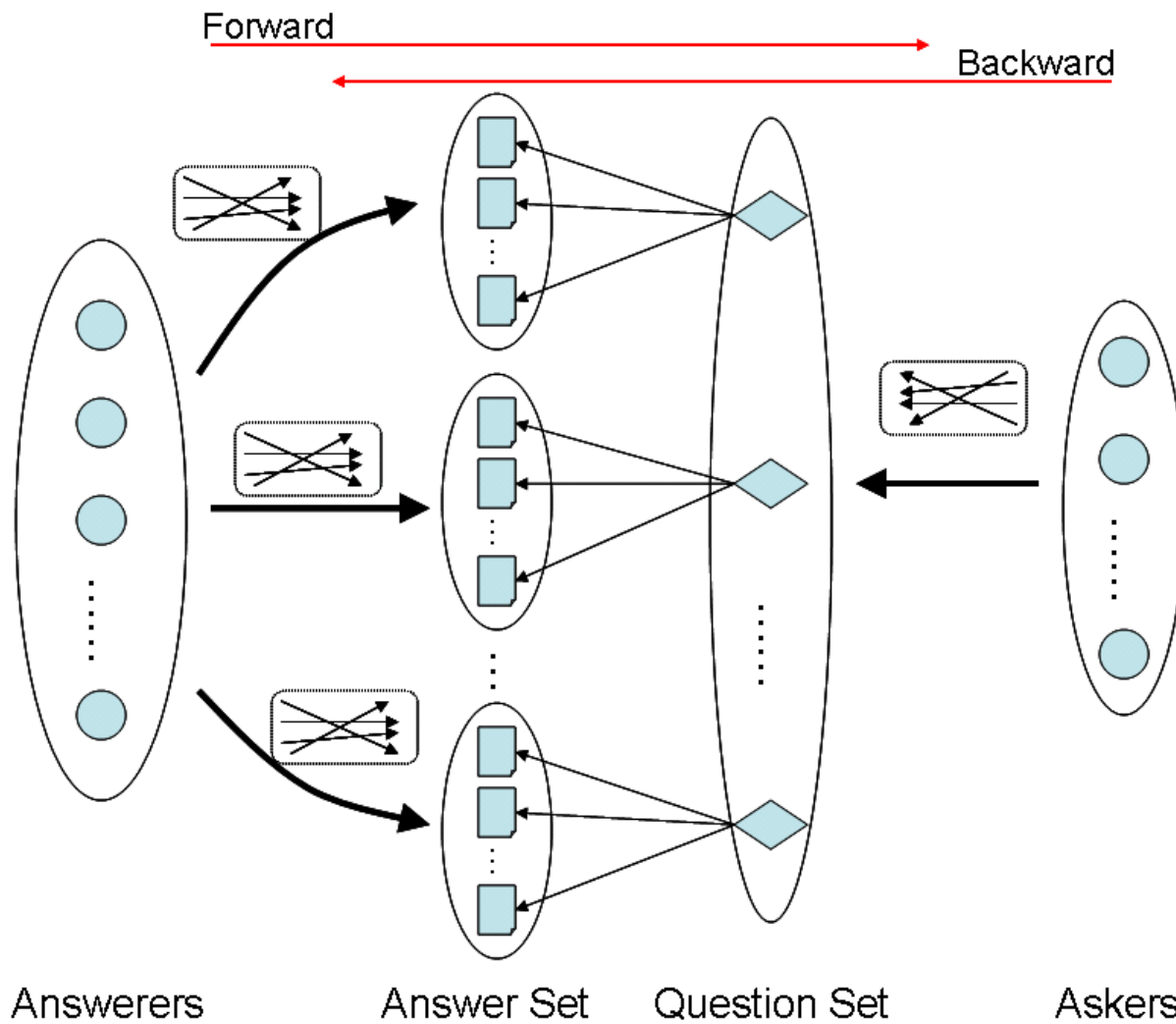


Content Quality & User Reputation

- Question Quality
 - A question's effectiveness at attracting high quality answers
- Answer Quality
 - The responsiveness, accuracy, and comprehensiveness of the answer to a question.
- Question Reputation
 - The expected quality of the questions posted by a user
- Answer Reputation
 - The expected quality of the answers posted by a user



CQA-MR Model



Mutual Reinforcement Principle

u's answer reputation

$$y_u^a \propto \sum_{u \sim a} m_{ua} y_a$$

the quality of answer a's question

a's quality

$$y_a \propto \alpha \sum_{a \sim u} m_{ua} y_u^a + (1 - \alpha) y_{q(\sim a)}$$

u's question reputation

$$y_u^q \propto \sum_{u \sim q} m_{uq} y_q$$

the question reputation of the user who ask question q

q's quality

$$y_q \propto \gamma \sum_{q \sim a} m_{aq} y_a + (1 - \gamma) y_{u(\sim q)}^q$$



Feature Space

| Question Feature Space $X(Q)$ | |
|-------------------------------|---|
| Q: subject length | Number of words of question subject |
| Q: detail length | Number of words of question detail |
| Q: posting time | Date and time when the question was posted |
| Q: question stars | Number of stars received earned for this question |
| Q: number of answers | Number of answers received for this question |
| Answer Feature Space $X(A)$ | |
| A: overlap | Words shared between question and answer |
| A: number of comments | Number of comments added by other participants |
| A: total thumbs up | Total number of thumb up votes for the answers |
| A: total thumbs down | Total number of negative votes for the answers |
| User Feature SPace $X(U)$ | |
| U: total points | Total points earned over lifetime community |
| U: questions asked | Number of questions asked |
| U: questions resolved | Number of questions resolved |
| U: total answers | Number of posted answers |
| U: best answer | Number of answers that were selected as “best answer” |
| U: stars | Number of stars the user receive |
| U: thumbs up ratio | The ratio of thumbs up votes the user posted before |
| U: thumbs down ratio | The ratio of thumbs down votes the user posted before |
| U: indegree | number of other users whose questions are answered by the user |
| U: outdegree | number of other users who answer the questions posted by the user |
| U: hub score | the hub score of the user computed by HITS |
| U: authority score | the authority score of the user computed by HITS |



Logistic Regression Model

- $P(\mathbf{x})$: probability of being “good” (\mathbf{x} can be a question, answer or user feature vector)

$$\log \frac{P(\mathbf{x})}{1 - P(\mathbf{x})} = \beta^T \mathbf{x}$$

$$LL(\mathcal{X}) = \sum_{\mathbf{x} \in \mathcal{X}} y \beta^T \mathbf{x} - \log(1 + e^{\beta^T \mathbf{x}})$$

$$LL(\mathbf{y}|Y_d) = - \sum_{i=1}^{|\mathcal{X}|} y(i) \log \frac{y(i)}{y'(i)} - (1 - y(i)) \log \frac{1 - y(i)}{1 - y'(i)}$$

- Object function

$$L(\mathcal{X}) = LL(\mathcal{X}) + \sigma LL(\mathbf{y}|Y_d)$$



Algorithm

input : questions, answers and users and their connection from CQA-network.

output: answer quality y_a ;
 answer-reputation of user y_u^a ;
 question quality y_q ;
 question-reputation of user y_u^q

Start with an initial guess, e.g. uniform values, for y_a , y_u^a , y_q and y_u^q ;

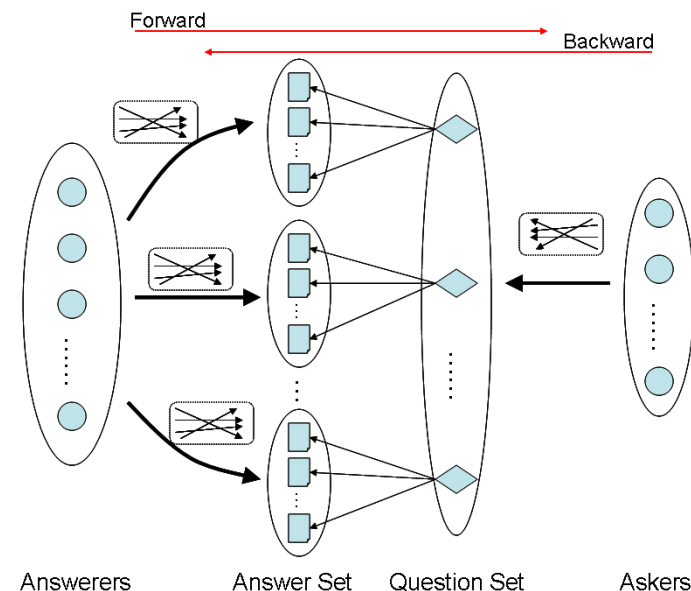
begin

while y_a, y_u^a, y_q, y_u^q not converge **do**

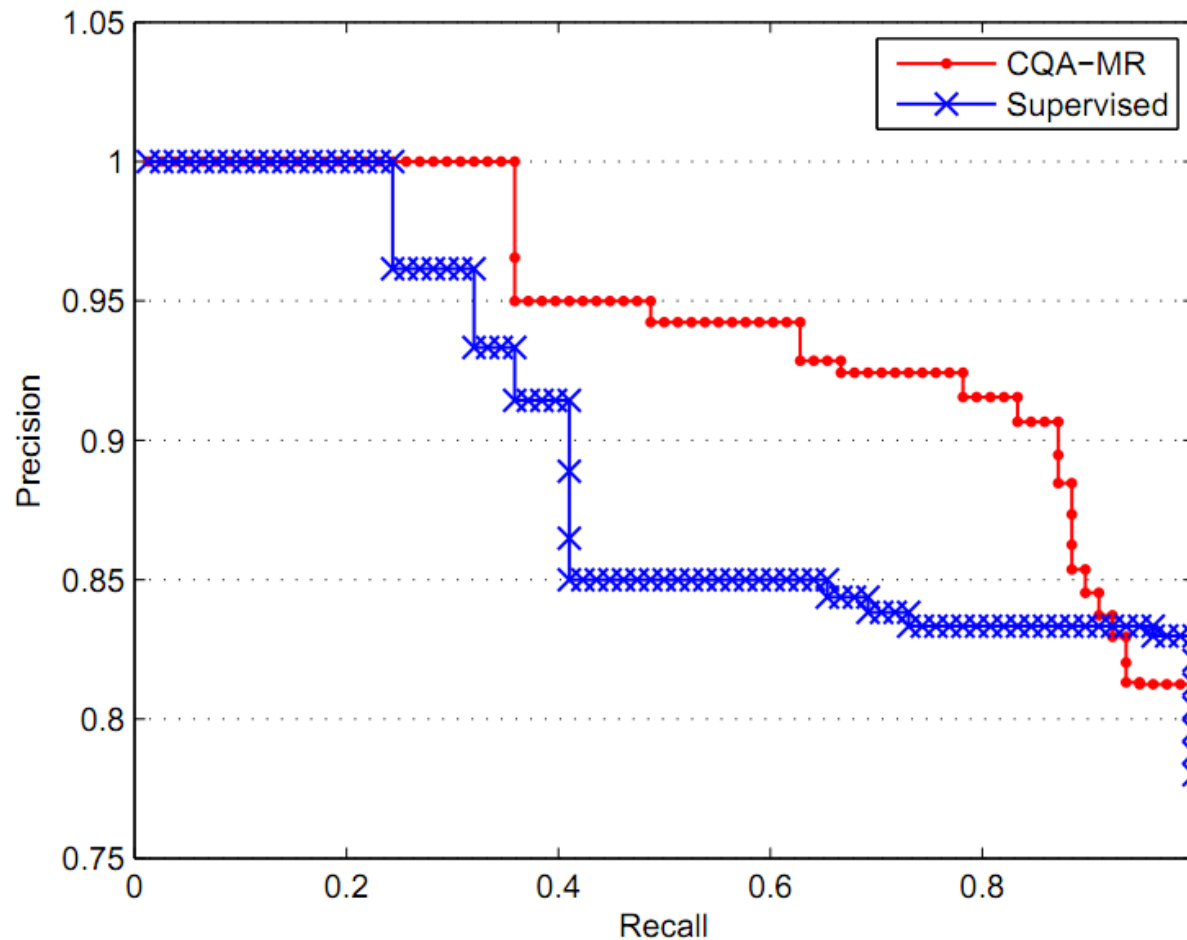
Forward fit the logistic regression models and calculate new values for y_a, y_q and y_u^q in sequence ;

Backward fit the logistic regression models and calculate new values for y_a, y_q and y_u^a in sequence

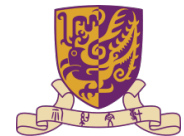
end



Experimental Result

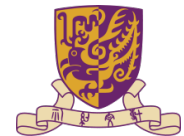


Precision-Recall curves for predicting question quality of CQA-MR and Supervised method



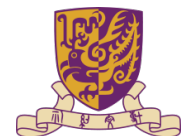
References

- B. Dom and D. Paranjpe. A Bayesian Technique for Estimating the Credibility of Question Answerers. *Proceedings of SIAM Conference on Data Mining (SDM'08)*, pages 399--409, 2008.
- Chirag Shah and Jefferey Pomerantz. 2010. Evaluating and predicting answer quality in community QA. In *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '10)*. ACM, New York, NY, USA, 411-418.
- Dredze, M.; Crammer, K.; and Pereira, F. 2008. Confidence-Weighted Linear Classification. In *Proceedings of the Twenty-Fifth International Conference on Machine Learning (ICML)*. Princeton, NJ: International Machine Learning Society.
- Ferrucci, D., and Lally, A. 2004. UIMA: An Architectural Approach to Unstructured Information Processing in the Corporate Research Environment. *Natural Language Engineering* 10(3-4): 327-348.
- GUO, J., XU, S., BAO, S., AND YU, Y. 2008. Tapping on the potential of q&a community by recommending answer providers. In *Proceeding of the 17th ACM conference on Information and knowledge management. CIKM '08*. ACM, New York, NY, USA, 921-930.
- Jiang Bian, Yandong Liu, Ding Zhou, Eugene Agichtein, and Hongyuan Zha. 2009. Learning to recognize reliable users and content in social media with coupled mutual reinforcement. In *Proceedings of the 18th international conference on World wide web (WWW '09)*. ACM, New York, NY, USA, 51-60.
- Jiwoon Jeon, W. Bruce Croft, Joon Ho Lee, and Soyeon Park. 2006. A framework to predict the quality of answers with non-textual features. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval (SIGIR '06)*. ACM, New York, NY, USA, 228-235.



References

- Jun Zhang, Mark S. Ackerman, and Lada Adamic. 2007. Expertise networks in online communities: structure and algorithms. In *Proceedings of the 16th international conference on World Wide Web (WWW '07)*. ACM, New York, NY, USA, 221-230.
- Lenat, D. B. 1995. Cyc: A Large-Scale Investment in Knowledge Infrastructure. *Communications of the ACM* 38(11): 33–38.
- Yunbo Cao, Huizhong Duan, Chin-Yew Lin, Yong Yu, and Hsiao-Wuen Hon . Recommending Questions Using the MDL-based Tree Cut Model, WWW, 2008.
- Kai Wang, Zhaoyan Ming, and Tat-Seng Chua. A Syntactic Tree Matching Approach to Finding Similar Questions in Community-based QA Services, SIGIR, 2009.
- Xin Cao, Gao Cong, Bin Cui, Christian Søndergaard Jensen, and Ce Zhang. The Use of Categorization Information in Language Models for Question Retrieval, CIKM, 2009.
- Tom Chao Zhou, Chin-Yew Lin, Irwin King, Michael R. Lyu, Young-In Song, and Yunbo Cao. Learning to Suggest Questions in Online Forums, AAI, 2011.
- LIU, M., LIU, Y., AND YANG, Q. 2010. Predicting best answerers for new questions in community question answering. In *Web-Age Information Management*, L. Chen, C. Tang, J. Yang, and Y. Gao, Eds. *Lecture Notes in Computer Science Series*, vol. 6184. Springer Berlin / Heidelberg, 127–138.
- Maggy Anastasia Suryanto, Ee Peng Lim, Aixin Sun, and Roger H. L. Chiang. 2009. Quality-aware collaborative question answering: methods and evaluation. In *Proceedings of the Second ACM International Conference on Web Search and Data Mining (WSDM '09)*, Ricardo Baeza-Yates, Paolo Boldi, Berthier Ribeiro-Neto, and B. Barla Cambazoglu (Eds.). ACM, New York, NY, USA, 142-151.



References

- Pawel Jurczyk and Eugene Agichtein. 2007. Discovering authorities in question answer communities by using link analysis. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management (CIKM '07)*. ACM, New York, NY, USA, 919-922.
- QU, M., QIU, G., HE, X., ZHANG, C., WU, H., BU, J., AND CHEN, C. 2009. Probabilistic question recommendation for question answering communities. In *Proceedings of the 18th international conference on World wide web. WWW '09*. ACM, New York, NY, USA, 1229–1230.
- Smith T. F., and Waterman M. S. 1981. Identification of Common Molecular Subsequences. *Journal of Molecular Biology* 147(1): 195–197.
- Xin-Jing Wang, Xudong Tu, Dan Feng, and Lei Zhang. 2009. Ranking community answers by modeling question-answer relationships via analogical reasoning. In *Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval (SIGIR '09)*. ACM, New York, NY, USA, 179-186.
- X. Liu, W. B. Croft, and M. Koll. Finding experts in community-based question-answering services. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 315–316, New York, NY, USA, 2005.
- ZHOU, Y., CONG, G., CUI, B., JENSEN, C. S., AND YAO, J. 2009. Routing questions to the right users in online communities. In *Proceedings of the 2009 IEEE International Conference on Data Engineering*.



References

- Fei Song and W. Bruce Croft . *A general language model for information retrieval, CIKM, 1999*
- John Lafferty and Chengxiang Zhai. *Document language models, query models, and risk minimization for information retrieval, SIGIR, 2001*
- Chengxiang Zhai and John Lafferty. *A study of smoothing methods for language models applied to information retrieval, TOIS*
- Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. *Finding Semantically Similar Questions Based on Their Answers, SIGIR, 2005*
- Jiwoon Jeon, W. Bruce Croft, and Joon Ho Lee. *Finding similar questions in large question and answer archives, CIKM, 2005*
- Gao Cong, Long Wang, Chin-Yew Lin, Young-In Song, and Yueheng Sun. *Finding question-answer pairs from online forums, SIGIR, 2008*
- Xiaobing Xue, Jiwoon Jeon, and W. Bruce Croft. *Retrieval Models for Question and Answer Archives, SIGIR, 2008*
- Mihai Surdeanu, Massimiliano Ciaramita, and Hugo Zaragoza. *Learning to Rank Answers on Large Online QA Collections. In Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL 2008), 2008.*

