A Data-Driven Approach to Question Subjectivity Identification in Community Question Answering

Tom Chao Zhou[1], Xiance Si[2], Edward Y. Chang[2],
Irwin King[3,1] and Michael R. Lyu[1]

[1]The Chinese University of Hong Kong
[2]Google Research
[3]AT & T Labs Research

AAAI 2012, Toronto, Canada

Sometimes the questions are complicated
and the answers are simple.
--Dr. Seuss

what is the capital of canada

Chao Zhou 0 + Share

About 622,000,000 results (0.22 seconds)

SafeSearch on

What is the capital city of **Canada**
wiki.answers.com › ... › Countries States and Cities › Canada - Cached +1
The **capital of Canada** is Ottawa, Ontario. It was named the national capital by Queen Victoria on December 31, 1857, but the government did not move there ...

Ottawa - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Ottawa - Cached
ˈɒtəwɑː/ or /ˈɒtəwə/) is the **capital of Canada**. It is the second largest city in Ontario and the fourth largest city in the country. The city is located on the south ...
↳ Demographics of Ottawa - Neighbourhoods - Geography of Ottawa - Ottawa River

Capital Cities of **Canada**
canadaonline.about.com/od/cities/a/**capital**-cities-**canada**.htm - Cached
The **capital** cities of **Canada** - the national **capital** and **capital** cities of the provinces and territories of **Canada**.

What is the Capital of Canada?
go**canada**.about.com/b/.../**what-is-the-capital-of-canada**.htm - Cached
20 Jan 2010 – Toronto and Montreal may be better known, but Ottawa, Ontario, is **Canada's capital** city. Ottawa is a charming city to visit; it has a cultured, yet ...

Map of **Canada** with Provincial **Capitals**

Ottawa

Ottawa is the capital of Canada. It is the second largest city in Ontario and the fourth largest city in the country. The city is located on the south bank of the Ottawa River in the eastern portion of Southern Ontario.
Wikipedia

Population: 812,135 (2006)
Area: 1,073 sq miles (2,779 km²)
Weather: 59° F, Wind S at 8 mph, 94% Humidity
Local time: 11:15pm Thursday (EDT)
Province: Ontario

3

what is the capital of canada

Chao Zhou 0 + Share

About 622,000,000 results (0.22 seconds)

SafeSearch on

What is the capital city of Canada
wiki.answers.com › ... › Countries States and Cities › Canada - Cached +1
The **capital of Canada** is Ottawa, Ontario. It was named the national capital by Queen Victoria on December 31, 1857, but the government did not move there ...

Ottawa - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Ottawa - Cached
ˈɒtəwɑː/ or /ˈɒtəwə/) is the **capital of Canada**. It is the second largest city in Ontario and the fourth largest city in the country. The city is located on the south ...
↳ Demographics of Ottawa - Neighbourhoods - Geography of Ottawa - Ottawa River

Capital Cities of Canada
canadaonline.about.com/od/cities/a/**capital**-cities-**canada**.htm - Cached
The **capital** cities of **Canada** - the national **capital** and **capital** cities of the provinces and territories of **Canada**.

What is the Capital of Canada?
go**canada**.about.com/b/.../**what-is-the-capital-of-canada**.htm - Cached
20 Jan 2010 – Toronto and Montreal may be better known, but Ottawa, Ontario, is **Canada's capital** city. Ottawa is a charming city to visit; it has a cultured, yet ...

Map of **Canada** with Provincial **Capitals**

Ottawa

Ottawa is the capital of Canada. It is the second largest city in Ontario and the fourth largest city in the country. The city is located on the south bank of the Ottawa River in the eastern portion of Southern Ontario.
Wikipedia

Population: 812,135 (2006)
Area: 1,073 sq miles (2,779 km²)
Weather: 59° F, Wind S at 8 mph, 94% Humidity
Local time: 11:15pm Thursday (EDT)
Province: Ontario

3

what is the capital of canada

Chao Zhou 0 + Share

About 622,000,000 results (0.22 seconds)

SafeSearch on

What is the capital city of Canada
wiki.answers.com › ... › Countries States and Cities › Canada - Cached +1
The **capital of Canada** is Ottawa, Ontario. It was named the national capital by Queen Victoria on December 31, 1857, but the government did not move there ...

Ottawa - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Ottawa - Cached
ˈɒtəwɑː/ or /ˈɒtəwə/) is the **capital of Canada**. It is the second largest city in Ontario and the fourth largest city in the country. The city is located on the south ...
↳ Demographics of Ottawa - Neighbourhoods - Geography of Ottawa - Ottawa River

Capital Cities of Canada
canadaonline.about.com/od/cities/a/**capital**-cities-**canada**.htm - Cached
The **capital** cities of **Canada** - the national **capital** and **capital** cities of the provinces and territories of **Canada**.

What is the Capital of Canada?
go**canada**.about.com/b/.../**what-is-the-capital-of-canada**.htm - Cached
20 Jan 2010 – Toronto and Montreal may be better known, but Ottawa, Ontario, is **Canada's capital** city. Ottawa is a charming city to visit; it has a cultured, yet ...

Map of **Canada** with Provincial **Capitals**

Ottawa

Ottawa is the capital of Canada. It is the second largest city in Ontario and the fourth largest city in the country. The city is located on the south bank of the Ottawa River in the eastern portion of Southern Ontario.
Wikipedia

Population: 812,135 (2006)
Area: 1,073 sq miles (2,779 km²)
Weather: 59° F, Wind S at 8 mph, 94% Humidity
Local time: 11:15pm Thursday (EDT)
Province: Ontario

3

Facebook | Google | Major Internet Companies | Technology Trends | The Internet Edit

What will be the next big trend on the Internet after "search" of Google and "social" of Facebook? Edit

Add Question Details

3+ Comments • Post (5) • Wiki • Options • Redirect Question

Answer Wiki

- Mobile
- Influence
- Connecting business to clients dynamically
- Personalization
- Offline-to-Online
- Social Discovery
- Connecting people through the interest graph

Edit

39 Answers

Facebook | Google | Major Internet Companies | Technology Trends | The Internet Edit

What will be the next big trend on the Internet after "search" of Google and "social" of Facebook? Edit

Add Question Details

3+ Comments • Post (5) • Wiki • Options • Redirect Question

Answer Wiki

- Mobile
- Influence
- Connecting business to clients dynamically
- Personalization
- Offline-to-Online
- Social Discovery
- Connecting people through the interest graph

Edit

39 Answers

Facebook | Google | Major Internet Companies | Technology Trends | The Internet | Edit

What will be the next big trend on the Internet after "search" of Google and "social" of Facebook? Edit

Add Question Details

3+ Comments • Post (5) • Wiki • Options • Redirect Question

Answer Wiki

- Mobile
- Influence
- Connecting business to clients dynamically
- Personalization
- Offline-to-Online
- Social Discovery
- Connecting people through the interest graph

Edit

39 Answers

Research Topic

Automatic Question Answering (AQA)

Research Topic

Information Retrieval

Natural Language Processing

Automatic Question Answering (AQA)

Research Topic

Automatic Question Answering (AQA)

Research Topic

Automatic Question Answering (AQA)

Automatic **Factual** Question Answering (AFQA)

Automatic **Subjective** Question Answering (ASQA)

Automatic Factual Question Answering

- AFQA
  - Fact, a fixed answer
  - Compute from fact resources, e.g. Wikipedia
  - Harabagiu et al. 2001, Demner-Fushman and Lin 2007

Automatic Factual Question Answering

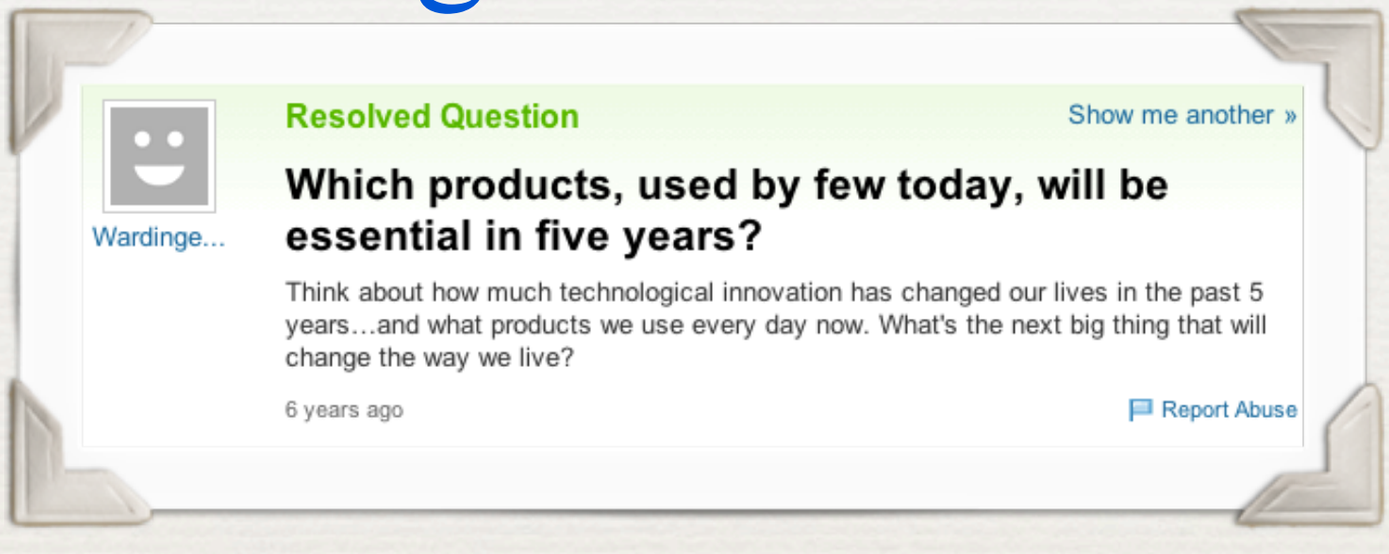
- AFQA
  - Fact, a fixed answer
  - Compute from fact resources, e.g. Wikipedia
  - Harabagiu et al. 2001, Demner-Fushman and Lin 2007
- IBM Watson
  - Win Jeopardy! Challenge
  - Q: He was president during the War of 1812.
  - A: Who is James Madison?

6

Automatic Subjective Question Answering

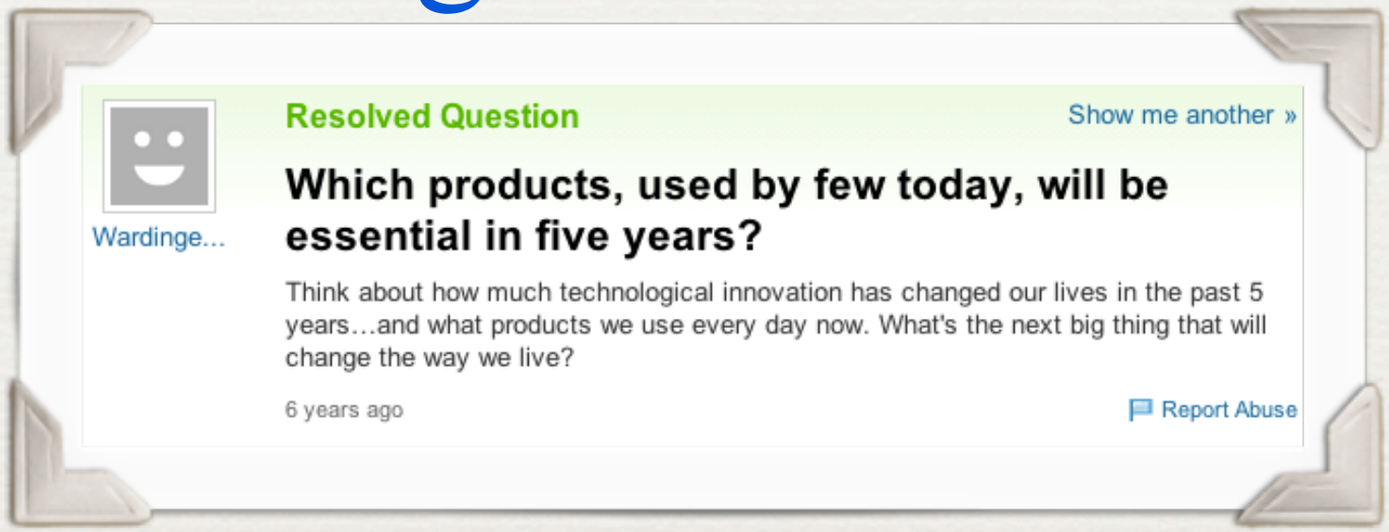
- ASQA
  - Opinions
  - Summarized answer, different perspectives
  - Soricut and Brill 2004, Li et al. 2008

Automatic Subjective Question Answering



- ASQA
- Opinions
- Summarized answer, different perspectives
- Soricut and Brill 2004, Li et al. 2008

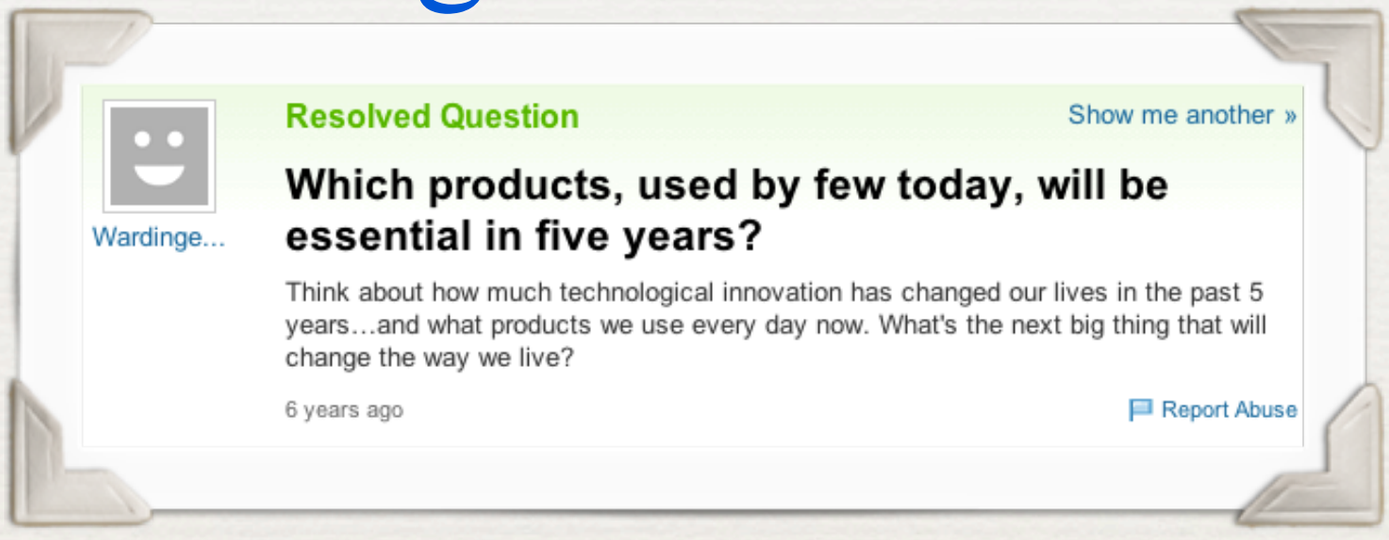
Automatic Subjective Question Answering



- ASQA
- Opinions
- Summarized answer, different perspectives

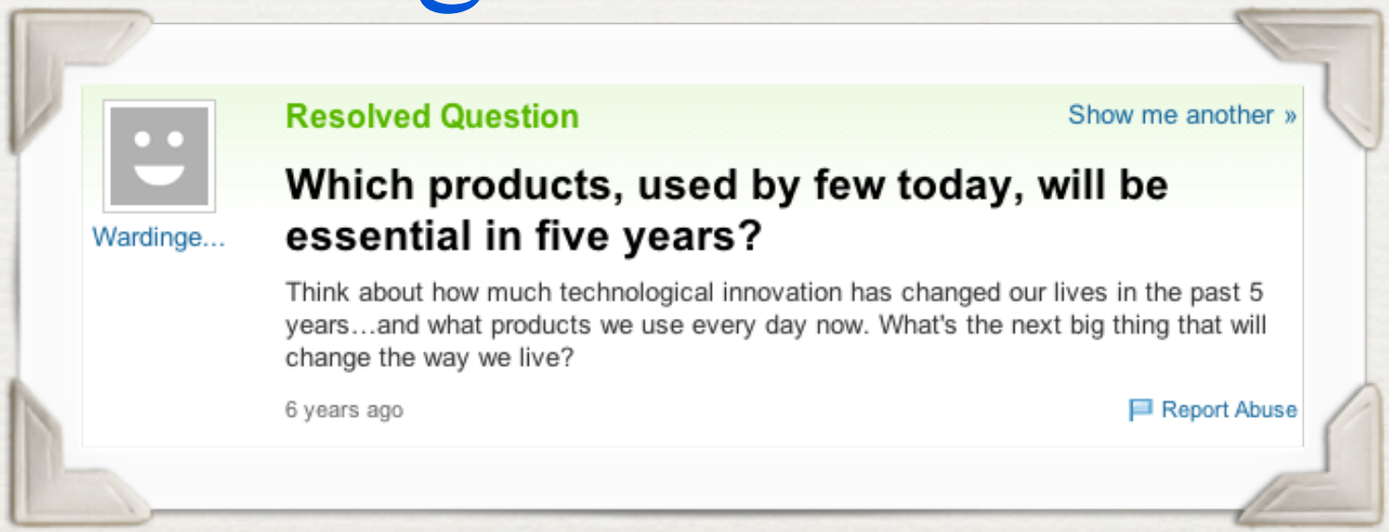
- Soricut and Brill 2004, Li et al. 2008

Automatic Subjective Question Answering



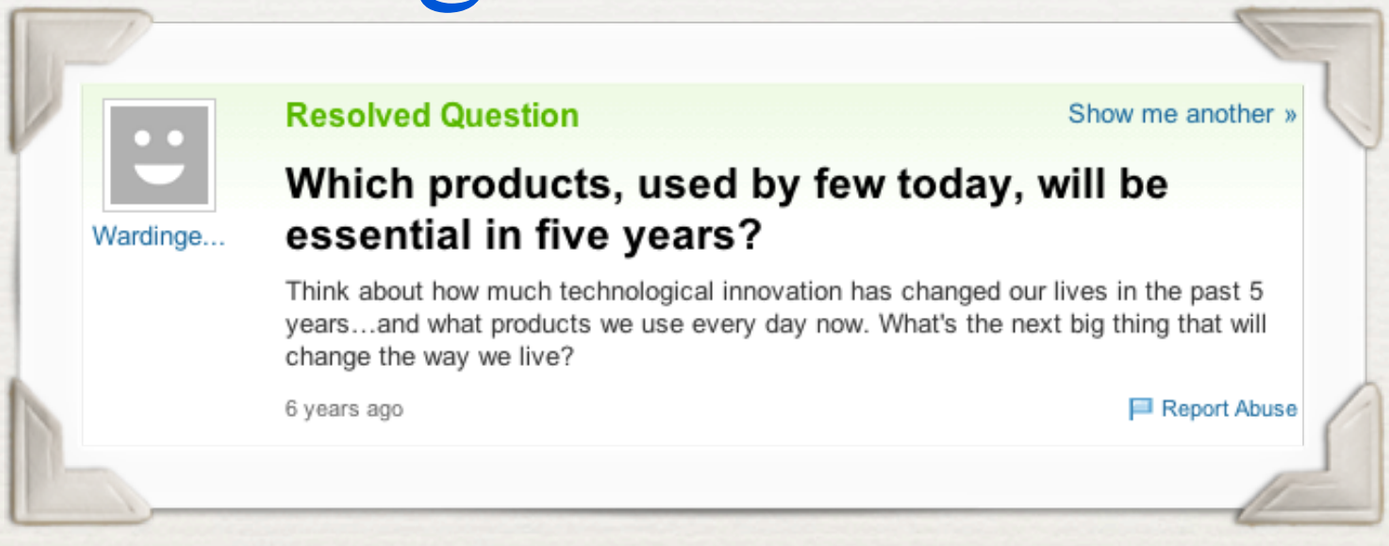
- ASQA
- Opinions
- Summarized answer, different perspectives
- Soricut and Brill 2004, Li et al. 2008

Automatic Subjective Question Answering



- ASQA
- Opinions
- Summarized answer, different perspectives
- Soricut and Brill 2004, Li et al. 2008

Automatic Subjective Question Answering



- ASQA
- Opinions
- Summarized answer, different perspectives
- Soricut and Brill 2004, Li et al. 2008

More challenge!

7

Outline

- Motivation
- Social Signal
- Feature
- Experiments
- Related Work
- Conclusions && Future Work

8

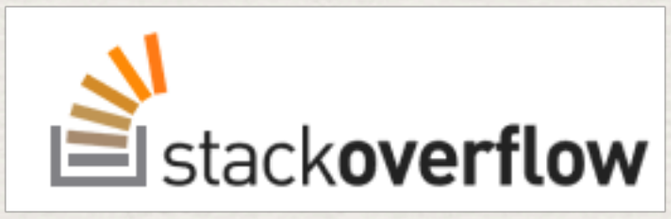
Motivation

- Alternative to ASQA?

Motivation

- Alternative to ASQA?
- Yes! Community Question Answering (CQA)

- Platform
- Post questions
- Answer question
- Give feedbacks



Challenge

- Challenge of using CQA for question analysis
  - Ill-phrased, vague and complex
  - Lack of labeled data

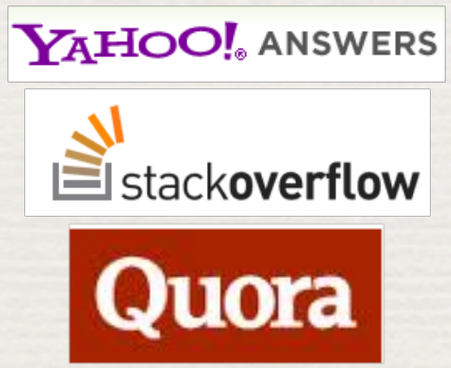
10

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

--Alon Halevy, Peter Norvig and Fernando Pereira

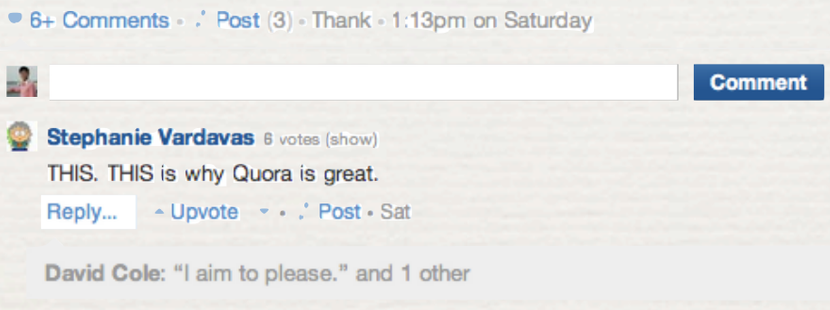
11

Social Signal



12

Social Signal



Commenting

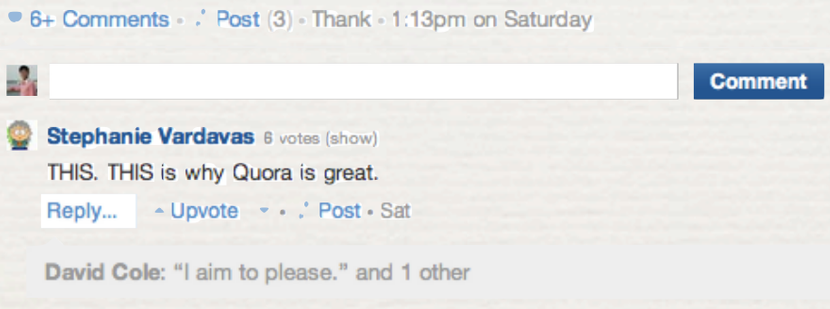
YAHOO! ANSWERS

stackoverflow

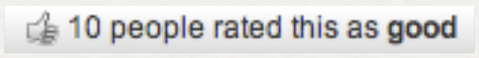
Quora

12

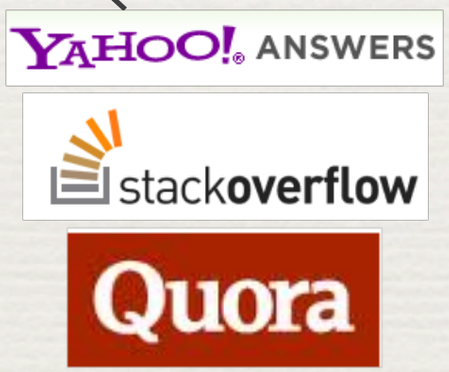
Social Signal



Commenting

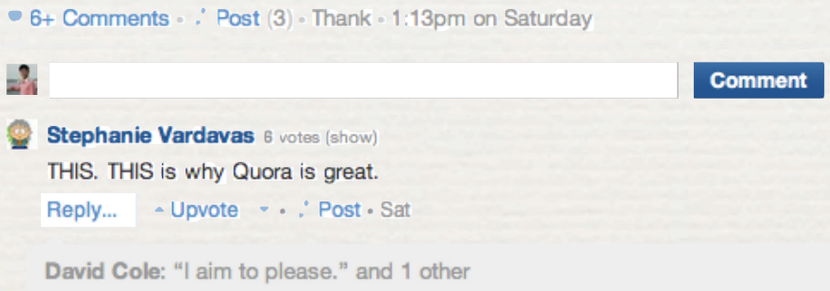


Rating

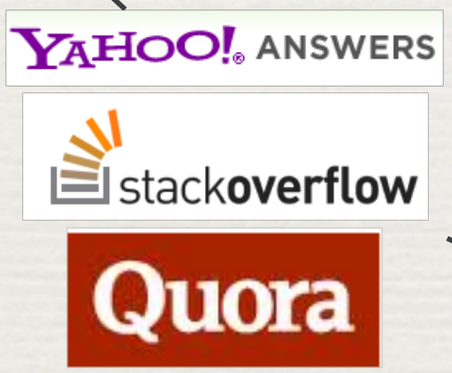


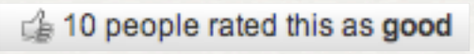
12

Social Signal



Commenting





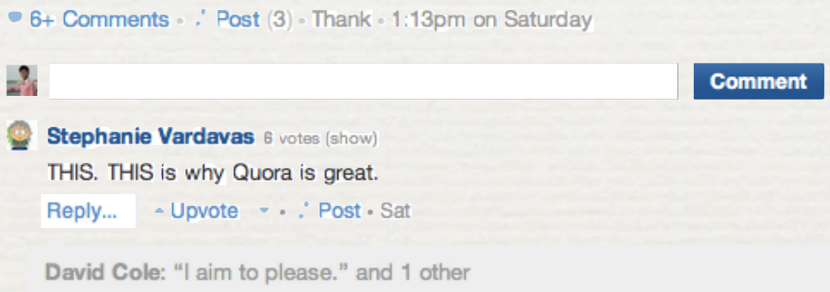
Rating

Voting

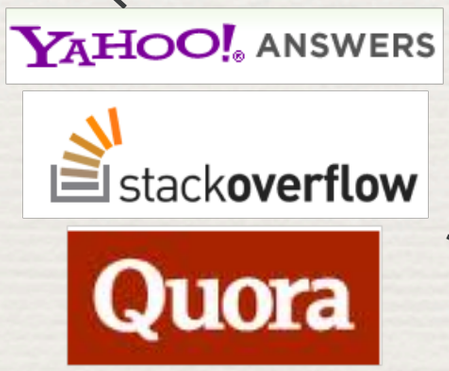
12

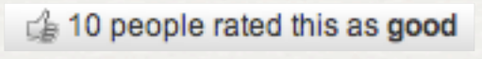
Social Signal

Social Signal



Commenting





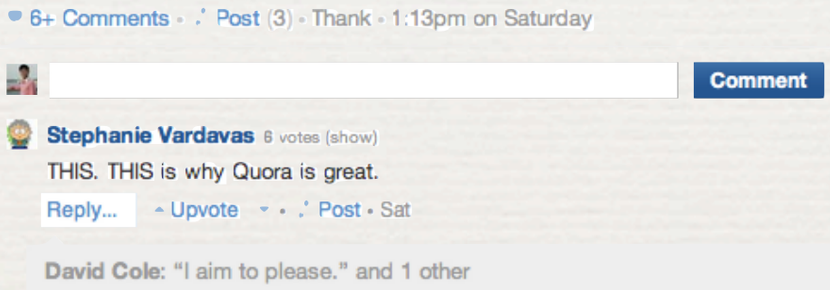
Rating

Voting

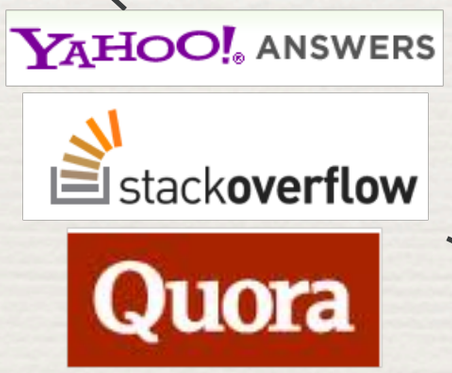
12

Social Signal

Social Signal



Commenting



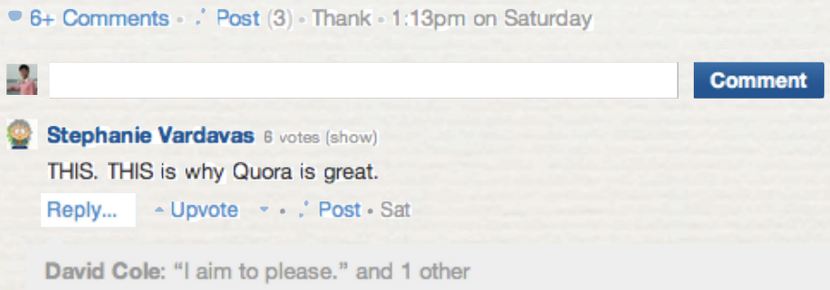
Rating

Voting

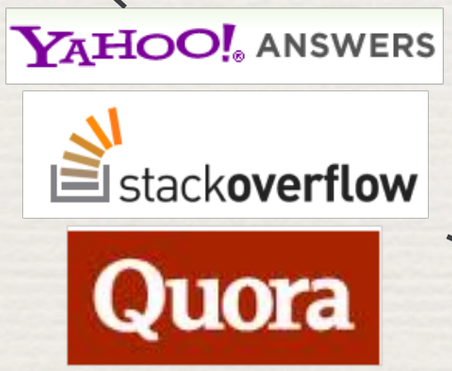
Community Wisdom

Social Signal

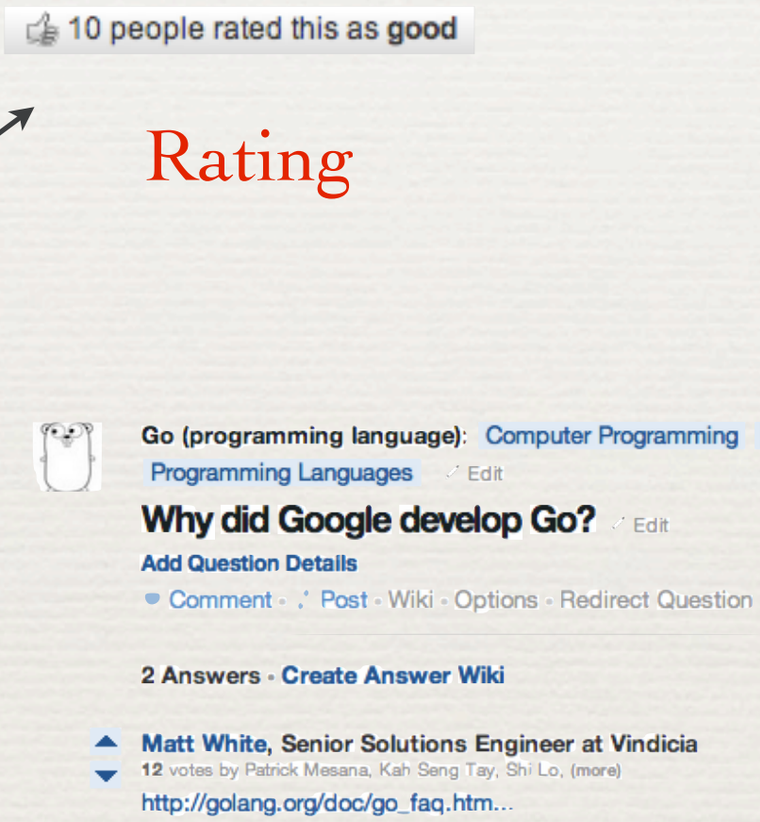
Social Signal



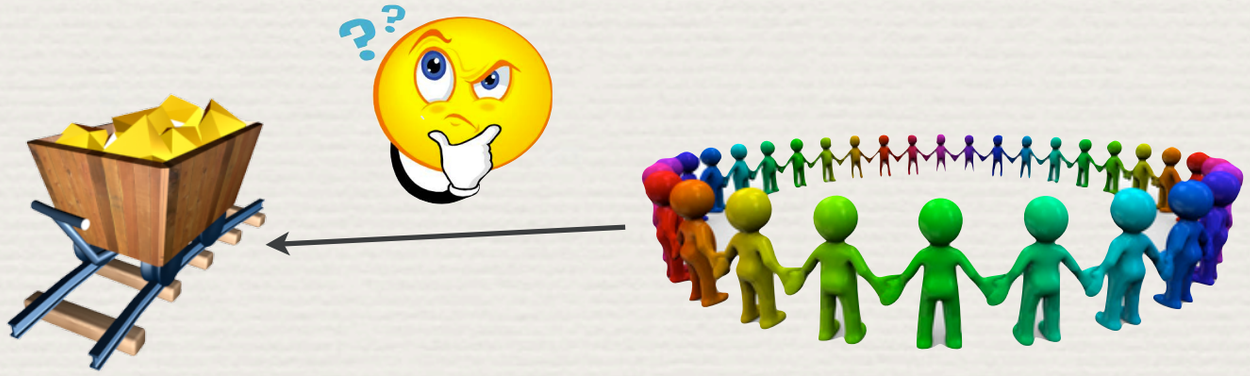
Commenting



Rating



Voting



Community Wisdom

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

13

Motivation

- Question Subjectivity Identification (QSI)
  - A test case

14

Motivation

- Question Subjectivity Identification (QSI)
  - A test case
- Subjective Question
  - One or more subjective answers

Motivation

- Question Subjectivity Identification (QSI)
  - A test case
- Subjective Question
  - One or more subjective answers
- Objective Question
  - Authoritative answer, common knowledge or universal truth

14

Examples

- Subjective
  - Does anyone remember a book called the “Robe” by Lloyd C. Douglas? What did you think of it?
  - What was your favorite novel that you read?
  - What are the ways to calm myself when flying?
- Objective
  - When and how did Tom Thompson die? He is one of the group of Seven.
  - What makes the color blue?
  - Was Roy Orbison blind?

15

Why Perform QSI?

- More accurately identify similar questions
- Better rank or filter the answers
- Crucial component of inferring user intent
- Subjective question --> Route to users
- Objective question --> Trigger AFQA

16

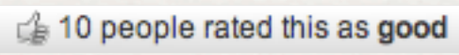
Social Signal

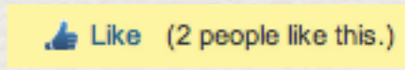
- QSI
  - Classification task
  - Subjective question: positive
  - Objective question: negative

17

Social Signal

- Like: like an answer if they find the answer useful

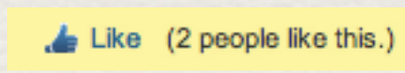




18

Social Signal

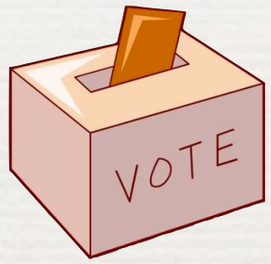
- Like: like an answer if they find the answer useful
- Intuition
  - Subjective
    - Answers are opinions, different tastes
    - Best answer receives similar number of likes with other answers
  - Objective
    - Like an answer which explains universal truth in the most detail
    - Best answer receives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



18

Social Signal

- Vote: users could vote for best answer



Quora

Michael Forrest Jones,
129 votes by Miles Dolphin, Tra

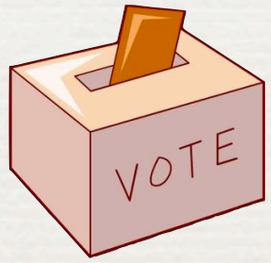
YAHOO! ANSWERS

Baidu 知道

19

Social Signal

- Vote: 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 contains the most fact
    - Percentage of votes of best answer is high



19

Social Signal

- Source: references to authoritative resources

- Intuition
  - Only available for objective question that has fact answer

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...

20

Social Signal

- Source: references to authoritative resources

- Intuition
  - Only available for objective question that has fact answer

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...

20

Social Signal

- Source: references to authoritative resources

- Intuition
  - Only available for objective question that has fact answer

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

20

Social Signal

- Source: references to authoritative resources

- Intuition
  - Only available for objective question that has fact answer

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

20

Social Signal

- Poll and Survey
- Intuition

  - User intent is to seek opinions
  - Very likely to be subjective

21

Social Signal

- Poll and Survey
- Intuition

  - 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?

21

Social Signal

- Answer Number: the number of posted answers to each question varies

Social Signal

- Answer Number: the number of posted answers to each question varies
- Intuition
  - Subjective

    - Post opinions even they notice there are other answers
  - Objective
    - May not post answers to questions that haveM received other answers since an expected answer is usually fixed
- A large answer number indicate subjectivity
- HOWEVER, a small answer number may be due to many reasons, such as objectivity, small page views

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

23

Feature

- Word
  - Effective in many question answering applications
  - Each word is represented with term frequency (tf)

Feature

- Word
  - Effective in many question answering applications
  - Each word is represented with term frequency (tf)
- Word n-gram
  - Li et al. 2008 (supervised), Li, Liu and Agichtein 2008 (semi-supervised) observe the performance gain of word n-gram compared with word is not large
  - May be due to sparsity of their small amount of training data
  - Each word n-gram is represented with term frequency

Feature

- Question Length
  - Information needs of subjective questions are complex, users use descriptions to explain, Wang et al. 2010 ==> larger question length
  - Divide into 10 buckets, corresponding bucket number is a feature

25

Feature

- Request Word
  - Particular words to explicitly indicate their request for seeking opinions
    - E.g. Should I buy the blackberry torch 9860 or the curve 9380 and why?
- should, might, anyone, can, shall, may, would, could, please
- Total number of request word as a feature

26

Feature

- Subjectivity Clue
  - External lexicon, Wilson et al. 2005, over 8000 clues
  - Manually compiled word list from news to express opinions

27

Feature

- Punctuation Density
  - Density of punctuation marks
  - Use short sentence segments when sharing their experiences

$$PDensity(Q) = \frac{\#\text{ punctuation marks}}{\#\text{ punctuation marks} + \#\text{ words}}$$

Feature

- Grammatical Modifier
  - Inspired by opinion mining research of using grammatical modifiers on judging users' opinions
  - Adjective and adverb are considered as grammatical modifiers

29

Feature

- Entity
  - Objective
    - Expect answer is fact, leading to less relationships among entities
  - Subjective
    - More descriptions, may involve relatively complex relations

30

Experiments

Experiments

- Comparison Methods
  - Supervised learning with manual labeling
  - CoCQA, co-training, state-of-the-art, Li, Liu and Agichtein 2008

31

Experiments

- Comparison Methods
  - Supervised learning with manual labeling
  - CoCQA, co-training, state-of-the-art, Li, Liu and Agichtein 2008
- Dataset
  - Yahoo! Answers, 4,375,429 questions with associated social signals
  - Ground truth: adapted from Li, Liu and Agichtein 2008

31

Experiments

- Comparison Methods
  - Supervised learning with manual labeling
  - CoCQA, co-training, state-of-the-art, Li, Liu and Agichtein 2008
- Dataset
  - Yahoo! Answers, 4,375,429 questions with associated social signals
  - Ground truth: adapted from Li, Liu and Agichtein 2008
- Classification Method
  - Naive Bayes with add-one smoothing, more effective than SVM
  - Parallelized using MapReduce

31

Experiments

Performance of word n-gram	
Method	Precision on Sub
Supervised	0.6596
CoCQA	0.6861 (+4.2%)
L+V+PS+AN+S	0.6626(+0.45%)
L	0.5714(-13.37%)
V+PS+AN+S	0.6981(+5.84%)
PS+AN+S	0.6915(+4.84%)
V+PS+AN	**0.7214(+9.37%)**
V+AN	0.7201(+9.17%)
AN+S	0.7038(+6.70%)

32

Experiments

Performance of word n-gram	
Method	Precision on Sub
Supervised	0.6596
CoCQA	0.6861 (+4.2%)
L+V+PS+AN+S	0.6626(+0.45%)
L	0.5714(-13.37%)
V+PS+AN+S	0.6981(+5.84%)
PS+AN+S	0.6915(+4.84%)
V+PS+AN	**0.7214(+9.37%)**
V+AN	0.7201(+9.17%)
AN+S	0.7038(+6.70%)

- CoCQA utilize some amount of unlabeled data, but it could only utilize a small amount (3, 000 questions)

32

Experiments

Performance of word n-gram	
Method	Precision on Sub
Supervised	0.6596
CoCQA	0.6861 (+4.2%)
L+V+PS+AN+S	0.6626(+0.45%)
L	0.5714(-13.37%)
V+PS+AN+S	0.6981(+5.84%)
PS+AN+S	0.6915(+4.84%)
V+PS+AN	**0.7214(+9.37%)**
V+AN	0.7201(+9.17%)
AN+S	0.7038(+6.70%)

- CoCQA utilize some amount of unlabeled data, but it could only utilize a small amount (3, 000 questions)
- Effectiveness of collecting training data using well-designed social signals

32

Experiments

Performance of word n-gram	
Method	Precision on Sub
Supervised	0.6596
CoCQA	0.6861 (+4.2%)
L+V+PS+AN+S	0.6626(+0.45%)
L	0.5714(-13.37%)
V+PS+AN+S	0.6981(+5.84%)
PS+AN+S	0.6915(+4.84%)
V+PS+AN	**0.7214(+9.37%)**
V+AN	0.7201(+9.17%)
AN+S	0.7038(+6.70%)

- CoCQA utilize some amount of unlabeled data, but it could only utilize a small amount (3, 000 questions)
- Effectiveness of collecting training data using well-designed social signals
- These social signals could be found in almost all CQA

32

Experiments

Performance of word n-gram	
Method	Precision on Sub
Supervised	0.6596
CoCQA	0.6861 (+4.2%)
L+V+PS+AN+S	0.6626(+0.45%)
L	0.5714(-13.37%)
V+PS+AN+S	0.6981(+5.84%)
PS+AN+S	0.6915(+4.84%)
V+PS+AN	**0.7214(+9.37%)**
V+AN	0.7201(+9.17%)
AN+S	0.7038(+6.70%)

- CoCQA utilize some amount of unlabeled data, but it could only utilize a small amount (3, 000 questions)
- Effectiveness of collecting training data using well-designed social signals
- These social signals could be found in almost all CQA
- Why Like signal is not effective? Complex of best answer selection criteria of the asker, social-emotional factor affect a lot, Kim et al. 2007

32

Experiments

Performance of Word and n-gram		
Method	Word	n-gram
Supervised	0.6380	0.6596 (+3.39%)
CoCQA	0.6432	0.6861 (+6.66%)
V+PS+AN	0.6707	**0.7214 (+7.56%)**
V+AN	0.6265	**0.7201 (+14.94%)**
AN+S	0.6157	0.7038 (+14.31%)

Experiments

Performance of Word and n-gram		
Method	Word	n-gram
Supervised	0.6380	0.6596 (+3.39%)
CoCQA	0.6432	0.6861 (+6.66%)
V+PS+AN	0.6707	**0.7214 (+7.56%)**
V+AN	0.6265	**0.7201 (+14.94%)**
AN+S	0.6157	0.7038 (+14.31%)

- Better performance using word n-gram compared with word

Experiments

Performance of Word and n-gram		
Method	Word	n-gram
Supervised	0.6380	0.6596 (+3.39%)
CoCQA	0.6432	0.6861 (+6.66%)
V+PS+AN	0.6707	**0.7214 (+7.56%)**
V+AN	0.6265	**0.7201 (+14.94%)**
AN+S	0.6157	0.7038 (+14.31%)

- Better performance using word n-gram compared with word
- Social signals achieve on average 12.27% relative gain

Experiments

Performance of Word and n-gram		
Method	Word	n-gram
Supervised	0.6380	0.6596 (+3.39%)
CoCQA	0.6432	0.6861 (+6.66%)
V+PS+AN	0.6707	**0.7214 (+7.56%)**
V+AN	0.6265	**0.7201 (+14.94%)**
AN+S	0.6157	0.7038 (+14.31%)

- Better performance using word n-gram compared with word
- Social signals achieve on average 12.27% relative gain
- Supervised, manually labeling, sparsity

33

Experiments

Performance of Word and n-gram		
Method	Word	n-gram
Supervised	0.6380	0.6596 (+3.39%)
CoCQA	0.6432	0.6861 (+6.66%)
V+PS+AN	0.6707	**0.7214 (+7.56%)**
V+AN	0.6265	**0.7201 (+14.94%)**
AN+S	0.6157	0.7038 (+14.31%)

- Better performance using word n-gram compared with word
- Social signals achieve on average 12.27% relative gain
- Supervised, manually labeling, sparsity
- CoCQA, use several thousand unlabeled data, tackled sparsity to some extent

33

Experiments

Performance of Word and n-gram		
Method	Word	n-gram
Supervised	0.6380	0.6596 (+3.39%)
CoCQA	0.6432	0.6861 (+6.66%)
V+PS+AN	0.6707	**0.7214 (+7.56%)**
V+AN	0.6265	**0.7201 (+14.94%)**
AN+S	0.6157	0.7038 (+14.31%)

- Better performance using word n-gram compared with word
- Social signals achieve on average 12.27% relative gain
- Supervised, manually labeling, sparsity
- CoCQA, use several thousand unlabeled data, tackled sparsity to some extent
- Training data collected by social signals, large amount, better solve data sparsity problem

33

Experiments

Performance of varying amount of training data				
Method	20%	40%	90%	100%
V+AN	0.6549	0.7004	0.7188	0.7201
AN+S	0.6550	0.6696	0.6842	0.7038
V+PS +AN	0.6640	0.6846	0.7037	0.7214

- Three best performing combinations of social signals
- Increase training data, performances improve

Experiments

Performance of Heuristic Features									
	ngram	+qlength	+rword	+sclue	+pdensity	+gmodifier	+entity	heuristic features	+heuristicf eatures
Precision	0.6596	0.6896	0.6834	0.6799	0.7000	0.6950	0.6801	0.6995	**0.7337 (+11.23%)**

- Adding any heuristic feature to word n-gram improve precision
- Combining heuristic feature and word n-gram achieves 11.23% relative gain over n-gram

35

Related Work

- Factual Question Classification
  - Stoyanov et al. 2005, Ferrucci et al. 2010

36

Related Work

- Factual Question Classification
  - Stoyanov et al. 2005, Ferrucci et al. 2010
- Automatic Question Answering
  - AFQA: Soricut and Brill 2004, Harabagiu et al. 2001, Demner-Fushman and Lin 2007, Ferruci et al. 2001
  - ASQA: Dang et al. 2007 TREC

36

Related Work

- Factual Question Classification
  - Stoyanov et al. 2005, Ferrucci et al. 2010
- Automatic Question Answering
  - AFQA: Soricut and Brill 2004, Harabagiu et al. 2001, Demner-Fushman and Lin 2007, Ferruci et al. 2001
  - ASQA: Dang et al. 2007 TREC
- Question Subjectivity Identification
  - Li et al. 2008, Li, Liu, and Agichtein 2008, Aikawa et al. 2011

Conclusions && Future Work

- Conclusions
  - Data-driven approach for QSI by utilizing social signals in CQA with NO manual labeling
  - Study various light-weight features for QSI
    - Experiments indicate effectiveness of proposed approaches
- Future Work
  - Feature investigation: semantic analysis using NLP
  - Characteristics of subjective questions, whether we could find popular semantic patterns for subjective questions

37

Acknowledgement

- Thanks AI Journal and Google for sponsoring my AAAI-12 Student Scholarship
- The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. CUHK 413210 and CUHK 415311) and two grants from Google Inc. (one for Focused Grant Project "mobile 2014" and one for Google Research Awards)

38

Thanks!

Q & A

39

FAQ

- How to define subjective or objective questions?
  - Ground truth data was created using Amazon's Mechanical Turk service. Each question was judged by 5 qualified Mechanical Turk workers. Subjectivity was decided using majority voting

40

FAQ

- Formula of like social signal
  - Positive $L(Q_{best\_answer}) \leq \frac{\sum L(Q_{answer})}{AN(Q)}$
    - L() is number of people like this answer
    - Q_{best_answer} is the best answer
    - Qanswer is an answer of Q
    - AN() is the number of answers of a question
  - Negative $L(Q_{best\_answer}) \geq \alpha \times MAX(L(Q_{other\_answer}))$
    - α is a parameter
    - Qother_answer is an answer except the best answer
    - MAX() is the maximum function

41

FAQ

- Formula of vote social signal
  - Positive $V(Q_{best\_answer}) \leq \beta$
    - V() is the percentage of votes of an answer
    - β is a parameter
  - Negative $V(Q_{best\_answer}) \geq \gamma$
    - ϒ is a parameter

42

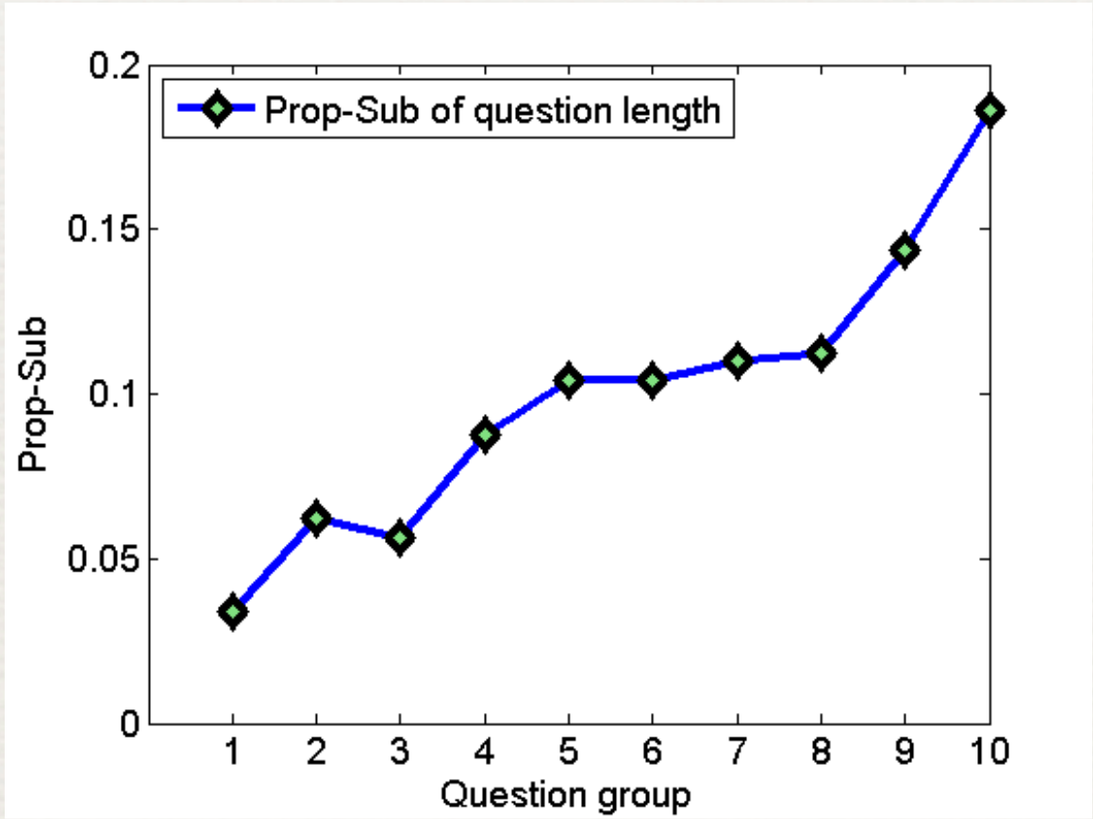
FAQ

- Formula of answer number social signal
  - Positive $AN(Q) \geq \theta$
    - AN() is the number of answers of a question
    - θ is a parameter

43

FAQ

- Proportion of subjective questions with respect to questions' lengths
- Objective question, express needs precisely
  - Which player has won the fa cup twice with 2 different teams?
- Subjective question, share his/her personal opinion together with the question
  - Has anyone read "Empire" by Orson Scott Card? This is scary. I especially liked the "Afterword" by him. Its amazing how close you can feel today to it coming true.



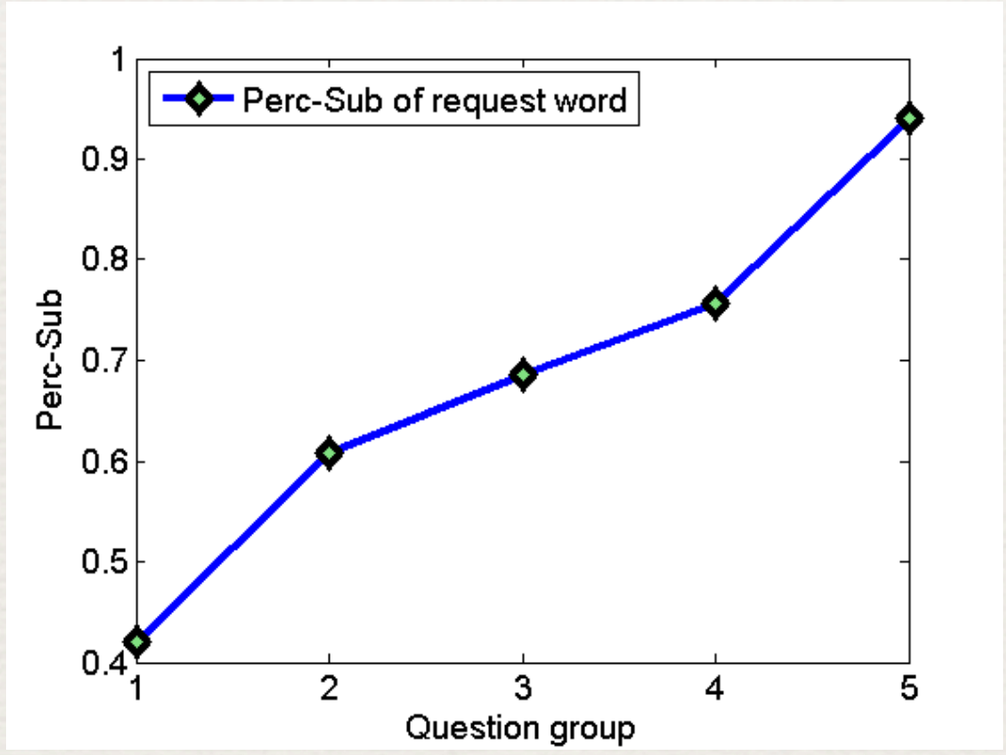
44

FAQ

- Group 1 contains questions that don't have any request word, groups 2 contains questions having 1 request word, group 3 contains 2 request words, group 4 contains 3 request words, and group 5 contains at least 4 request words

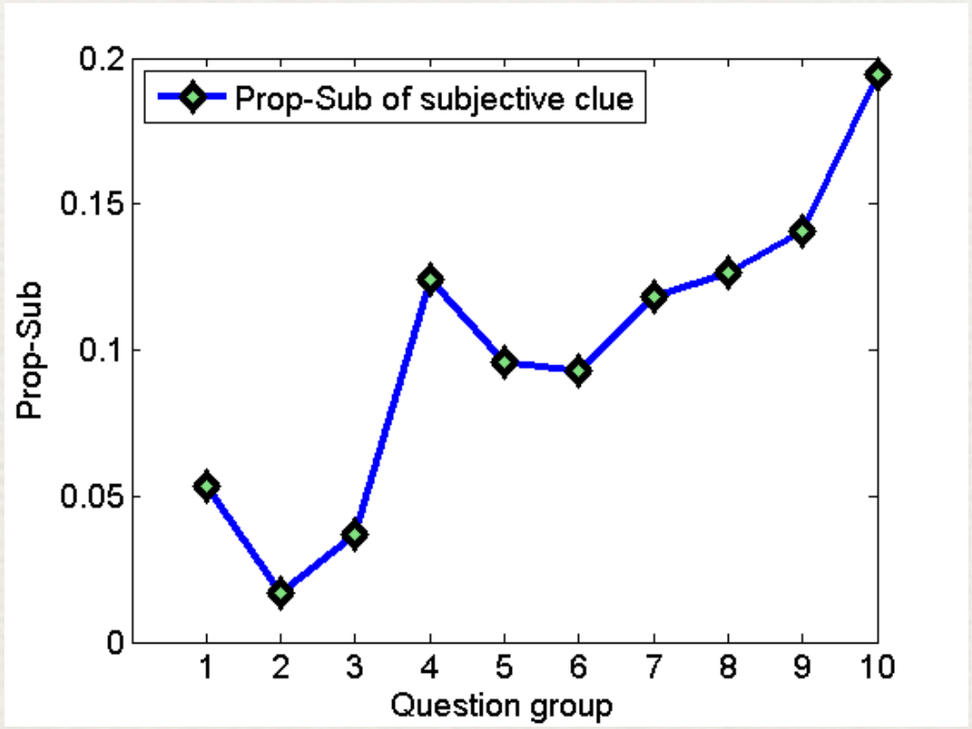
- Percentage of subjective questions among all questions in each group

- Subjective questions
  - Complicated background or detailed opinions
  - Question relatively long
  - Attract potential answerers



FAQ

- External lexicon still help distinguish between subjective and objective questions to some extent



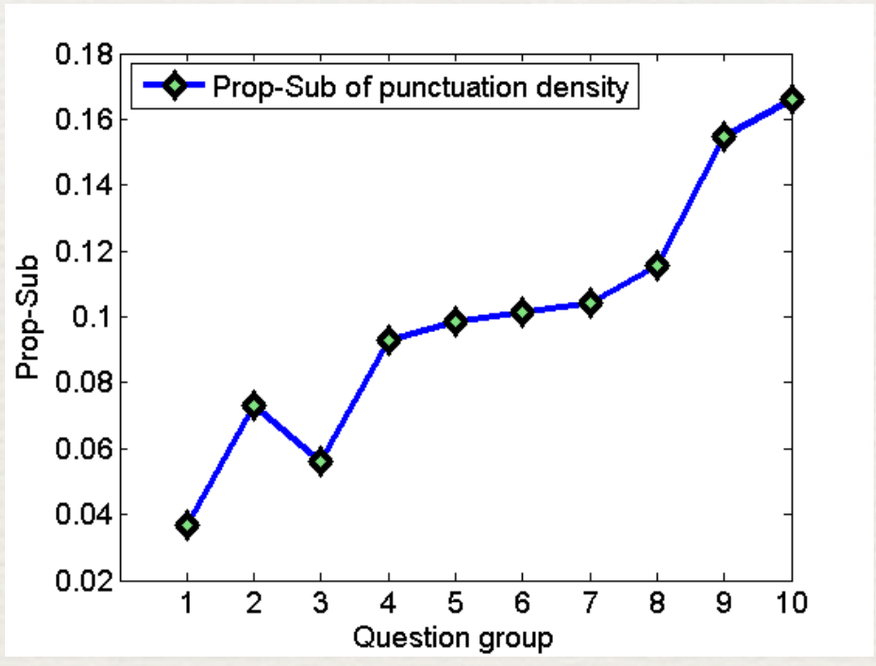
46

FAQ

- Punctuation mark density of subjective question is higher than that of objective questions

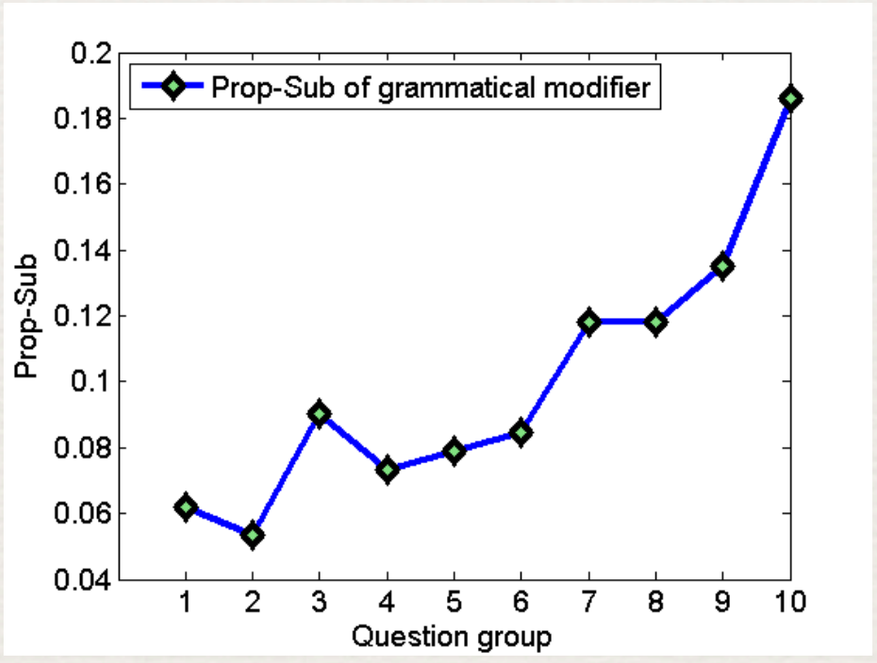
- Use short sentence segments when sharing their experiences in subjective questions

- Short sentences help better express users' feelings and opinions in asking subjective questions



FAQ

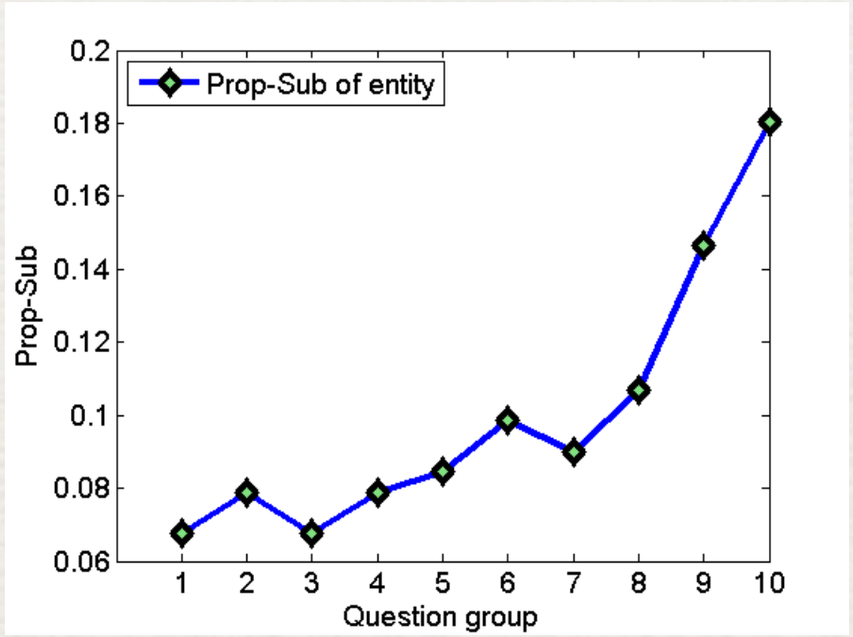
- Grammatical modifiers are commonly used to describe users' feelings, experiences, and opinions in subjective questions



FAQ

- Objective questions involve fewer entities

- Subjective questions involve more descriptions, which also contain entities



49