The Synergy Between Voting and Acceptance of Answers on StackOverflow, or the Lack thereof

Neelamadhav Gantayat, Pankaj Dhoolia, Rohan Padhye, Senthil Mani and Vibha Singhal Sinha
IBM Research
{neelamadhav, pdhoolia, ropadhye, sentmani, vibha.sinha}@in.ibm.com

***Abstract*—StackOverflow's primary goal is to serve as a platform for users to solicit answers regarding programming questions, though its archives are often used by other users who face similar issues and thus it serves a secondary purpose of documenting common problems. The two driving mechanisms for filtering out low quality posts and highlighting the best answers are community votes and the mark of acceptance by the original question asker. But does the asker's choice always match the popular vote? If so, is the asker's choice influenced by the community vote or is the community vote biased towards the accepted answer? And if the asker and community disagree, then can we determine any particular characteristics of posts that influence the choice of the asker and community differently, such as its size, readability, presence of code snippets and external links as well as similarity to the original question? In this paper, we explore the answers to these questions by studying a data-set of all posts on StackOverflow from its launch in September 2008 to September 2014.**

I. Introduction

StackOverflow.com is one of the most active Q&A forums on the World Wide Web today. Up to September 2014, there were about 8 million questions asked on this forum, with about 7 million of these having at least one reply. This activity involves some 3.4 million users, which is a large community by any standard.

User participation in StackOverflow is encouraged by providing incentives in the form of reputation points for good questions and answers. The *goodness* of a post for this purpose is determined by up-votes it has received from registered community members; about 5.26 million questions (66%) have some voting activity on their answers. Additionally, question askers can also mark one of the received replies as an *accepted answer*, and are encouraged to do so. About 4.58 million questions (57%) have an accepted answer; 2.43 million of these had multiple answers.

Many questions on StackOverflow continue to be voted-on and even receive additional answers long after the original asker has accepted an answer and perhaps even stopped using the forum. Thus, many archived StackOverflow posts serve as community-endorsed *de-facto* documentation for common issues that may help future answer-seekers.

An interesting question to ask, therefore, is whether the accepted answer for any question is the same as the answer that has received the highest community votes? Of the 2.43 million questions which have both multiple replies as well as an accepted answer, it appears that in about almost 2 million (81%) of the cases, the accepted answer is also the reply with most number of votes for that question. This is a very high association that motivates the first two research questions explored in this paper, about whether community votes influence accepted answers or vice versa.

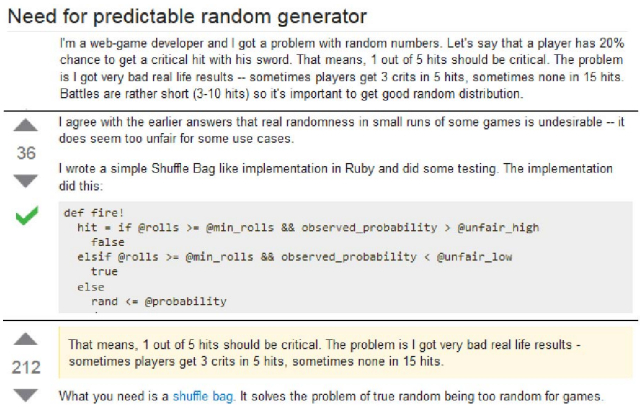

Fig. 1. Example StackOverflow post where the accepted answer differs from the top-voted answer.

While the correlation between acceptance and community votes is certainly high, it is also interesting to consider the other 12% of the cases where the accepted answer differs from the top-voted answer today. Figure 1 shows a snapshot of such a post on StackOverflow. If this *better* answer, as deemed by public opinion, also existed at the time that the asker accepted a different answer, what could have been the reason for this? Our third research question explores these cases by looking at different features such as the size of the answer, its similarity to the question, readability of text, presence of external links and code snippets.

To summarize, we address three research questions in this paper by analyzing a data-set of all posts on StackOverflow up to September 2014 [3]. The research questions are:

1) When faced with multiple answers, are question askers influenced by existing community votes while accepting one answer?
2) Does the act of accepting an answer bias future readers into up-voting that answer even if another comparable answer exists?
3) Are there any distinctive characteristics about answers that a question asker accepts for themselves and those that are top-voted by the public?

978-0-7695-5594-2/15 $31.00 © 2015 IEEE
DOI 10.1109/MSR.2015.50

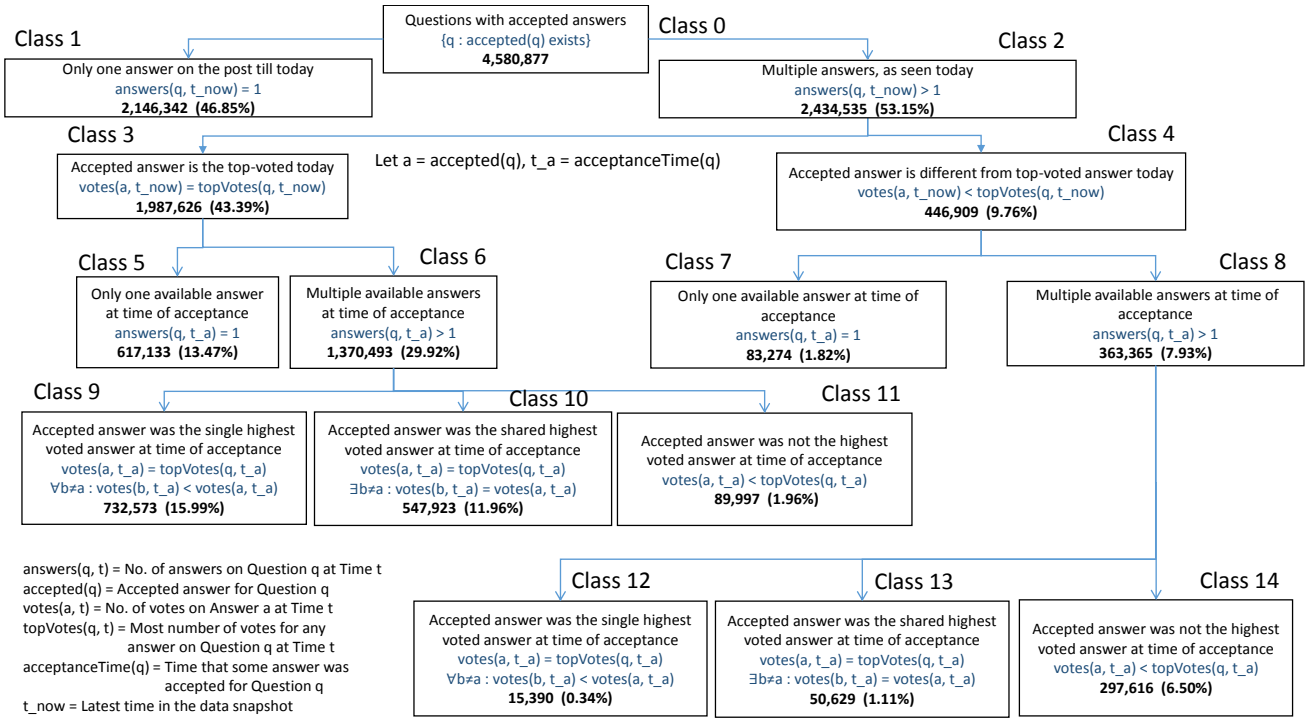

Fig. 2. Classification of StackOverflow questions based on the nature of accepted and top-voted answers.

The paper is organized as follows. Section II analyzes the influence of community votes on a user's choice of accepted answers. Section III investigates whether community votes themselves are influenced by the act of a user marking an answer as accepted. Section IV considers the cases where public opinion differs from the asker, and attempts to find general reasons to explain these cases. We summarize our findings in Section V.

II. Influence of Community Votes on Asker Acceptance

Figure 2 depicts a classification of StackOverflow questions having accepted answers, based on the relationship between their accepted answers and top-voted answers. In the rest of this paper we refer to different sub-sets of the StackOverflow questions by using the labels on nodes in this hierarchy.

The first study is motivated by the fact that of these questions that have multiple answers today[1], (Class 2), about 81% share the property that the highest-voted answer is the same as the accepted answer (Class 3). This creates a situation of an undisputed *best* answer for a particular question, since both the question asker as well as the community of registered users agree on the same post. But did the asker and the community reach this conclusion independently or did one influence the other?

RQ1: When faced with multiple answers, are question askers influenced by existing community votes while accepting one answer?

To answer the first research question, we compute, for all questions on StackOverflow with an accepted answer, the number of votes on every answer posted for this question that it received on or before the day of acceptance[2].

A large portion of these cases (62.14%) constitutes the situation in which only one answer is available (Class 1, 5, 7), and the asker accepts it. In about 16.33% of the cases, the asker picks the single highest-voted answer (Class 9, 12), and in 13.07% of the cases, the accepted answer shares the top-voted spot with some other answer (Class 10, 13). In about 8.46% of the questions, the asker accepts an answer despite there being another answer having more votes available at the time (Class 11, 14).

Thus, when presented with multiple replies, a user has picked the highest voted answer thrice as many times as some other answer, which indicates that the correlation between acceptance and community votes is consistent at the time of acceptance as well. However, this does not necessarily indicate causality, since there may have just been one good answer among the multiple replies. This brings us to consider the next research question, which considers the reverse influence of acceptance on community votes, which we *can* measure.

[1]We use the term "today" loosely to refer to the latest point in time in our data snap-shot, which corresponds to September 30, 2014

[2]Since the data-set does not contain a time-of-day component, we are constrained by the granularity of calendar days.

III. Influence of Asker Acceptance on Community Votes

The second research question is motivated by the fact that of the questions in which the accepted answer has more votes than other answers today (Class 3), in about 30% of the cases there was a comparable or better voted answer available at the time of acceptance (Class 10, 11).

> RQ2: Does the act of accepting an answer bias future readers into up-voting that answer even if another comparable answer exists?

Since this question concerns the *change* of votes on replies to a question after an answer was accepted for it, it cannot be evaluated by looking at the size of classes in Figure 2 alone. For example, if this bias turns out to exist, it is possible that many questions in Class 4 as observed today are moving to Class 3 in the future.

We thus attempt to determine this bias by analyzing all questions where there was a higher-voted answer at the time of acceptance (Class 11, 14), and apply an independent t-test on the difference in votes on accepted and comparable answers before and after the time of acceptance.

However, as our data-set contains all activity up to September 2014, the time elapsed after acceptance would vary based on whether the question was posted close to September 2014 or several years before that. Hence, we restrict our analysis to questions posted before September 1, 2013 and for each question consider two time-windows: (1) the time period between posting of the question (t_q) and acceptance of an answer (t_a), say $T_1 = [t_q, t_a]$, (2) the time period of one year after the acceptance of an answer on the question, $T_2 = [t_a, t_a + 1yr]$. The final data-set consisted of 220,550 questions having 220,550 accepted and 236,898 comparable answers.

The two-sample t-test was applied to the difference in votes across both groups and found that the gain is significant for the *accepted* group with t-value of -127.36 and p-value $< 2.2 \times 10^{-16}$.

We may thus infer that the act of accepting an answer influences the community to vote for that answer in the future, even though it was not clearly the most popular answer before acceptance. This is perhaps one explanation for the high correlation between accepted and top-voted answers in the entire data-set.

IV. Disagreement between Asker and Community

The final research question is concerned with the 12% of cases (Class 4) where the accepted answer is different from the community top-voted answer.

In 114,331 of these cases (about 25%), the top-voted answer as observed today was actually posted after the acceptance of an answer by the asker. This is intuitive because answers to many questions about programming practices, libraries or API usage can change over time, as the languages and libraries themselves evolve. The original asker may not even be active on StackOverflow to accept the new answer, but the community can most certainly increase its visibility by up-voting it.

In the remaining 332,578 cases where both the accepted answer and the current top-voted answer were available to the user at the time of acceptance, only in 10,056 cases was the accepted answer having higher votes at the time of acceptance. In a majority of the cases, the current top-voted answer had equal or more votes than the current accepted answer even at the time of acceptance by the asker. What, then, led the asker to choose a different answer than what public opinion might find better?

> RQ3: Are there any distinctive characteristics about answers that are preferred by question asker and the community?

We first describe five different characteristics of a post that we considered may be influential in this split of opinions. These characteristics are partly inspired by previous work [4], [5] on the quality of answers in Q&A sites.

1) *Size of the post (in words)*: The larger the answer, the more likely it is to contain lots of minute details. We chose this feature with the intuition that question askers might prefer direct, concise to-the-point answers that can quickly solve their issue at hand, while the community might prefer it to be more elaborate and self contained with all necessary details, so that any reader of that post can understand it without any prior context.
2) *Size of code snippets*: Code snippets provide examples or pseudo-code that may improve the quality of an answer, and may be especially useful to the asker if they are looking for a quick-fix solution to an immediate problem. We calculated the size (word count) of block-level `<code>` snippets in each post.
3) *Number of external links*: References to documentation or other sources of information prove are surely useful to the asker, but may be more valued by future readers since the information sources themselves might be changing over time.
4) *Readability*: The language, grammar and structure of non-code text in an answer might influence either or both the asker and the community. We used the Flesch-Kincaid score [1] (English only) to compute readability of each post.
5) *Similarity*: Question askers often like answers that are tailored to the context of the question [4], and hence we computed a similarity score for every answer with respect to the text in the question using a cosine similarity index [2].

We addressed the research question by analyzing the 322,522 questions where at the time of acceptance, the user had a choice between multiple replies, and chose to accept an answer that had equal or lesser votes than the answer which ultimately has the highest number of votes as observed today. This class is a subset of the union of Class 13 and 14, which

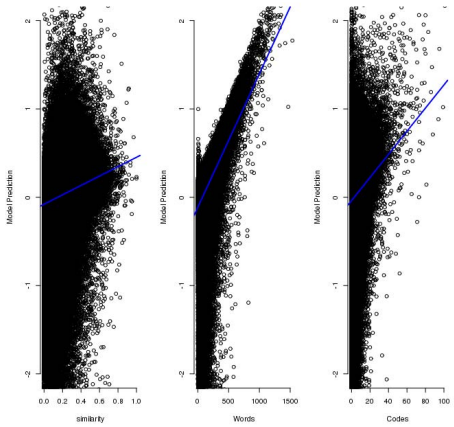

Fig. 3. Significant Predictors of Acceptance

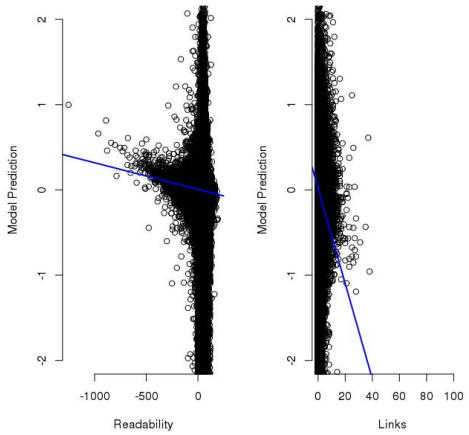

Fig. 4. Significant Predictors of Top Voted

would consider *all* other answers that may have more votes than the accepted answer at the time of acceptance, even if some third answer is the top-voted today. The reason for using this sub-set to address this research question is that we are interested in those cases where the community seems to have consistently disagreed with the asker right from when the asker had the consider all replies all the way up to the current point in time.

We ran a binary logistic regression to infer how the characteristics of the post influenced the asker's acceptance and popularity of that answer within the community. We constructed a model with the five independent variables for size of post, size of code blocks, number of external links, readability and similarity. The dependent variable in our model captured the asker's decision about the acceptance as value 1 and a value of 0 if it is top-voted by the community (these are mutually exclusive classes in our filtered data-set).

The resulting model had an accuracy of 56.24% and indicates that all five parameters significantly influence its prediction. Figure 3 shows the prediction of the model on factors that were positively correlated with acceptance: similarity, size of text and code snippets. Figure 4 shows the prediction of the model on factors that were negatively correlated with acceptance (equivalently, positively correlated with community voted, since in our selected sample, they are mutually exclusive): readability and number of links. Each factor's influence was significant with p-values < 0.05.

As it turns out, the similarity of an answer's text with respect to the question and the presence of code snippets in it influences the asker's choice more while the readability of text and presence of external links influences third-party readers, which is consistent with our intuition. Surprisingly, the size of the answer was correlated with acceptance indicating that the community votes up short concise answers more often than the original question asker.

V. Conclusion

The two determining factors of the quality of answers to a question on StackOverflow are the number of votes it received from registered community members and whether or not the asker has accepted it. In this paper, we have studied the synergy between the act of acceptance of an answer, and voting by the community.

We observed that in a large fraction (81%) of questions that have multiple answers, the accepted answer is also the top-voted answer. When faced with more than one reply, and given some answers marked with up votes, askers accept the answer with the most votes 77.65% of the time, and thus the number of votes on a post may be an influencing factor on their choice of acceptance. Interestingly, we observed that the act of acceptance of an answer influences the community to vote for that answer instead of another answer that had comparable or more number of votes before the asker had accepted any reply.

We also studied the cases in which the asker's choice and the public opinion disagreed consistently. We considered several features of a post that contribute to its quality, as asserted by previous work, such as size, code snippets, external links, readability and similarity to the question text. We found that in the cases where the accepted answer is not the top-voted answer, the asker is biased towards longer posts that use similar terms and include code snippets, while community votes lean towards short, concise answers that include external links and have a better readability score.

References

[1] J. Kincaid, R. Fishburne, R. Rogers, and B. Chissom, "Derivation of new readability formulas (automated readability index, fog count, and flesch reading ease formula) for navy enlisted personnel," *Research Branch Report 8-75. Chief of Naval Technical Training: Naval Air Station Memphis.*, 1975.

[2] A. Singhal, "Modern information retrieval: a brief overview," *Bulletin of the IEEE Computer Society Technical Committee on Data Engineering*, vol. 24(4), pp. 35–43, 2001.

[3] A. T. T. Ying, "Mining challenge 2015: Comparing and combining different information sources on the stack overflow data set," in *The 12th Working Conference on Mining Software Repositories*, 2015, p. to appear.

[4] S. Nasehi, J. Sillito, F. Maurer, and C. Burns, "What makes a good code example?: A study of programming Q&A in StackOverflow," in *Software Maintenance (ICSM), 2012 28th IEEE International Conference on*, Sept 2012, pp. 25–34.

[5] Z. Zhu, D. Bernhard, and I. Gurevych, "A multidimensional model for assessing the quality of answers in social Q&A sites." in *Technical Report TUD-CS-2009-0158. Technische Universitt Darmstad.*, 2009.