Research Proposal: Predicting Community Engagement with Questions Across Online Question-Answer Fora

Candidate Number: 10140, Seminar Group 6, Word Count: 3990 words

29 April 2019

Contents

1 Introduction

Formulating high-quality questions and obtaining answers to these questions is not only a key part of the learning and information gathering processes, but to scientific and human progress. Online social networks and fora have given users unprecedented ability to comment on and question the world, and while attention has been devoted to finding the right answers (just ask Google), relatively little attention has been given to how we can ask better questions. Producing relevant, legible and well-researched questions concerning technical concepts is especially valuable in online social question-answer communities where expert resources are often scarce.

One way to improve the functioning of these communities would be to provide an indication to users of how "well" their questions will be received by a community, potentially nudging questioners to improve aspects of their question before they add demand to a question-answering system. The goal of this analysis therefore is to scrutinise work aimed at questions in question-answering communities, and build on this by constructing and validating a selection of community engagement metrics and consequent models. Unlike most previous research, I will extend my analysis to a large range of diverse communities for further robust validation.

2 Research Question

Put explicitly, the goal of this research is to answer the following question:

To what extent can community engagement of question-answering communities in response to heterogeneous questions be captured by various metrics and accurately predicted?

Here, my conceptualisation of community engagement is a community interacting with user questions through answering, voting, commenting, favouriting, editing, closing down etc. I define positive engagement as constructive, amiable activity across these interactions. Questions are also heterogeneous in the sense that their construction and different levels of "quality" evoke varying levels of positive and negative community interaction.

The research question above necessitates two tasks: Capturing, and then predicting community engagement. The first task can be addressed by considering a number of response variables that relate to, or are proxies for community engagement. The second task involves using these response variables in a predictive classification model to ascertain how well each forum's model can predict "good" (positive) and "bad" (negative) community engagement.

While this research question fits into how communities react to and value questions, the focus is on validating the methodological measurement of community engagement quantitatively, rather than qualitative, causal or correlational explanatory analysis. It is hoped that the methodology validated in this analysis will be used as a stepping stone for further research to more precisely explore how and why diverse communities value and engage with questions differently.

3 Previous Work

3.1 Question-Answer Communities

A substantial amount of research has investigated question-answer communities. Work has been done on quality of answers (Jeon *et al.*, 2006; Shah and Pomerantz, 2010; Tian, Zhang and Li, 2013), questioner satisfaction (Liu, Bian and Agichtein, 2008) and behaviour of expert users (Riahi *et al.*, 2012; Sung, Lee and Lee, 2013). Work on question-answer communities has also often been placed in the framework of a matching process where the goal is optimising the routing of user questions to expert resources (Li and King, 2010; Li, King and Lyu, 2011; Zhou, Lyu and King, 2012; Shah *et al.*, 2018), or routing questions according to answerers' interests as a recommendation system (Wu, Wang and Cheng, 2008; Qu *et al.*, 2009; Szpektor, Maarek and Pelleg, 2013).

My approach differs to this previous work in two main aspects. Firstly, rather than focusing on answer or user characteristics, I focus on questions. As Agichtein *et al.* (2008) show, question quality can significantly impact answer quality. Trivially, questions also serve as the initial event in the question-answering process, upon which all community engagement follows. Thus, maximising the positive community engagement of questions will most certainly enhance the functioning and development of a community.

The second difference is the framework in which I place the research. I steer clear of the systems-based framework of optimising matching/routing mechanisms and instead centralise how users can be nudged to improve aspects of their question before exerting demand on expert resources. This framework coincides strongly with a large body of research on codifying question quality in question-answer communities, and so this is discussed next.

3.2 Question Quality

Intuitively, question quality and community engagement are two sides of the same coin - high-quality questions assuredly lead to positive community engagement. As will be argued later, I consider community engagement a more robust characterisation of what is being measured in the following literature, however I will refer to "question quality"" rather than community engagement for the sake of this discussion.

Question quality has been investigated using the online question-answer community Yahoo! Answers (Agichtein *et al.*, 2008; Bian *et al.*, 2009; Li *et al.*, 2012), but a considerable limitation of this dataset is that it lacks objective and clear-cut metrics for question quality. To address this, Agichtein *et al.* (2008) use semantic features (punctuation, typos, lexical complexity etc.) to define quality questions, Bian *et al.* (2009) make use of manual labels to identify quality in a sample of 250 questions, and Li *et al.* (2012) use a combination of tag count, answer count and time-to-first-answer along with domain expert and author judgement to ascertain ground truth.

Fortunately, the StackExchange fora that I will explore are richer datasets in the sense that a greater number of proxies for question quality and community engagement are available. More importantly, these metrics are more objective as they are derived from the data itself rather than manual human labeling. In this way, it is also possible to "automatically" define question quality with a robust[1] response variable, vastly increasing the number of ground truth cases.

Another substantial difference that has evolved over time in the question-quality literature is the chosen predictive model. Previous work (Agichtein *et al.*, 2008; Bian *et al.*, 2009; Anderson *et al.*, 2012; Li *et al.*, 2012) chose to use features such as the question category, lexical properties of the question (length, level of misspelling, words-per-sentence) and question-asker's reputation for prediction.

By ignoring words, topics and other inherent question content and focusing instead on features of the question-asker, users asking questions for the first time would be largely ignored by models of this kind. This is directly opposed to my specific goal of predicting how communities will engage with all users, especially new and inexperienced users who are likely to not have their questions answered at all, let alone positively. For my modeling task, I therefore only focus on explanatory features derived from the contents of questions, not user attributes, history or other data.

---

[1]The key issue being the assumptions made and the manner in which this labeling occurs

3.3 Topic Modeling

Focusing on the textual content of questions alone leads me to a discussion on how this text will be modeled. Bayesian models have gained popularity in solving diverse structured prediction tasks in Natural Language Processing (NLP) (Chiang *et al.*, 2010), and Latent Dirichlet Allocation (LDA) (Blei, Ng and Jordan, 2003) is a form of topic modeling where generative Bayesian models are applied to documents to model hidden topics as a probability distribution over words. LDA can thus be used to uncover the underlying semantic structure of documents and infer topics associated with these documents.

Indeed, Ravi *et al.* (2014) show that using latent topics derived from LDA modeling allows for powerful prediction (accuracy scores of up to 72%) of "question quality" using the coding question-answer site, StackOverflow. My analysis will mirror not only the modeling methodology of Ravi *et al.* (2014), but also critique and build on the methodology they use to define their metric for question quality.

The final predictive model that Ravi *et al.* (2014) employ is based on a model proposed by Allamanis and Sutton (2013), who themselves analysed StackOverflow questions but did not use their model to predict "question quality". Allamanis and Sutton (2013) use LDA models on three tiers: First across the entire question body, secondly on the code chunks in the body, and lastly on the body with noun phrases removed.

Ravi *et al.* (2014) on the other hand choose the following three levels over which to model latent topics: Globally to capture topics over the entire question content, locally to capture sentence-level topics, and a Mallows model (Fligner and Verducci, 1986) global topic structure to enforce structural constraints on the sentence topics in questions. This is the same model that I will employ in task two of my research question.

4 Data

4.1 StackExchange Fora

The data I will use are the 10 largest community question-answering fora from the family of websites, StackExchange, of which StackOverflow analysed by Ravi *et al.* (2014) is the largest (10 million registered users and 18 million questions at time of writing). There are over 170 different StackExchange sites ranging from vegetarianism to robotics to Russian computer coding and all the data are publicly available for download in compressed XML format at archive.org.

The following files are available for each site:

- `PostHistory.xml`: A history of the versions of each question and answer posted (questions can be edited)
- `Posts.xml`: The final, up to date version of each question and answer posted
- `Users.xml`: Data on registered users
- `Votes.xml`: Data on different votes cast (moderator-related, offensiveness, spam, votes to close, delete etc.)
- `Comments.xml`: The final, up to date version of each comment posted
- `Badges.xml`: Log of when badges (an incentive mechanism) are awarded to users for specific achievements
- `PostLinks.xml`: Connects posts marked as linked or duplicated
- `Tags.xml`: Data on the tags put forth with each question posted

For my analysis I will only use the `Posts.xml` data for each community, and these are the variables recorded per post in this file:

- `Id`: A chronological post identity variable
- `PostTypeId`: Mainly an indicator of whether the post is an question (=1) or an answer (=2)
- `ParentId`: An indicator of which question an answer belongs to (only answers)
- `AcceptedAnswerId`: An indicator of which answer the original question poster selected as accepted (only questions)

- `CreationDate`: A date variable relating to when the post was initially posted

- `Score`: The difference between up-votes and down-votes for posts

- `ViewCount`: The number of times that a post has been viewed by registered and non-registered users alike

- `Body`: The main post content

- `OwnerUserId`: An identity variable for the post owner

- `LastEditorUserId`: An identity number for the last registered user that edited the post

- `LastEditDate`: A date variable relating to when the post was last edited

- `LastActivityDate`: A date variable relating to when last there was activity on the post

- `Title`: The post title (questions only)

- `Tags`: The collection of tags linked to a question

- `AnswerCount`: The number of answers that a question receives (questions only)

- `CommentCount`: The number of comments that a post receives

- `FavoriteCount`: The number of times that users have favourited a question (questions only)

- `ClosedDate`: A date variable relating to if the question was closed (questions only)

Note: The word "post" above refers to both questions and answers.

4.2 Noteworthy Aspects

In order to develop a deeper understanding of the data, a brief discussion of the functioning of StackExchange websites is worthwhile. Questions on StackExchange fora are publicly available for anyone on the internet to view, however in order to ask a question on a specific forum one has to register an account with that forum using an email address and display name. Once a user is registered, they start off with 1 reputation[2]. Most importantly for my analysis, having 15 reputation or more allows you to "up-vote" a post, a level of 50 or more allows you to comment on questions and answers, at or above 125 gives you the ability to "down-vote" a post, and finally reputation of 2000 allows you to edit any question or answer.

---

[2]https://meta.stackexchange.com/questions/7237/how-does-reputation-work

There are a number of methodological complexities created by the structure of the sites. Firstly, since all questions are publicly viewable, not everyone who views a question (and thus adds to the `ViewCount`) is a registered user and thus has the ability to vote. The differing reputation levels for up- and down-voting privileges (15 and 125 respectively) also lead to the `Score` variable being highly negatively skewed (more high-scored questions and answers). Lastly, the fact that questions can be edited by the original posters and by users with 2000 reputation complicates most of the questioner-community engagement since there is no data on if answers, comments, votes, views etc. occur before or after question edits.

5   Methodology

5.1   The `Score` Variable and Question Quality

In addressing the first task of the research question, I examine a number of response variables relating to community engagement. Ravi *et al.* (2014) choose to use the `Score` variable as an indicator of question quality, however I believe that a question's `Score` is better characterised as a proxy for community engagement since there are a number of confounding factors regarding this variable.

One factor is that there seems to be a less-than-full consensus of when exactly to up- or down-vote[3] despite general guidelines on StackExchange sites stating that up-votes should be given if a question shows prior research, is clear and useful, and down-voting the opposite. As mentioned in section 4.2, another potential issue with `Score` is its negatively skewed distribution - up-votes for questions can be orders of magnitude higher than down-votes, making it appear as if there are many more "good" questions on StackExchange sites versus "bad" ones. There is also the likely possibility that the `Score` variable is subject to bias via "rep-whoring"[4].

Lastly, I am of the opinion that question quality is a highly subjective issue, with different communities desiring different aspects of a question (closed-end for hard sciences or promoting-discussion in social sciences for example), hence why I choose to view the `Score` variable as a measure of community engagement instead of question quality. I further believe that this is not the only proxy for community engagement evident in the data, and so that it was I investigate next.

---

[3]https://meta.stackexchange.com/questions/12772/should-i-upvote-bad-questions
[4]A derogatory term describing users' chasing reputation points by any means necessary

5.2   Additional Response Variables

I propose 3 additional response variables to capture positive community engagement for questions:

- The number of times a post is edited (conditional on non-closed down), `EditCount`

- The sentiment of the answers, `AnswerSentiment`

- The questioner has accepted an answer (conditional on positive `Score`), `AcceptedAnswer`

My main assumption for the first variable, `EditCount`, is that the number of times a post is edited is inversely proportional to its positive community engagement, since more work is needed by the community to get the question into an answerable state. My central assumption for `AnswerSentiment` is that positive- and neutral-sentiment answers indicate positive community engagement as amicable responses to a question, and my main assumption for `AcceptedAnswer` is that a questioner accepting an answer is not only an indication that the community has responded to their question, but also indicates the questioner has had their request for information met (i.e. positive community engagement and questioner fulfillment).

One important finding from Ravi *et al.* (2014) is that highly-viewed questions are more likely to receive a higher `Score` (at least on the StackOverflow forum), and so they address this by considering the quantity `Score`/`ViewCount` to eliminate a "popularity-bias". Since `EditCount` is the only variable that would be affected by a popularity bias, I too normalise it by dividing by `ViewCount`. A last adjustment that Ravi *et al.* (2014) make is to only consider questions above a certain `ViewCount` threshold (1000 in their case for StackOverflow) to eliminate the chance that questions are not viewed by users qualified to vote (or edit), and I replicate this for all response variables for the same reason.

5.3 Final Binary Response Variables

The last step for the "capturing" task of the research question is to code the response candidates into binary variables (i.e. "good" questions versus "bad"[5]). This step is a simplification that ignores the fact that community engagement is continuous rather than discrete, however it serves the purposes of the classification model employed later and it is left to further research to extend the analysis to consider a range of community engagement.

In order to convert the `Score`/`ViewCount` variable to a binary response, Ravi *et al.* (2014) appear to decide arbitrarily on a threshold of 0.001 (the average for `Score`/`ViewCount` being 0.002), resulting in approximately 60% of questions being classified as "good questions". I believe that this is a crucial step that shouldn't be overlooked, and thus needs to be validated before I too decide on a threshold to convert `EditCount`/`ViewCount` and `AnswerSentiment` into binary variables (`AcceptedAnswer` is already a binary response variable).

5.4 Validation

In order to make the split of good and bad questions as robust as possible, I plan to validate the splitting across all four response variables. This will be done by computing linguistic features of the good/bad split to ascertain the degree of differentiation between samples. The assumption here is of course that linguistic features represent a strong way in which questions that receive positive community engagement differ from those that don't. Furthermore, this also allows me to validate the thresholds of `Score`/`ViewCount`, `EditCount`/`ViewCount` and `AnswerSentiment` by uncovering which thresholds maximise the differentiation between the the two sets of questions.

Linguistic features that I will examine include the following:

- Textual Euclidean Distance, Cosine Similarity and Jacquard Similarity
- Grammatical and semantic features such punctuation, URL-text, stop-word usage, number of words
- Readability and lexical complexity
- Dictionary methods to examine sentiment, appeal to moral values, emotion association etc.

---

[5]To be clear, when I refer to good and bad questions for my analysis, I am referring to questions attracting positive and negative community engagement respectively

The differences between the above linguistic features across good and bad questions will also be tested for statistical significance using non-parametric difference-of-means tests such as the Kruskal-Wallis and Wilcoxon signed-rank test.

5.5   Modeling

Having completed task one of my research question and ascertained robustly defined response variables, task two involves building a classification model to predict on these response variables. The prediction problem is linear classification regression with a binary dependent variable, and as previously mentioned the only explanatory variables are sourced from the user input when asking a question in a community (i.e. a question's `Title` and `Body`).

Here I precisely replicate the methodology of Ravi *et al.* (2014) by using the three-tiered LDA model discussed in section 3.3, balancing the sample of good and bad questions and obtaining prediction results. While the model is complex and word constraints limit my detailing of the underlying mechanics, the work done forming robust good/bad question samples ensures that the predictions of this model accurately relate to community engagement, given the assumptions already discussed.

Implementing this model on a range of large, diverse StackExchange sites including StackOverflow itself, I will be able to compare how prediction accuracies vary across the four response variables, how they vary across communities and how they compare to the 72% accuracy attained by Ravi *et al.* (2014).

6 Predicted Data Outcomes

For the first task of capturing community engagement through four different responses, the outcomes are a set of linguistic differences (with statistical significance) between good and bad questions for each response variable. Furthermore, the optimal thresholds of where positive and negative community engagement is defined for `Score`/`ViewCount`, `EditCount`/`ViewCount` and `AnswerSentiment` will be ascertained to improve robustness in this step of the methodology.

Subsequently, employing the predictive model in task two of the research question will yield predictive accuracy results for the robust response variables across a multitude of large, diverse communities. I believe that comparing these accuracies to the 72% attained in Ravi *et al.* (2014), comparing across response variables and comparing across fora will yield invaluable insight into A) how well this methodology is able to predict positive community engagement and B) if and how diverse communities engage with questions differently.

7 Limitations, Impact and Further Research

7.1 Limitations

One limitation regarding my methodology concerns the proposed `AcceptedAnswer` variable. It is quite possible that questioners may find utility or even a solution to a question posed, yet neglect to make the effort to formally select which answer solved their question on the forum. This would bias the number of formally solved questions down and potentially confound good and bad questions. Furthermore, answers being posted as comments[6] and vice-versa is a common occurrence (at least on StackOverflow) and this too would confound the predictive results for this variable.

An assumption to mitigate these issues with the `AcceptedAnswer` variable would be that these anomalies are minimal and not biased to occur on specific posts - thus with being small, even and randomly distributed across the data it would not have a significant effect on the results.

A major concern is the fact that questions are editable[7], not only by the original poster but by anyone with 2000 reputation. Editing guidelines include fixing grammar and spelling, clarifying, correcting minor mistakes or adding related resources and links - the consequence of this being that users could vote, comment and answer on essentially different questions as a question is continually edited. This would no doubt confound the prediction of community engagement for all four response variables, thus a decision would need to be made on whether to use the original text of questions across all models, or to use the final edited version of questions.

Since the editing issue is particularly problematic for the `EditCount` response variable where it wouldn't make sense to link the final edited text and the number of edits for prediction (thus eliminating any explanatory features in the `Body` text that specifically led to the edits), I would choose to use the original text of questions across all response variables and models. The issue would now be that a substantial amount of views, votes and answers might occur after a question has been edited multiple times, thus it would be assumed that the majority of these events occur before edits take place. A temporal of analysis of average waiting times for answers, views and votes would be beneficial to support or disprove this assumption.

---

[6]https://meta.stackexchange.com/questions/17447/answer-or-comment-whats-the-etiquette
[7]https://stackoverflow.com/help/privileges/edit

7.2 Impact and Further Research

As previously discussed, there is definitely a spectrum of community engagement rather than two good and bad classes. Predicting more than just "good" or "bad" would most definitely assist in the formulation of more constructive questions for users, thus this is one recommended area for further research. Another suggestion is to investigate the issue of question-editing by examining the average wait times for events like answers, votes and views to occur.

There is also a large amount of data on the StackExchange fora that hasn't been explored, so there may yet be stronger response variable candidates that would increase the robustness of the analysis even further. While this analysis should be highly reliable, valid and generalisable owing to the robust methodology constructed, it still does not address causal or even correlational relationships. It is therefore strongly recommended that further research address the *how* and *why* of community engagement on question-answer communities, as opposed to just the *what* that was explored in this analysis.

To my knowledge, no previous research has used the methodology here proposed in the respective framework to capture and measure community engagement. At the very least, this research will improve on the extent to which we can capture and predict positive community engagement from online question-answering communities, and yield insight into how homogeneously engagement lies across diverse communities with different subject matters. If it holds that we can accurately predict positive community engagement on a strong, robust and well-validated response variable, then there is room to implement this model on any collection of question-answering sites, including Massive Open Online Courses, to improve the efficacy and productiveness of these communities.

References

Agichtein, E. *et al.* (2008) 'Finding high-quality content in social media', in *Proceedings of the 2008 international conference on web search and data mining.* ACM, pp. 183–194. doi: 10.1145/1341531.1341557.

Allamanis, M. and Sutton, C. (2013) 'Why, when, and what: Analyzing stack overflow questions by topic, type, and code', in *2013 10th working conference on mining software repositories (msr).* IEEE, pp. 53–56. doi: 10.1109/MSR.2013.6624004.

Anderson, A. *et al.* (2012) 'Discovering value from community activity on focused question answering sites: a case study of stack overflow', in *Proceedings of the 18th acm sigkdd international conference on knowledge discovery and data mining.* ACM, pp. 850–858. Available at: http://dl.acm.org/citation.cfm?id=2339665.

Bian, J. *et al.* (2009) 'Learning to recognize reliable users and content in social media with coupled mutual reinforcement', in *Proceedings of the 18th international conference on world wide web.* ACM, pp. 51–60. doi: 10.1145/1526709.1526717.

Blei, D. M., Ng, A. Y. and Jordan, M. I. (2003) 'Latent Dirichlet Allocation', *Journal of Machine Learning Research*, 3, pp. 993–1022.

Chiang, D. *et al.* (2010) 'Bayesian Inference for Finite-State Transducers', in *Human language technologies: The 2010 annual conference of the north american chapter of the association for computational linguistics.* Association for Computational Linguistics (June), pp. 447–455. Available at: http://www.isi.edu/{~}sravi/pubs/naacl2010{_}bayes-fst.pdf.

Fligner, M. and Verducci, J. S. (1986) 'Distance based ranking models', *Journal of the Royal Statistical Society: Series B (Methodological)*, 48(3), pp. 359–369.

Jeon, J. *et al.* (2006) 'A framework to predict the quality of answers with non-textual features', in *Proceedings of the 29th annual international acm sigir conference on research and development in information retrieval.* ACM, pp. 228–235. doi: 10.1145/1148170.1148212.

Li, B. and King, I. (2010) 'Routing questions to appropriate answerers in community question answering services', in *Proceedings of the 19th acm international conference on information and*

knowledge management. ACM, pp. 1585–1588. doi: [10.1145/1871437.1871678](10.1145/1871437.1871678).

Li, B. *et al.* (2012) 'Analyzing and predicting question quality in community question answering services', in *Proceedings of the 21st international conference on world wide web.* ACM, pp. 775–782. doi: [10.1145/2187980.2188200](10.1145/2187980.2188200).

Li, B., King, I. and Lyu, M. R. (2011) 'Question routing in community question answering', in *Proceedings of the 20th acm international conference on information and knowledge management.* ACM, pp. 2041–2044. doi: [10.1145/2063576.2063885](10.1145/2063576.2063885).

Liu, Y., Bian, J. and Agichtein, E. (2008) 'Predicting information seeker satisfaction in community question answering', in *Proceedings of the 31st annual international acm sigir conference on research and development in information retrieval.* ACM (Section 2), pp. 483–490. doi: [10.1145/1390334.1390417](10.1145/1390334.1390417).

Qu, M. *et al.* (2009) 'Probabilistic question recommendation for question answering communities', in *Proceedings of the 18th international conference on world wide web.* ACM (2), pp. 1229–1230. doi: [10.1145/1526709.1526942](10.1145/1526709.1526942).

Ravi, S. *et al.* (2014) 'Great Question! Question Quality in Community Q&A.', in *Eighth international aaai conference on weblogs and social media.* (1), pp. 426–435.

Riahi, F. *et al.* (2012) 'Finding expert users in community question answering', in *Proceedings of the 21st international conference on world wide web.* ACM, pp. 791–798. doi: [10.1145/2187980.2188202](10.1145/2187980.2188202).

Shah, C. and Pomerantz, J. (2010) 'Evaluating and predicting answer quality in community QA', in *Proceedings of the 33rd international acm sigir conference on research and development in information retrieval.* ACM (March 2008), pp. 411–418. doi: [10.1145/1835449.1835518](10.1145/1835449.1835518).

Shah, V. *et al.* (2018) 'Adaptive matching for expert systems with uncertain task types', in *2017 55th annual allerton conference on communication, control, and computing (allerton).* IEEE, pp. 753–760. doi: 10.1109/ALLERTON.2017.8262814.

Sung, J., Lee, J.-g. and Lee, U. (2013) 'Booming Up the Long Tails: Discovering Potentially Contributive Users in Community-Based Question Answering Services', in *Seventh international aaai conference on weblogs and social media*, pp. 602–610.

Szpektor, I., Maarek, Y. and Pelleg, D. (2013) 'When relevance is not enough: promoting diversity

and freshness in personalized question recommendation', in *Proceedings of the 22nd international conference on world wide web.* ACM, pp. 1249–1260.

Tian, Q., Zhang, P. and Li, B. (2013) 'Towards Predicting the Best Answers in Community-Based Question-Answering Services', in *Seventh international aaai conference on weblogs and social media*, pp. 725–728.

Wu, H., Wang, Y. and Cheng, X. (2008) 'Incremental probabilistic latent semantic analysis for automatic question recommendation', in *Proceedings of the 2008 acm conference on recommender systems.* ACM, p. 99. doi: 10.1145/1454008.1454026.

Zhou, T. C., Lyu, M. R. and King, I. (2012) 'A classification-based approach to question routing in community question answering', in *Proceedings of the 21st international conference on world wide web.* ACM, pp. 783–790. Available at: http://www2012.wwwconference.org/proceedings/companion/p783.pdf.