



Predicting Information Seeker Satisfaction in Community Question Answering

Yandong Liu, Jiang Bian and Eugene Agichtein
Proceedings of SIGIR2008

April 13, 2010

Yi ZHU



Outline

Introduction

CQA

Motivation

Problem

Model

Classifiers

Features

Experiment

Setting

Result

Conclusion



Community Question Answering

Resolved Question

[Show me another »](#)

How to get rid of a tattoo?

I don't like my tattoo anymore, are there any DIY ways to get rid of tattoos?

5 hours ago

 Report Abuse

Best Answer - Chosen by Asker

Buy an exfoliating soap... after exfoliating the area use a SKIN LIGHTENER.... Repeat as necessary

4 hours ago

   Report Abuse

Asker's Rating: *****

thanks for all the answers, laser sounds too expensive and painful :(I really hope the exfoliating method works, I'll definitely give it a shot!

Figure: Yahoo! answer



Lifecycle of a Question

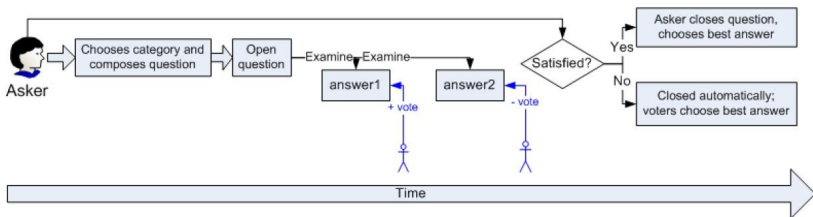


Figure: Lifecycle for a question posted in the CQA site



Motivation

- **Predict the satisfaction of the information seeker(asker);**
- **Find the benefits and drawbacks of the CQA site;**
- **Potential application include: answer ranking, user intent inference and query suggestion.**



Problem define

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



Asker Satisfaction Prediction system

- **Standard classification framework;**
- **Feature selection;**



Classifiers

- **Decision Tree:** C4.5, RandomForest
- **SVM**
- **Boosting:** AdaBoost
- **Naive Bayes**



Features

- **Question** (wh-type, length of subject and detail)
- **Question-Answer Relationship** (number of candidate answers, overlap)
- **Asker User History** (the satisfaction to previous questions)
- **Answerer User History** (number of answers provided, number of the best answers)
- **Category Features** (average asking rating)
- **Textual Features** (text of question and answer)



Features

Feature (72 total)	Description	Available at Question Time?
<i>Question (from 32 total)</i>		
Q: Subject length	Number of words in question subject	yes
Q: Posting time	Time(in hours) of the day when the question was posted	yes
Q: Number of answers	Number of answers received for this question	no
Q: Question stars	Number of stars received earned for this question	no
Q: Wh-type	Wh-word introducing the question title (e.g., "what", "where", etc.)	yes
Q: Number of comments	Number of comments added by other participants	no
Q: Total positive/negative votes	Total number of positive/negative votes for the answers	no
Q: Average of positive/negative votes	Average number of positive/negative votes for each answer	no
Q: Max positive/negative votes	Max number of position/negative votes among all the answers	no
Q: Most_vote answer positive/negative votes	Number of positive/negative votes for the answer which received most votes	no
Q: QA Overlap	Words shared between question and the answer which received most votes	no
Q: Average past rating	Average rating given when closing questions in the past	yes
Q: Most recent rating	Most recent rating given for last question	yes
Q: Question KL-Divergence features	Question subject/content KL-Divergence values with Wikipedia/TREC/Y! Categories	yes
Q: Question visual quality features	Question subject/content punctuation/typo/space density	yes
<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	no
QA: Most_vote answer's time difference	The time difference between the question and the answer which received most votes	no
QA: Top ten answers KL-Divergence features	Top ten answers content KL-Divergence values with Wikipedia/TREC/Y! Categories	no
QA: Top ten answers visual quality features	Top ten answers content punctuation/typo/space density	no
<i>Asker User History (4 total)</i>		
UH: Questions resolved	Number of questions resolved in the past	yes
UH: Total answers received	Number of all answers this user received in the past	yes
UH: Member since	How long since last registration	yes
UH: Answer/Question ratio	Ratio of Answers to Questions posted	yes
<i>Answerer User History (from 21 total)</i>		
AH: Total points received	Total points received from all the answerers/max answerer/most_vote answerer	no
AH: Questions resolved	Number of questions resolved from all the answerers/max answerer/most_vote answerer	no
AH: Total answers received	Number of total answers received from all the answerers/max answerer/most_vote answerer	no
AH: Best answers received	Number of best answerers received from all the answerers/max answerer/most_vote answerer	no
AH: Member since	How long since last registration from the all answerers/max answerer/most_vote answerer	no
AH: Best answer ratio	Ratio of best answers over all answers from all the answerers/max answerer/most_vote answerer	no
AH: Answer question ratio	Ratio of answers over questions from all the answerers/max answerer/most_vote answerer	no
<i>Category Features (6 total)</i>		
CA: Average time to close a question	Average interval between opening and closing for that category	yes
CA: Average answers per question	Average number of answers per question for that category	yes
CA: Average asker rating	Average rating given by asker for an answer from that category	yes
CA: Average voter rating	Average votes given by voters for an answer from that category	yes
CA: Average number of questions per hour	Average number of questions per hours from that category	yes
CA: Average number of answers per hour	Average number of answers per hours from that category	yes



Data set

- **Question:** 216,170
- **Answers:** 1,963,615
- **Askers:** 158,515
- **Category:** 100
- **Satisfied:** 50.7%



Data set(Sample)

- **Question:** 5,000
- **Answers:** 25,063
- **Category:** 90
- **Satisfied:** 50.7%



Method Compared

- **Human** (Mechanical Turk, majority vote)
- **Heuristic** (A question with many answers)
- **Baseline** (Majority class)
- **ASP-SVM**
- **ASP-RandomForest**
- **ASP-C4.5**
- **ASP-Boosting**
- **ASP-NB**



MTurk

<i>Rater group</i>	<i>Precision</i>	<i>Recall</i>	<i>F1</i>	<i>Accuracy</i>
<i>Expert (strict)</i>	0.36	0.68	0.47	0.45
<i>Casual (majority=3/5)</i>	0.43	1.0	0.60	0.47
<i>Casual (majority=4/5)</i>	0.44	1.0	0.61	0.48
<i>Casual (majority=5/5)</i>	0.41	0.75	0.53	0.46

Figure: Comparing casual human raters (Mechanical Turk Workers) with expert raters (130 randomly sampled questions)

Result

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				

Figure: Accuracy of ASP-SVM, ASP-C4.5, ASP-RandomForest, ASP-Boosting, and ASP-NB for varying parameters (5-fold cross validation).



Result

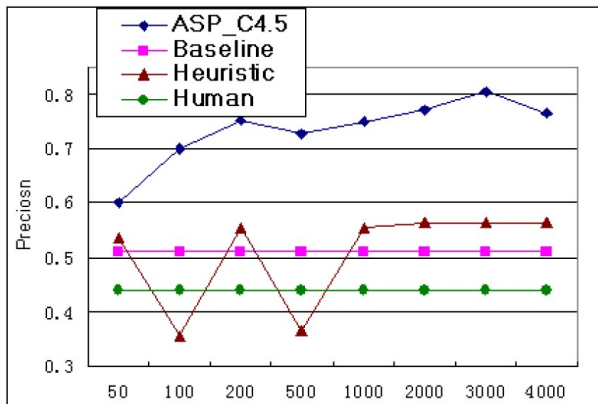


Figure: Precision of ASP, Human, Baseline, and Heuristic for varying amount of training data.



Feature

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

Figure: Top 15 features with Highest Information Gain (IG)



Online V.S. Offline

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

Figure: On-line vs. off-line prediction of satisfaction



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

Figure: Prediction accuracy with feature ablation.



Conclusion

- **Introduce the problem of predict asker satisfaction;**
- **Prediction framework for offline and online setting;**
- **Experimental feature selection.**



Thanks!
Q & A