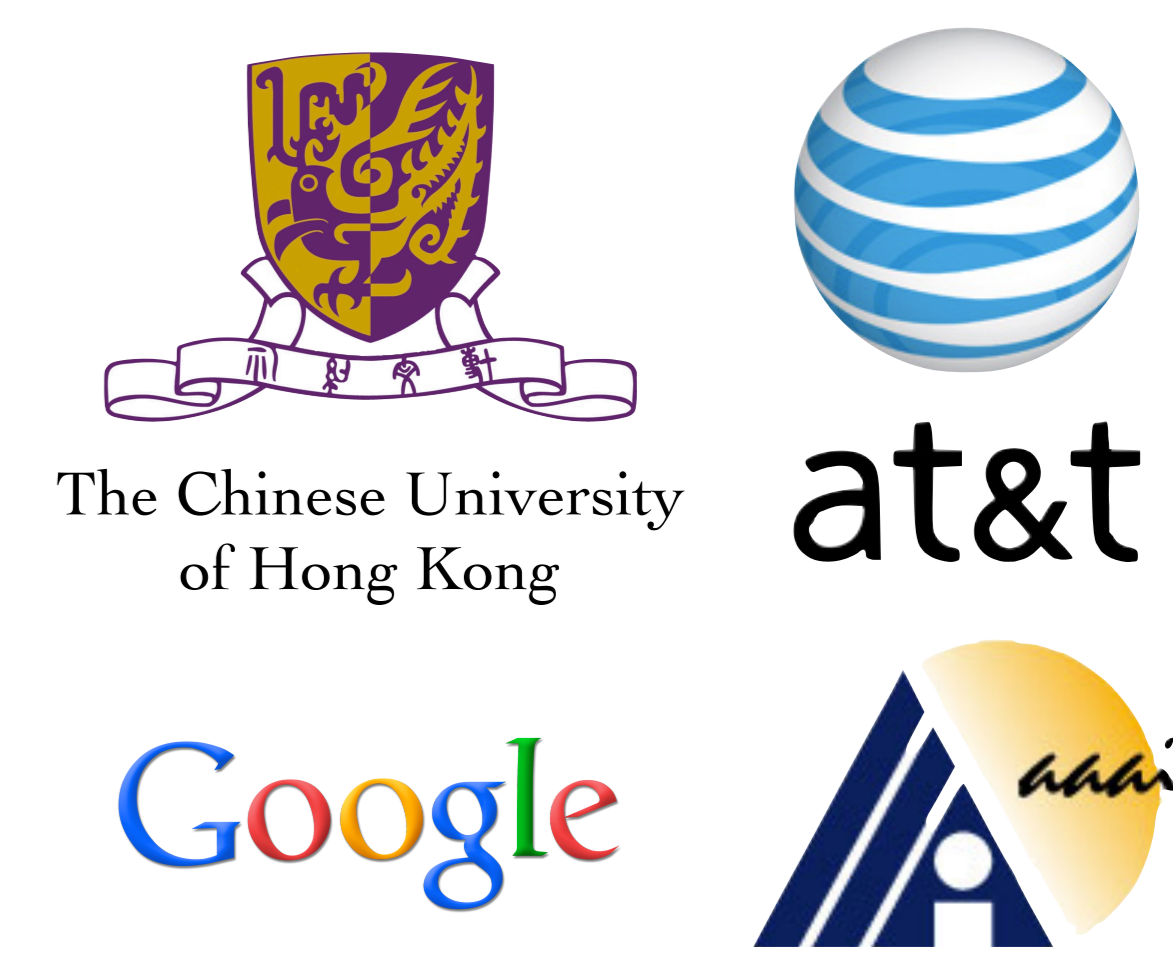
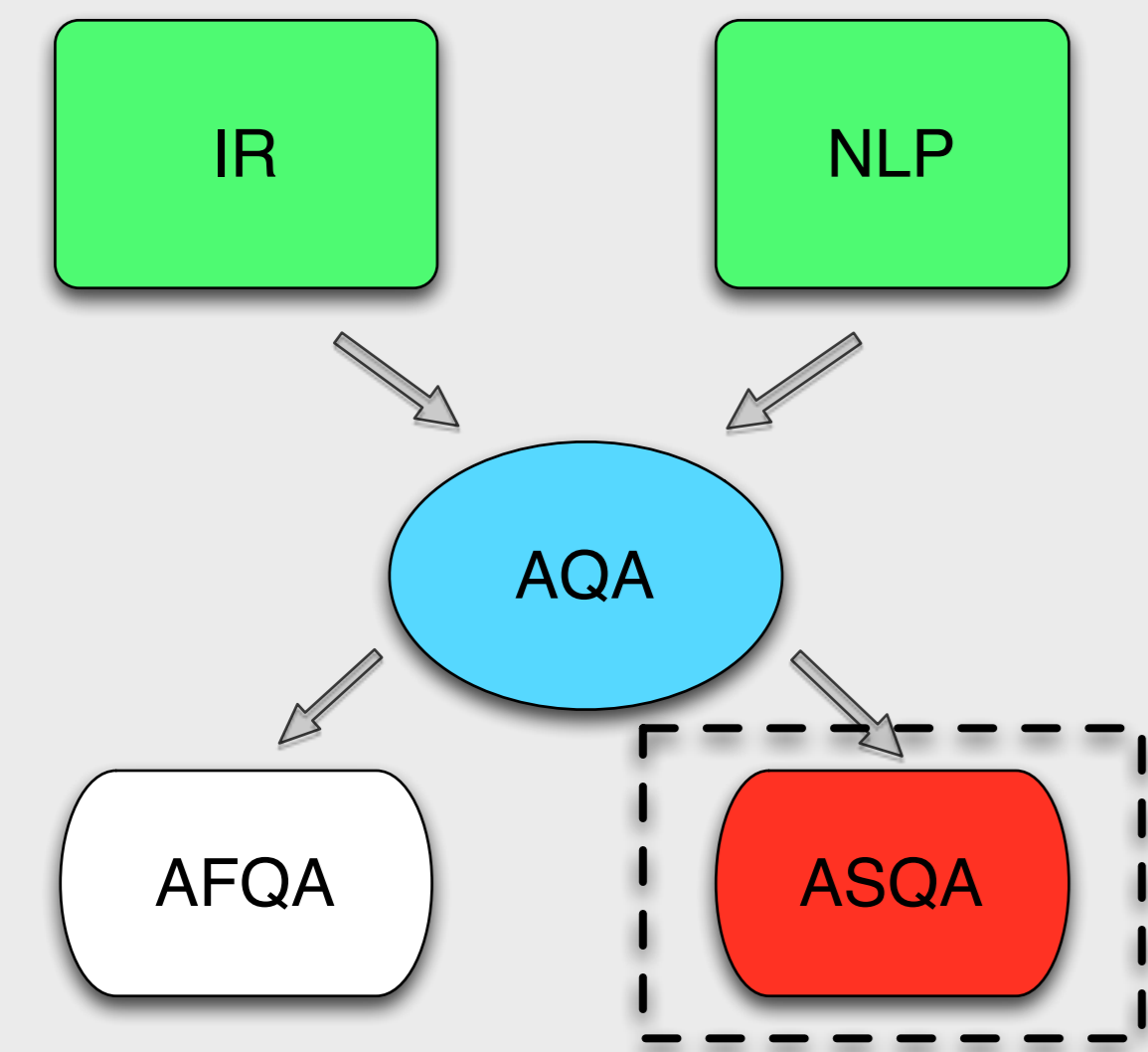


# A Data-Driven Approach to Question Subjectivity Identification in Community Question Answering



Tom Chao Zhou, Xiance Si, Edward Y. Chang, Irwin King and Michael R. Lyu

## Background and Motivation



AFQA <sup>1</sup>	ASQA <sup>2</sup>
Fact; fixed answer	Opinions; summarized from different perspectives
E.g.: What is the capital of Canada?	E.g.: Which products, used by few today, will be essential in five years?
Computed from fact resources	More challenge! Alternative?



## Social Signal

Summary of Social Signals			
Name	Description	Subjective	Objective
Like	Like an answer if find useful	Answers are opinions, different tastes; Best answer receives similar number of likes with other answers	Like an answer which explains fact in most detail; Best answer receives higher likes than other answers
Vote	Users could vote for best answer	Support different opinions; Low percentage of votes on best answer	Easy to identify the answer contains the most fact; High percentage of votes on best answer
Source	References to authoritative resources	N.A.	Only available for objective question that has fact answer
Poll and Survey	Use CQA sites as platforms to conduct poll and survey	Users' intent is to seek opinions; Very likely to be subjective	N.A.
Answer Number	The number of posted answers to each question varies	Post opinions even they notice there are other answers; A large answer number	N.A.

### Subjective Question Examples

- 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 Question Examples

- When and how did Tom Thompson die? He is one of the group of Seven.
- What makes the color blue?
- Was Roy Orbison blind?

## Feature

Summary of Features		
Name	Why	How
Word	Effective in many QA applications	Term Frequency
Word n-gram	Performance gain of word n-gram compared with word is not large in previous works; Maybe due to sparsity of their small amount of training data	Term Frequency
Question Length	Information needs of subjective questions are complex, users use descriptions to explain; Larger question length	Divide into 10 buckets, corresponding bucket number is a feature
Request Word	Particular words to explicitly indicate their request for seeking opinions	Total number of request word
Subjectivity Clue	Investigate whether external lexicon would help	Total number of subjectivity clues
Punctuation Density	Use short sentence segments when sharing their experiences	Density of punctuation marks
Grammatical Modifier	Inspired by opinion mining research of using grammatical modifiers on judging users' opinions	Total number of adjective and adverb
Entity	Expected answer of an objective question is a fact, less relationships among entities; Subjective questions have more descriptions, may involve complex relations	Total number of entities

## Results

Performance of Word n-gram		Performance of Word and n-gram		
Method	Precision on Sub	Method	Word	n-gram
Supervised	0.6596	Supervised	0.6380	0.6596 (+3.39%)
CoCQA	0.6861 (+4.2%)	CoCQA	0.6432	0.6861 (+6.66%)
L+V+PS+AN+S	0.6626(+0.45%)	V+PS+AN	0.6707	0.7214 (+7.56%)
L	0.5714 (-13.37%)	V+AN	0.6265	0.7201 (+14.94%)
V+PS+AN+S	0.6981(+5.84%)	AN+S	0.6157	0.7038 (+14.31%)
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%)			

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

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

## 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)

## Contact

• czhou@cse.cuhk.edu.hk, Tom Chao Zhou