

Predicting Information Seeker Satisfaction in Community Question Answering

Yandong Liu, Jiang Bian, Eugene Agichtein

Speaker: Yi-Lin, Hsu

Advisor: DR. Koh, Jia-lin

SIGIR'08

Date: 2009/1/15

Outline

- INTRODUCTION
- THE ASKER SATISFACTION PROBLEM
- PREDICTING ASKER SATISFACTION
- EXPERIMENTAL SETUP
- EXPERIMENTAL RESULTS
- CONCLUSIONS

INTRODUCTION

Yahoo! | My Yahoo! | Mail | More ▾ Make Y! My Homepage New User? Sign Up | Sign In | Help

YAHOO! ANSWERS Search **WEB SEARCH**

ask.

Can't find it with search? Ask

Post Question

answer.

Share knowledge
Help others
Earn points

What people think of Answers
How does it work?

discover.

Featured Topic

See what people are saying in:
Holidays

Search for questions: Search Advanced My Profile

Ready to Participate?
Get Started!

Categories

- Best of Answers
- Arts & Humanities
- Beauty & Style
- Business & Finance
- Cars & Transportation
- Computers & Internet
- Consumer Electronics
- Dining Out
- Education & Reference
- Entertainment & Music
- Environment
- Family & Relationships
- Food & Drink
- Games & Recreation
- Health

Best of Answers

 **How do you make a veggie burger?**
11 ☆ Asked by [The E - Vegetarian & Vegan](#)

 This is a basic recipe to make a vegetarian burger. You can add cheese, herbs, chopped bell peppers, or sauteed mushrooms over the burger. Things You'll Need: 1 package (12 ounces) frozen...
[More >](#)

0 👍 0 🗨️ - Answer by [Kila](#)

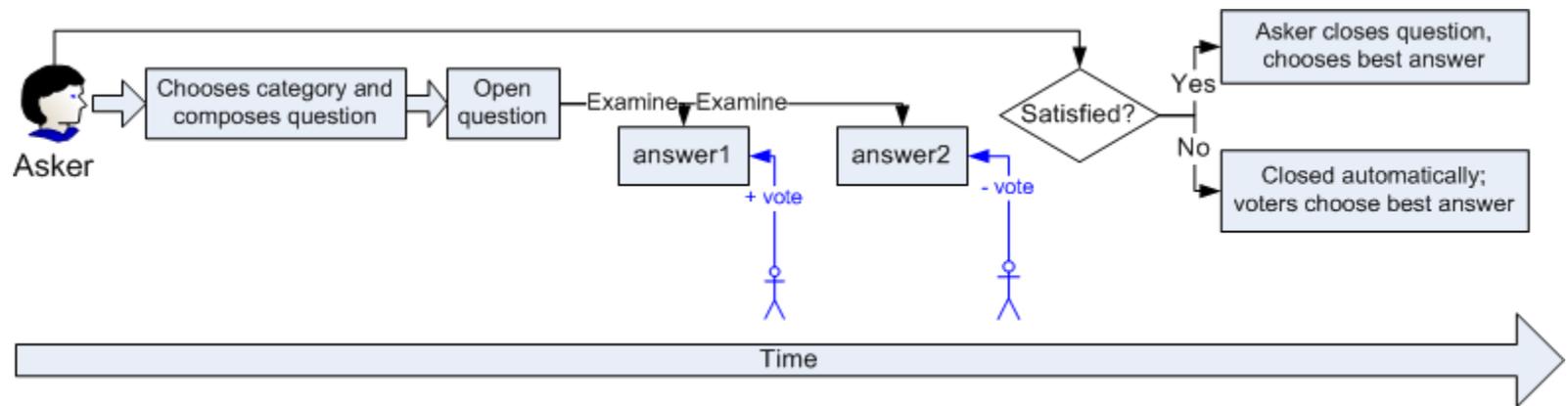
[More Best of Answers >](#) | [Send to a friend](#) ◀ || ▶

Answer Questions

RecentPopular

-  **Solving x with exponents?**
☆ In [Homework Help](#) - Asked by [/lalala/](#) - 2 seconds ago
-  **What is this song called?**
☆ In [Lyrics](#) - Asked by [Amber](#) - 2 seconds ago
-  **What did you have for lunch?**
☆ In [Other - Food & Drink](#) - Asked by [good answerer](#) - 3 seconds ago
-  **What is a shameful event in canadian history?**
☆ In [History](#) - Asked by [Anonymous <3](#) - 4 seconds ago

THE ASKER SATISFACTION PROBLEM



Intuition and Motivation

- Question Answering communities are both an important application by itself
- If many of the askers in CQA are not satisfied with their experience, they will not post new questions and will rely on other means of finding information.
- Predicting, understanding and monitoring asker satisfaction is at the core of maintaining an active and healthy QA community

Problem Statement

- **DEFINITION 1.** An asker in a QA community is considered satisfied iff: the asker personally has closed the question, selected the best answer, and provided a rating of at least 3 “stars” for the best answer quality. Otherwise, we define the asker to be unsatisfied.

Problem Statement

- **The Asker Satisfaction Problem:** Given a question submitted by an asker in CQA, predict whether the user will be satisfied with the answers contributed by the community.

Problem Statement

- the offline setting, where the posted question has already obtained some answers
- the online setting, where we attempt to predict immediately whether a user will be satisfied with the answers at some intermediate point in the process (e.g., while answers are still arriving), or even before any answers arrive

PREDICTING ASKER SATISFACTION

- Features
- Classification Algorithms

Features

- Question
- Question-Answer Relationship
- Asker User History
- Answerer User History
- Category Features
- Textual Features

Features

Feature (72 total)	Description
<i>Question (from 32 total)</i>	
Q: Subject length	Number of words in question subject
Q: Posting time	Time(in hours) of the day when the question was posted
Q: Number of answers	Number of answers received for this question
Q: Question stars	Number of stars received earned for this question
Q: Wh-type	Wh-word introducing the question title (e.g., “what”, “where”, etc.)
Q: Number of comments	Number of comments added by other participants
Q: Total positive/negative votes	Total number of positive/negative votes for the answers
Q: Average of positive/negative votes	Average number of positive/negative votes for each answer
Q: Max positive/negative votes	Max number of position/negative votes among all the answers
Q: Most_vote answer positive/negative votes	Number of positive/negative votes for the answer which received most votes
Q: QA Overlap	Words shared between question and the answer which received most votes
Q: Average past rating	Average rating given when closing questions in the past
Q: Most recent rating	Most recent rating given for last question
Q: Question KL-Divergence features	Question subject/content KL-Divergence values with Wikipedia/TREC/Y! Categories
Q: Question visual quality features	Question subject/content punctuation/typo/space density

Features

<i>Question-Answer Relationship (from 9 total)</i>	
QA: Most_vote answer content length	Number of words of the content from the answer which received most votes
QA: Most_vote answer's time difference	The time difference between the question and the answer which received most votes
QA: Top ten answers KL-Divergence features	Top ten answers content KL-Divergence values with Wikipedia/TREC/Y! Categories
QA: Top ten answers visual quality features	Top ten answers content punctuation/typo/space density
<i>Asker User History (4 total)</i>	
UH: Questions resolved	Number of questions resolved in the past
UH: Total answers received	Number of all answers this user received in the past
UH: Member since	How long since last registration
UH: Answer/Question ratio	Ratio of Answers to Questions posted
<i>Answerer User History (from 21 total)</i>	
AH: Total points received	Total points received from all the answerers/max answerer/most_vote answerer
AH: Questions resolved	Number of questions resolved from all the answerers/max answerer/most_vote answerer
AH: Total answers received	Number of total answers received from all the answerers/max answerer/most_vote answerer
AH: Best answers received	Number of best answerers received from all the answerers/max answerer/most_vote answerer
AH: Member since	How long since last registration from the all answerers/max answerer/most_vote answerer
AH: Best answer ratio	Ratio of best answers over all answers from all the answerers/max answerer/most_vote answerer
AH: Answer question ratio	Ratio of answers over questions from all the answerers/max answerer/most_vote answerer
<i>Category Features (6 total)</i>	
CA: Average time to close a question	Average interval between opening and closing for that category
CA: Average answers per question	Average number of answers per question for that category
CA: Average asker rating	Average rating given by asker for an answer from that category
CA: Average voter rating	Average votes given by voters for an answer from that category
CA: Average number of questions per hour	Average number of questions per hours from that category
CA: Average number of answers per hour	Average number of answers per hours from that category

Classification Algorithms

- Decision Trees(C4.5,RandomForest)
- SVM
- Boosting
- Naive Bayes

Evaluation Metrics

- Precision : the fraction of the predicted satisfied asker information needs that were indeed rated satisfactory by the asker.
- Recall : the fraction of all rated satisfied questions that were correctly identified by the system
- F1 : the geometric mean of Precision and Recall measures, computed as $\frac{2PR}{P+R}$
- Accuracy: the overall fraction of instances classified correctly into the proper class

Human Judgements

<i>Rater Group</i>	<i>Redundancy</i>	<i>Agreement</i>
Experts	3	0.82
Mechanical Turk Workers	5	0.9

Table 2: Ratings for 130 questions (54 satisfied/76 unsatisfied)

Human Judgements

The screenshot shows a web browser window with the URL <https://www.mturk.com/mturk/welcome>. The page features a navigation bar with tabs for 'Your Account', 'HITS', and 'Qualifications'. Below the navigation bar, there is a main heading 'Mechanical Turk is a marketplace for work.' followed by a sub-heading 'We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.' and a call to action '36,953 HITS available. [View them now.](#)'

The page is divided into two main sections:

- Make Money by working on HITs**
HITs - *Human Intelligence Tasks* - are individual tasks that you work on. [Find HITs now.](#)
As a Mechanical Turk Worker you:
 - Can work from home
 - Choose your own work hours
 - Get paid for doing good work

The process flow is: Find an interesting task (represented by a gear icon) → Work (represented by a gear icon) → Earn money (represented by a dollar sign icon).
- Get Results from Mechanical Turk Workers**
Ask workers to complete HITs - *Human Intelligence Tasks* - and get results using Mechanical Turk. [Get started.](#)
As a Mechanical Turk Requester you:
 - Have access to a global, on-demand, 24 x 7 workforce
 - Get thousands of HITs completed in minutes
 - Pay only when you're satisfied with the results

The process flow is: Fund your account (represented by a stack of money icon) → Load your tasks (represented by a document icon) → Get results (represented by a star icon).

The Windows taskbar at the bottom shows the Start button and several open applications: Google, Windows Explorer, Microsoft Office Word, and a PDF file named 'qa.pdf'. The system clock shows the time as 10:36 AM.

Dataset

- Our data is based on a snapshot of Yahoo! Answers
- In the early 2008
- Categories :
- “Health”, “Education & Reference”, “Sports”, “Science and Mathematics” and the “Arts”.
- we selected a random subset of 5,000 questions from the most recent 10,000 questions in the snapshot above.

Dataset

Questions	Answers	Askers	Categories	Satisfied (%)
216,170	1,963,615	158,515	100	50.7

Table 3: Statistics of the complete data crawled from the Yahoo! Answers site.

Dataset

Category	Questions	Answers	Answers per Question	Freq.	Satisfied	Time to close	Closed by Asker		Closed by Voters	
							Asker rating	Time to close	Voter rating	Time to close
Mathematics	651	2,329	3.58	13.0%	44.5%	3 days 20 hours	4.48	33 minutes	1.76	6 days
Diet & Fitness	450	2,436	5.41	9.0%	68.4%	2 days 17 hours	4.30	1.5 days	4.46	6 days
Women's Health	277	1,824	6.58	5.5%	62.8%	2 days 23 hours	4.28	35 minutes	1.98	6 days
Chemistry	236	508	2.15	4.7%	37.3%	4 days 7 hours	4.39	1 day 13 hours	1.19	6 days
Biology	176	589	3.35	3.5%	34.1%	4 days 5 hours	4.06	28 minutes	1.33	6 days
Books & Authors	161	645	4.01	3.2%	42.2%	4 days 6 hours	4.35	1 day 20 hours	2.13	6 days
Football (American)	152	1,722	11.33	3.0%	55.3%	3 days 11 hours	4.29	1 day 13 hours	2.05	6 days
Mental Health	151	1,159	7.68	3.0%	70.9%	2 days 16 hours	4.30	1 day 13 hours	1.32	6 days
Physics	149	428	2.87	3.0%	48.3%	3 days 13 hours	4.29	35 minutes	1.48	6 days
General Health	135	737	5.46	2.7%	70.4%	2 days 17 hours	4.49	1 day 13 hours	1.31	6 days
<i>Cumulative (10 Cat.)</i>	2,538	12,377	4.88	50.8%	53.4%		4.32		1.90	
<i>Overall (90 Cat.)</i>	5,000	25,063	5.01	100%	50.7%	3 days 15 hours	4.32	1 day 12 hours	1.87	6 days

Table 4: Selected statistics for the top 10 most popular categories in our dataset (together comprising 51% of questions in dataset).

Predicting Asker Satisfaction

- Human
- Heuristic
- Baseline
- ASP_SVM
- ASP_RandomForest
- ASP_C4.5
- ASP_Boosting
- ASP_NB

Predicting Asker Satisfaction

Classifier	With text		Without Text		Selected Features	
	F1	Accuracy	F1	Accuracy	F1	Accuracy
ASP_SVM	0.69	0.70	0.72	0.73	0.62	0.70
ASP_C4.5	0.75	0.74	0.76	0.75	0.77	0.77
ASP_RandomForest	0.70	0.67	0.74	0.73	0.68	0.68
ASP_Boosting	0.67	0.72	0.67	0.72	0.67	0.72
ASP_NB	0.61	0.63	0.65	0.68	0.58	0.67
<i>Human</i>	0.61	0.48				
<i>Baseline</i>	0.66	0.51				

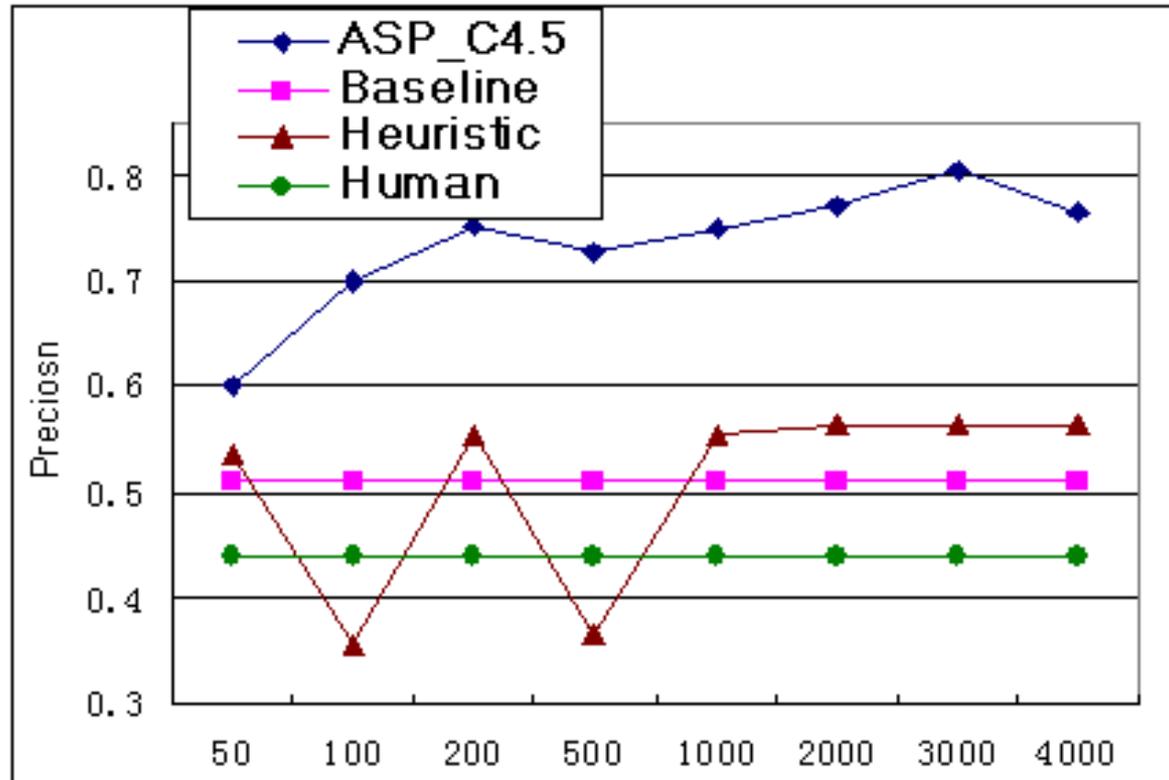
Table 6: Accuracy of ASP_SVM, ASP_C4.5, ASP_RandomForest, ASP_Boosting, and ASP_NB for varying parameters (5-fold cross validation).

Predicting Asker Satisfaction

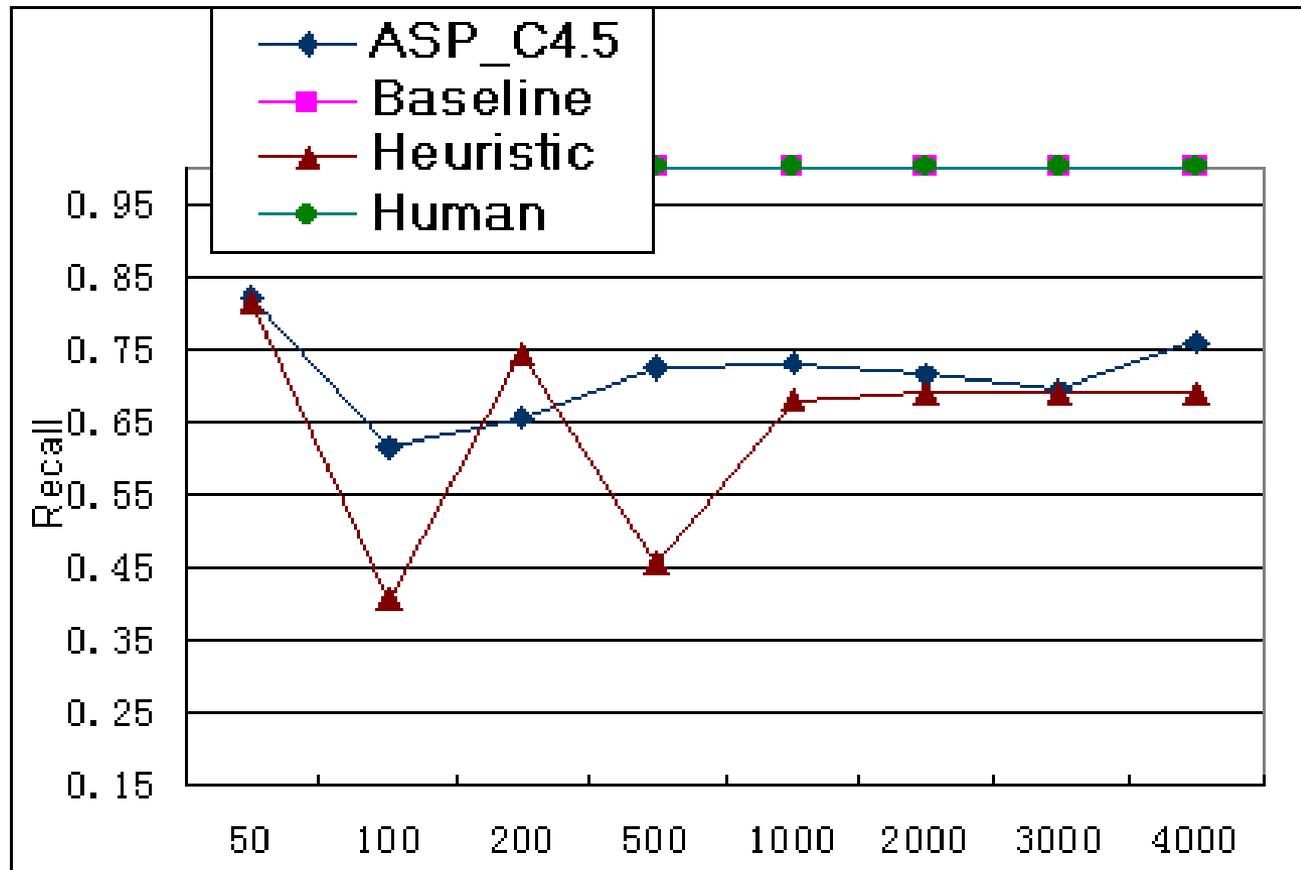
IG	Feature
0.14219	Q: Askers' previous rating
0.13965	Q: Average past rating by asker
0.10237	UH: Member since (interval)
0.04878	UH: Average number of answers attracted by past questions
0.04878	UH: Previous questions resolved for the asker
0.04381	CA: Average asker rating for the category (i.e., category prior)
0.04306	UH: Total number of answers received
0.03274	CA: Average voter rating
0.03159	Q: Question posting time
0.02840	CA: Average answers per question for the category
0.02633	AH: Answerer with most positive votes: member since (interval)
0.02080	AH: The highest best answer ratio for any answerer
0.02046	AH: The average best answer ratio of all answerer
0.01747	CA: Average number of answers per hour for the category
0.01531	QA: KL-Divergence of the top ten answers LM from Wikipedia LM

Table 7: Top 15 features with Highest Information Gain (IG)

Predicting Asker Satisfaction



Predicting Asker Satisfaction



Analysis and Discussion

- Online vs. Offline Prediction

	Precision	Recall	F1
On-line	0.78	0.70	0.74
Off-line	0.78	0.76	0.77

Table 8: On-line vs. off-line prediction of satisfaction

Analysis and Discussion

- Feature Ablation

	Precision	Recall	F1
Selected features	0.80	0.73	0.77
No question features	0.68	0.72	0.70
No question-answer features	0.76	0.74	0.75
No asker features	0.72	0.69	0.71
No answerer features	0.76	0.75	0.75
No category features	0.75	0.76	0.75

Table 9: Prediction accuracy with feature ablation.

Analysis and Discussion

- Textual Features

IG	Feature
0.003734	"i don't" in question
0.003335	"i was" in question
0.003147	"i have" in question
0.002595	"you are" in answer
0.002581	"to your" in answer
0.002543	"to get" in question
0.002536	"that i" in question
0.002532	"and i" in question
0.00238	"a few" in answer
0.002342	"but i" in question

Table 10: Textual features with high Information Gain (IG)

Analysis and Discussion

- Asker Satisfaction Varying with Past Experience

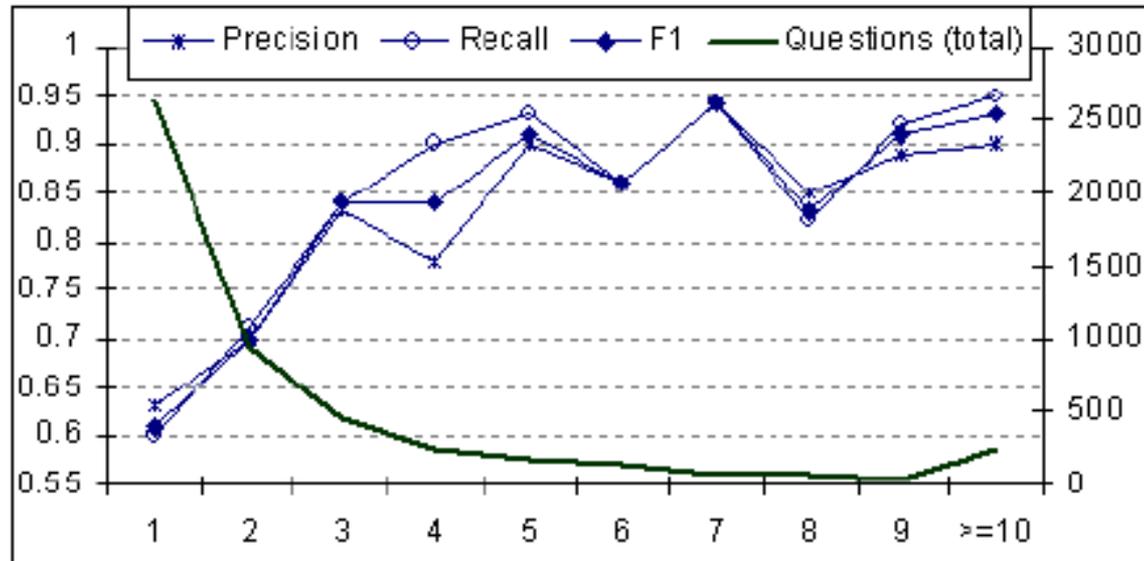


Figure 3: Satisfaction prediction accuracy for groups of askers with varying number of posted questions, and the corresponding number of questions posted by askers in each group.

CONCLUSIONS

- We have shown the importance of asker history to this highly personal, difficult, and subjective task, and demonstrated that our system can outperform human assessors who do not benefit from knowing the prior asker history.

CONCLUSIONS

- Our work opens a promising direction towards modeling user intent, expectations, and satisfaction, and can potentially result in practical improvements to the quality of question answering communities.