Exploring Question Subjectivity Prediction in Community QA

Baoli Li[1] **Yandong Liu**[1] **Ashwin Ram**[2] **Ernest V. Garcia**[1] **Eugene Agichtein**[1]

[1]Emory University
{baoli.li,yliu49,euhevg,eugene.agichtein}@emory.edu

[2]Georgia Institute of Technology
ashwin@cc.gatech.edu

ABSTRACT
In this paper we begin to investigate how to *automatically* determine the subjectivity orientation of questions posted by real users in community question answering (CQA) portals. Subjective questions seek answers containing private states, such as personal opinion and experience. In contrast, objective questions request objective, verifiable information, often with support from reliable sources. Knowing the question orientation would be helpful not only for evaluating answers provided by users, but also for guiding the CQA engine to process questions more intelligently. Our experiments on Yahoo! Answers data show that our method exhibits promising performance.

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: On-line Information Services – *Web-based services.*

General Terms
Experimentation

Keywords
Subjectivity Analysis, Question Classification

1. INTRODUCTION
Although much progress has been made in automatic Question Answering (QA), answering complex and realistic questions in open domain automatically is still beyond the state-of-the-art of automatic QA systems. This is one of the reasons that caused the emerging popularity of Community Question Answering (CQA) services – which allow users to post questions for other users to answer. Yahoo! Answers (http://answers.yahoo.com) now dominates the CQA market with hundreds of millions of answers posted by millions of participants on thousands of topics.

The reason for existence of these services is that people are usually better at interpreting and answering questions than automatic QA systems or general-purpose web search engines. Unfortunately, QA communities are not perfect, and due to a variety of factors such as incentives, abuse, or vandalism the provided answers may not be what the asker requested. Therefore, it is increasingly important to identify high-quality and responsive answers to questions. One approach is to automatically evaluate the quality of answers (and questions) provided by users of community QA systems. Another approach may be to automatically estimate "correctness" of the provided answers by trying to verify them against reliable sources. Our approach is complementary to the above: our goal is to understand the intent behind the question posted, which can then be used to infer whether proposed answers match the intent of the question. We focus on one important aspect of intent detection: *subjectivity analysis*. We attempt to predict whether a question is subjective or objective. Objective questions are expected to be answered with reliable or authoritative information, typically published online and possibly referenced as part of the answer, whereas subjective questions seek answers containing private states such as personal opinion, judgment, and experience. Consider two questions crawled from Yahoo! Answers:

> Question 1. what's the difference between chemotherapy and radiation treatments? (*objective*)
>
> Question 2. 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? (*subjective*)

As the examples above illustrate, objective and subjective questions require different types of answers. If we could automatically predict the orientation of a question, we would be able to better identify appropriate answers for the question. For objective questions, we could try to find a few highly relevant articles as references, whereas for subjective questions, most (if not all) of answers are not expected to be found in authoritative sources. Furthermore, question orientation information may be useful for evaluating user-provided answers. For example, we could rank or filter answers based on whether an answer matches the question orientation. Finally, learning how to identify question orientation is a crucial component of inferring user intent, a long-standing problem in web search. While previous work exists on differentiating objective and subjective contents for answer extraction [1, 2], ours is the first study on automatically predicting question subjectivity in CQA.

Compared to the questions in traditional QA research (e.g. TREC), questions asked by web users are prone to be ill-formatted (e.g., word capitalization may be incorrect or missing, or consecutive words may be concatenated), ungrammatical, and include common online idioms (e.g., using "u" to mean "you" and "2" to mean "to"). These properties make online questions more difficult to be analyzed with current NLP techniques, even for the basic step of tokenization. Moreover, one or more answers may be available for these questions, and might be used to help subjectivity analysis, as good answers are assumed to have the same subjectivity as questions. We explore what question and answer features could be most helpful for identifying subjectivity orientation of real questions. Our

SIGIR'08, July 20-24, 2008, Singapore.
ACM 978-1-60558-164-4/08/07.

experiments on Yahoo!Answers data show that our method exhibits promising performance.

2. EXPERIMENTSAL RESULTS

Supervised learning algorithms are dominant in both question classification [3] and sentiment analysis [4]. We follow this practice and focus on finding more effective features for our target scenario. Features that we consider are listed in Table 1. We use case-insensitive features and character n-grams to overcome spelling errors and poor formatting, and POS features to attempt to capture simple grammatical patterns. We experimented with three term weighting schemes: binary, term frequency (TF), and TF*IDF, and chose TF as the most effective. As questions are often accompanied by one or more answers, we also explore ways to obtain good performance by considering both questions and answers. As the classifier we use a robust SVM implementation, LIBSVM (http://www.csie.ntu.edu.tw/~cjlin/libsvm), with linear kernel.

Table 1: Features used to predict question subjectivity

Feature type	*Description*
Char 3-gram	Character tri-grams, lower-cased
Word	Word features, lower-cased
Word+Char 3-gram	Include both Word and Char 3-gram features
Word-n-gram	Word uni-, bi-, and tri-grams: W_i, W_iW_{i+1}, $W_iW_{i+1}W_{i+2}$
Word POS n-gram	Mix of word and POS tri-grams, e.g. W_i, W_iW_{i+1}, W_iPOS_{i+1}, POS_iW_{i+1}, $W_iW_{i+1}W_{i+2}$, $W_iPOS_{i+1}W_{i+2}$, $W_iW_{i+1}POS_{i+2}$, $POS_iW_{i+1}W_{i+2}$, $W_iPOS_{i+1}POS_{i+2}$, $POS_iW_{i+1}POS_{i+2}$, $POS_iPOS_{i+1}W_{i+2}$

Dataset: We created a labeled dataset consisting of 978 resolved questions randomly chosen from Yahoo! Answers under the following 5 top-level categories: Arts, Education, Health, Science, and Sports. For annotation, we employed Amazon's Mechanical Turk service (http://www.mturk.com). Each question was annotated by 5 Mechanical Turk workers, and we derived the final annotation by majority strategy with our judgment on marginal cases. The overall average percentage agreement between Mechanical Turk workers and the final annotation is 0.795. 646 questions (about 66%) are subjective, which indicates that CQA users tend to ask subjective questions.

Metrics: we use the macro-averaged F-1 (the average of F-1 for predicting both the subjective and objective classes). A naïve baseline that always picks the majority *subjective* class would result in F-1 value of 0.392.

Table 2 reports our experimental results. The first five rows show results using the text of the question (**question**), the text of the best answer (**best_ans**), the text of all the answers (**all_ans**), the text of both the question and the best answer (**q+bestans**), and the text of the question with all the answers (**q+allans**), respectively. Interestingly, the text of the best answer itself is not as effective as the text of the question, nor is using the text of all of the answers. One possible reason is that many best answers (about 40%) are chosen by the community, and not the asker him/herself, are hence do not necessarily represent the asker's intent. With character 3-gram, our system achieves performance comparable with word as feature, but combining them together does not improve performance. We observe a slight gain with more complicated features, e.g. word and POS n-gram, but the gain is not worth the increased time and space complexity.

Table 2: Accuracy of subjectivity prediction for varying feature sets (Macro-averaged F-1, 5-fold cross validation).

Feature set / Unit	**Char 3-gram**	**Word**	**Word+ Char 3-gram**	**Word n-gram (n<=3)**	**Word POS n-gram (n<=3)**
question	0.700	0.704	0.694	0.716	0.720
best_ans	0.587	0.583	0.578	0.580	0.565
all_ans	0.603	0.623	0.607	0.648	0.630
q+bestans	0.681	0.672	0.662	0.687	0.712
q+allans	0.679	0.678	0.676	0.708	0.689
q+bestans(S)	0.711	0.705	0.713	0.728	0.722
q+allans(S)	0.727	0.705	0.722	0.737	**0.742**

While combining question text with answer text directly does not improve performance (rows 4 and 5), we do get a small, but significant gain when incorporating the answer text in separate feature spaces (rows 6 and 7), where the same term from question and answer is treated as different features. We conjecture that separately modeling distributions of terms in questions and answers is helpful, as we plan to explore this further in future work.

3. CONCLUSION AND FUTURE WORK

We introduce the problem of *automatically* identifying subjectivity orientation of questions in QA communities, and explore a supervised machine learning solution with different features designed for this task. Our experiments demonstrate that case-insensitive character 3-gram feature is simple yet effective representation for this task, and that exploiting the answers provided by other users for a question indeed can improve the prediction accuracy. Our method is significantly more accurate than a naïve baseline, and will serve as a good starting point for future work. In the future we plan to explore semi-supervised learning methods, and explore further how to more effectively use the available answers to derive a more powerful classifier.

4. REFERENCES

[1] Yu, H., and Hatzivassiloglou, V. Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. In *proceedings of EMNLP,* 2003.

[2] Somasundaran, S., Wilson, T., Wiebe, J., and Stoyanov, V. QA with Attitude: Exploiting Opinion Type Analysis for Improving Question Answering in On-line Discussions and the News. In *proceedings of ICWSM,* 2007.

[3] Zhang, D., and Lee, W.S. Question Classification Using Support Vector Machines. In *proceedings of SIGIR,* 2003.

[4] Pang, B., and Lee, L. A Sentimental Education: Sentiment Analysis Using Subjective Summarization Based on Minimum Cuts. In *proceedings of ACL,* 2004.