Analysis of Redditor Reliability

Kashev Dalmia[1], Ryan Freedman[1], and Terence Nip[1],
{dalmia3, rtfreed2, nip2}@illinois.edu
[1]University of Illinois at Urbana-Champaign

ABSTRACT
In this paper, we present a system by which to evaluate the reliability of the users of the popular social network Reddit. Though Reddit is many things to many people, increasingly, through the efforts of both the company running Reddit and the userbase itself, it is becoming a place where users come to read and discuss news. Thus, there is a growing need to evaluate the realibility of the suppliers of information on Reddit. We first collect features of reliable and unreliable users based on their contributions, and importantly, the reaction of the community to their contributions. We then use machine learning techniques to train a regression model to give a reliability score to an arbitrary user, with promising results.

1. INTRODUCTION
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Something something boston bombing [3].

Something something news doesn't always go to the top [1]

2. RELATED WORK
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Something something only a single subreddit [2]

3. DESIGN
Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

4. EVALUATION AND RESULTS
From our trained random forest regression model, we get a picture of the way that Redditors are. We get a glimpse of what useful, contributing Redditors look like, and what bad, non-contributing Redditors look like.

We collected data on around two thousand Redditors, and ran their data through our regressive model to get a reliability score $-1 \leq s_r \leq 1$. Then, we re-correlate this score with input features to intuitively see what features are important or not, and what features indicated useful and not-useful Redditors.

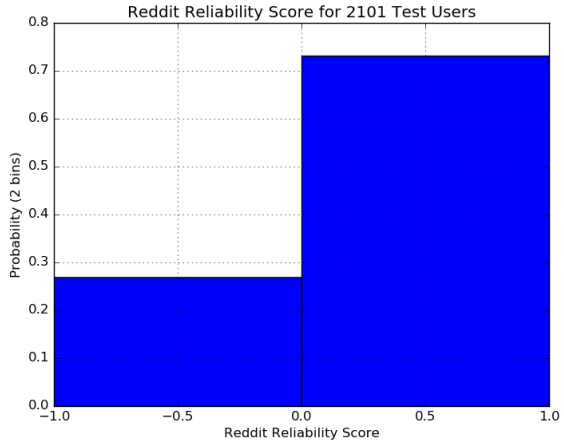

Figure 1: The distribution of the reliability score s_r of the sampled Redditors, binned into two bins.

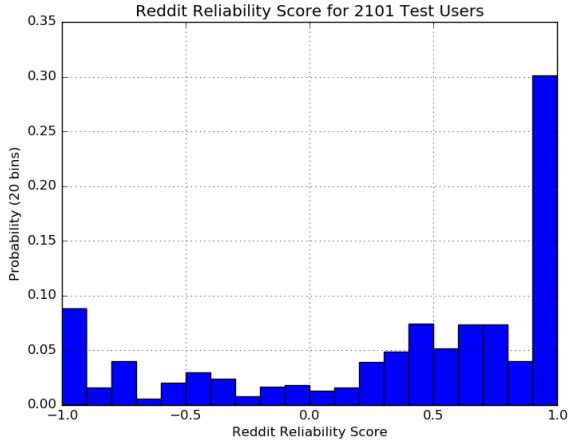

Figure 2: The distribution of the reliability score s_r of the sampled Redditors, binned into twenty bins.

TODO

5. CONCLUSION

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

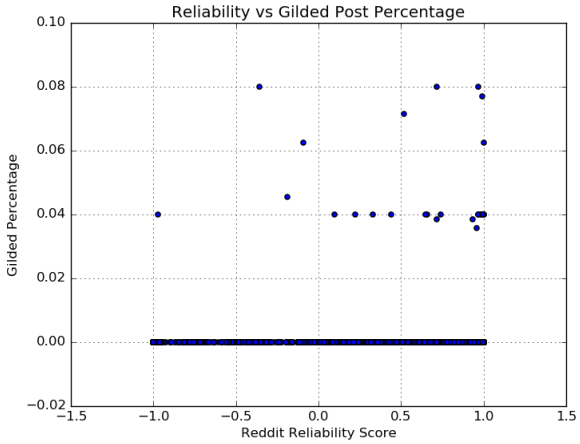

Figure 3: The reliability score s_r plotted against the percentage of gilded posts.

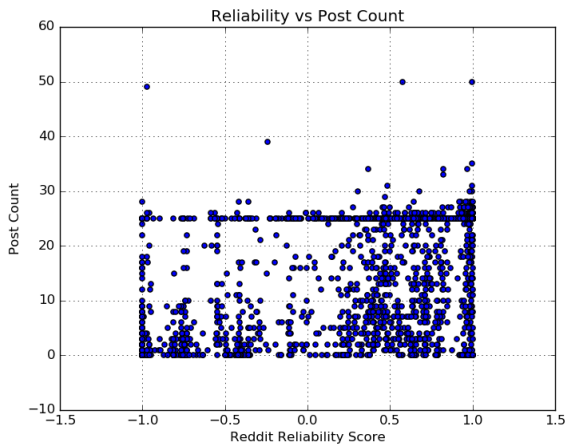

Figure 4: The reliability score s_r plotted against the number posts the Redditor has made.

6. ACKNOWLEDGMENTS

The authors would like to thank Professor Tarek F. Abdelzaher, of the University of Illinois at Urbana-Champaign, for his support in this project.

7. REFERENCES

[1] E. Gilbert. Widespread underprovision on reddit. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, pages 803–808, New York, NY, USA, 2013. ACM.

[2] A. Leavitt and J. A. Clark. Upvoting hurricane sandy: Event-based news production processes on a social news site. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '14, pages 1495–1504, New York, NY, USA, 2014. ACM.

[3] L. Potts and A. Harrison. Interfaces as rhetorical constructions: Reddit and 4chan during the boston marathon bombings. In *Proceedings of the 31st ACM International Conference on Design of Communication*,

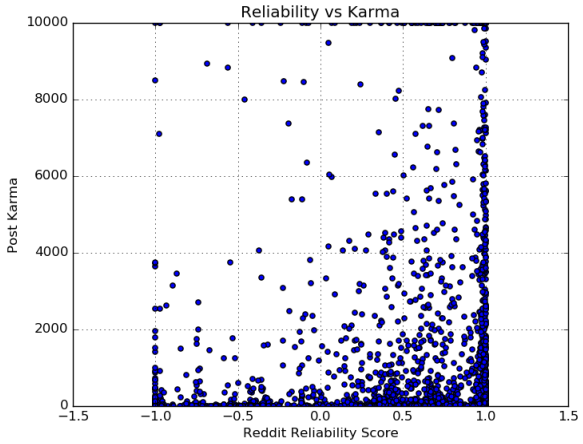

Figure 5: The reliability score s_r plotted against the average Karma per post the Redditor has made.

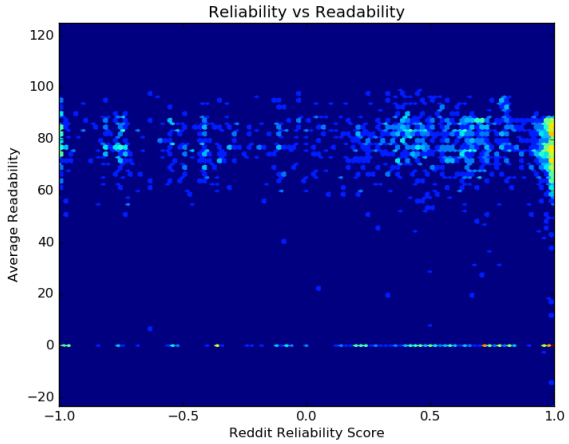

Figure 6: The reliability score s_r plotted against the Flesch–Kincaid readability of their comments.

SIGDOC '13, pages 143–150, New York, NY, USA, 2013. ACM.