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. In the past, reddit has had numerous identity issues. However, through the efforts of both 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 reliability of the suppliers of information on reddit. We first collect features of both reliable and unreliable users based on their contributions, and most importantly, the reaction of the community to their contributions. We then use machine learning techniques to train a regression model to assign a reliability score to an arbitrary reddit user.

1. INTRODUCTION

reddit, the self-proclaimed 'frontpage of the internet', is slowly becoming just that; many popular news sources, such as The New York Times and CNN, have begun to use reddit as a primary source of information in an effort to leverage the power of the large numbers of users which gather and try to provide and cross-validate information others have gathered.

In most cases, the information gathered is rather inoccuous. For example, a subset of reddit users (or *redditors*) gather information from popular news site and aggregate them on individual, topic-based forums (called *subreddits*) like `r/news` or `r/worldnews`, subreddits for current events within the United States and the world, respectively.

However, reddit's track record is not necessarily completely spotless when it comes to dispersing correct, cross-validated information. For example, during the Boston bombings of the Boston Marathon in April of 2013, reddit falsely accused individuals of having been part of the bombings simply because of circumstantial evidence. Many news sites, including CNN, picked up the story and publicized it. However, it turned out that the individuals reddit picked out were completely innocent and were only identified as being potentially involved because of circumstantial evidence (such as having been missing for long periods of time).[7].

reddit is also noteworthy for allowing users to *upvote* or *downvote* user-submitted content, indicating either agreement/interest or disagreement/disinterest, respectively, with the material at hand. As a side effect of this, users ultimately promote material that they find to be interesting to them, and hide material that they find to be out of scope, irrelevant, or simply not interesting with respect to the subreddit they're in.[2] Additionally, every upvote and downvote is connected with a measurement of the community's net happiness/satisfaction with the content submitted. A user gains *karma* for every upvote they receive, and loses karma for every downvote they get.

2. RELATED WORK

TODO

Something something only a single subreddit [6]

3. DESIGN

In this section, we show the design of our system. Additionally, we discuss some of the challenges, limitations, and design decisions that went into making it. Finally, we discuss some of the details of the implementation.

3.1 Gathering reddit Usernames

The first limitation of the reddit API is that usernames are an 'open secret'. If one has the username of an account, public information about that account can be retrieved, but there is no way to directly get usernames. Instead, what we were forced to do was scrape the posts of popular subreddits and get the usernames of the author of every post and comment. In doing so, we were able to collect over 150K usernames. Of the usernames we collected, we randomly selected around 3K to fully gather data on and run our regression model on.

One issue with this approach is that this makes it impossible to identify non- participants. If a user never comments on a post, or posts a post themselves, there is no way to know that that user exists. This is unfortunately an insurmountable limitation. Instead, we chose only to find good and bad users to train our classifier, and ignore non-participants.

3.2 reddit API Limits

reddit has an API limit of 30 requests per minute when one hasn't authorized their script properly, and 60 requests per minute when authorized. Given our timeframe, we eschewed the authorization method and decided to go with the less time-efficient method of not authorizing our script. In gathering data, however, we discovered this limit is not strictly enforced, but in order to be good citizens and as to not get our access revoked, we knew we had to design around these constraints. As such, in order to both speed up our ability to access user data, dynamically change the features that we used (see Section 3.4), and reduce load on reddit's servers, we used a MongoDB instance to cache the data we were gathering. Not were we able to store raw data from reddit API calls, we were also able to cache results from more computationally intensive features that we were generating.

3.3 Establishing Ground Truth

Establishing ground truth was done in two stages. The first stage was to find reliable users. Finding these users was relatively trivial as the site rewards positive behavior through karma, which leads to increased visibility. From there the users could be filtered by their level of contribution manually. Moderators from various communities were also taken for their work in helping the community. These users were used in our training set for a reliability score $s_r = 1$.

The second stage was to find users who were unreliable or detrimental to the community. Eventually, we discovered subreddits dedicated to weeding out users that didn't contribute and various posts that detailed accounts that were used to abuse the community. These were used as our training set for a reliability score $s_r = -1$.

3.4 Feature Selection

Before having a trained regression model, there was no way to tell what characteristics of reddit users would be important. However, one can't get this regression model without these features. So, we decided to pick a wide variety of features generated from the raw data, and figured out what was important later by looking at our results. Here, we present features which we suspected would be important. A discussion of what features were actually important is in Section 4.

3.4.1 reddit Karma and Derivatives

The most important and useful feature of the reddit platform is the fact that it has a built in voting system. We hypothesized that users whose posts and comments are well received would have positive correlation with that users reliability. However, not all content on reddit is news-worthy. There are many, many subreddits in which there is everything from pets to porn. Thus, we established several subcriteria in order to differentiate between karma earned by posting reliable news and karma earned by posting picture of cats.

First, we broke down karma earned in the top 100, 50, 25, and 10 subreddits as reported by redditlist [9]. This way that users who participate mostly in small, niche subreddits would be somewhat penalized for not participating in 'useful' subreddits. However, it is possible that one would want to know the reliability of users within a certain community. For the purposes of this project, we picked a small list of 'trusted' subreddits, and picked good, reliable users from those communities as training examples. Then, the amount of karma accrued from those subreddits was another feature that we used.

3.4.2 reddit Gold and Gilded Content

Another interesting feature of reddit which differentiates it from other social networks is reddit gold. reddit gold is a premium membership which users may pay for, and comes with many perks, like an ad-free experience, a members only subreddit, deals from partner companies, and more [8]. What makes reddit's gold program so unique is that users are encouraged not only to buy this for themselves, but also to give reddit gold membership as a gift to other users who produce content that they enjoy. Furthermore, this content is marked as 'gilded', and can be tracked. We suspected that this action would be extremely important in determining reliable content, and only content which was reliable would be worth a user spending real money to reward.

3.4.3 Natural Language Processing

Finally, we suspected that the actual writing patterns of reliable users would separate them from unreliable ones. Specifically, we tried three things that we suspected might have some sort of correlation with reliability.

1. **Unique Word Percentage**: We suspected that users who were reliable would use more unique words than those who were not.
2. **Readability**: We analyzed the Flesch–Kincaid readability [4] of the comments users posted, which can be calculated using the formula in Equation (1).

$$206.835 - 1.015\left(\frac{\#\text{ words}}{\#\text{ sentences}}\right) - 84.6\left(\frac{\#\text{ syllables}}{\#\text{ words}}\right) \quad (1)$$

3. **Swearing Rate**: We took a list of 'bad' words which Google compiled [1], and calculated the percentage of words that users used in comments which appeared on that list.

3.5 Picking a Regression Model

In picking a regression model, two things needed to be considered, namely: (1) the robustness of the model, and (2) the ease with which we could garner insights from the model generated.

Ultimately, random forests [3] was chosen based on the notion that the results were more interpretable than other popular modeling algorithms such as neural networks, and would provide greater reliability than just a simple decision tree given that random forests has boosting built in.

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.

Table 1: The features used to create the regression model

Feature Description	Importance %
Link Karma	53.13
Comment Karma	18.41
Average Karma per Post	13.34
Number of Total Posts	9.45
% of Comment Karma - Top 100 Subreddits	2.08
Has Verified Email	0.93
Average Comment Karma per Comment	0.75
% of Comment Karma - Top 10 Subreddits	0.33
Unique Words / Total Number Words	0.29
Number of Total Comments	0.20
Time Account Created	0.17
% of Comment Karma - Trusted Subreddits	0.15
Flesch–Kincaid Readability of Comments	0.15
Is Reddit Gold	0.14
% of Post Karma - Top 50 Subreddits	0.10
% of Post Karma - Top 100 Subreddits	0.10
% of Comment Karma - Top 50 Subreddits	0.08
% of Post Karma - Trusted Subreddits	0.07
% of Comment Karma - Top 25 Subreddits	0.06
% of Swear Words Used in Comments	0.03
% of Post Karma - Top 25 Subreddits	0.03
% of Post Karma - Top 10 Subreddits	0.02
Number of Gilded Posts	0.00
% of Posts Gilded	0.00
Number of Gilded Comments	0.00
% of Comments Gilded	0.00

On our test set of the data from about 3K redditors, we used our 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.

TODO

5. CONCLUSION

TODO

5.1 Future Work

Something something many weaknesses

Given the set of users we have obtained in conjunction with the variability that exists among the various subsets of reddit users there exists weaknesses in our analysis. However, given the current state of research of reddit and the increasing reliance of individuals on reddit as a source of reliable information our contribution noteworthy for the purpose of exploration with respect to reddit as an up-and-coming source of information.

Something something exploratory/first exploration of field

something something Apollo [5].

TODO

5.2 Open Issues

Something something hard to establish ground truth

Voting on reddit will always be somewhat ambiguous

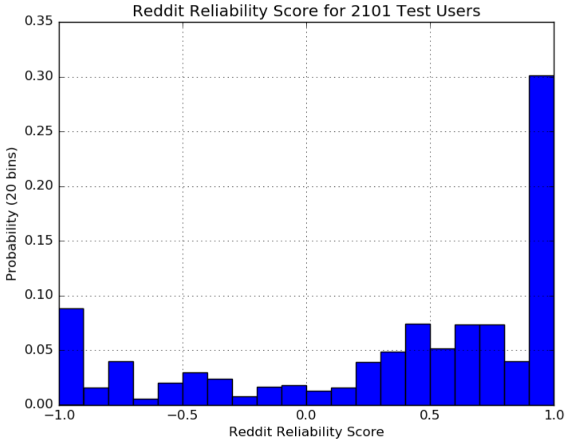

Figure 1: The distribution of the reliability score s_r of the sampled redditors, binned into twenty bins.

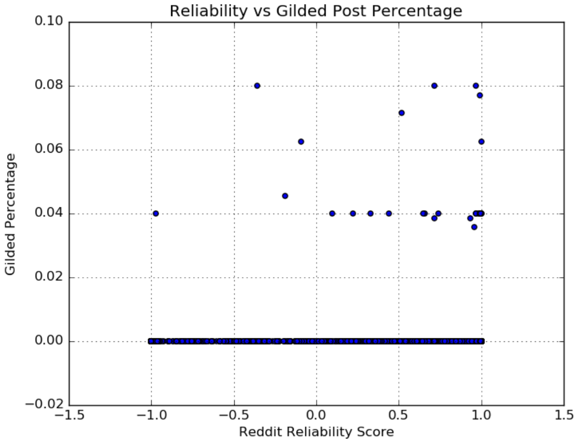

Figure 2: The reliability score s_r plotted against the percentage of gilded posts.

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] J. Dubs. Google's official list of bad words, 2011.

[2] 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.

[3] T. K. Ho. Random decision forests. In *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*, volume 1, pages 278–282 vol.1, Aug 1995.

[4] J. P. Kincaid, R. P. Fishburne Jr, R. L. Rogers, and B. S. Chissom. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel.

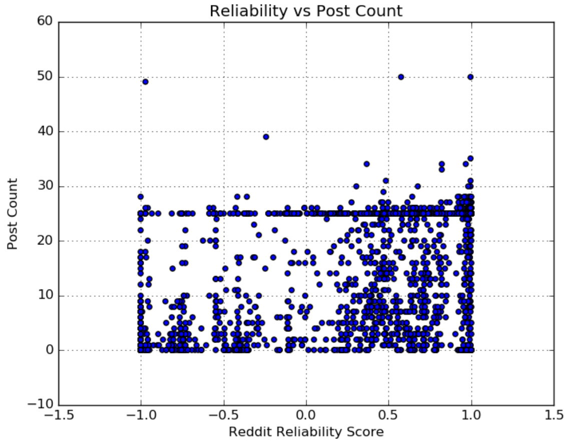

Figure 3: The reliability score s_r plotted against the number posts the redditor has made.

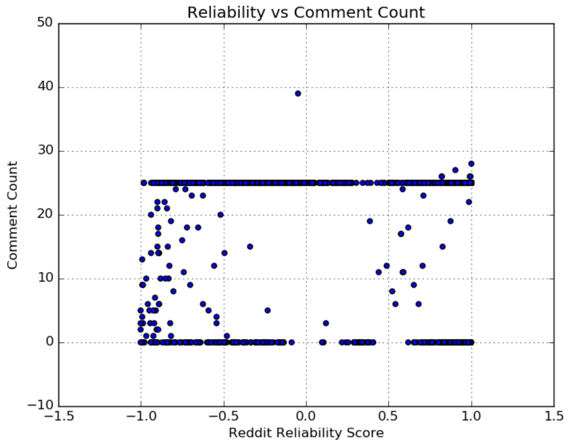

Figure 4: The reliability score s_r plotted against the number posts the redditor has made.

Technical report, DTIC Document, 1975.

[5] H. Le, D. Wang, H. Ahmadi, Y. S. Uddin, B. Szymanski, R. Ganti, and T. Abdelzaher. Demo: Distilling likely truth from noisy streaming data with apollo. In *Proceedings of the 9th ACM Conference on Embedded Networked Sensor Systems*, SenSys '11, pages 417–418, New York, NY, USA, 2011. ACM.

[6] 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.

[7] 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*, SIGDOC '13, pages 143–150, New York, NY, USA, 2013. ACM.

[8] reddit incorporated. reddit.com: gold, 2015.

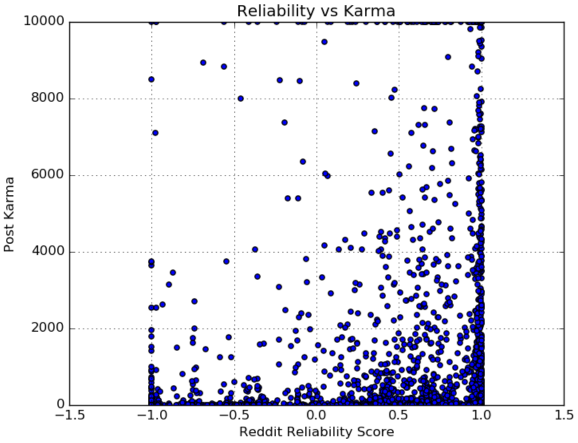

Figure 5: The reliability score s_r plotted against the average Karma per post the redditor has made.

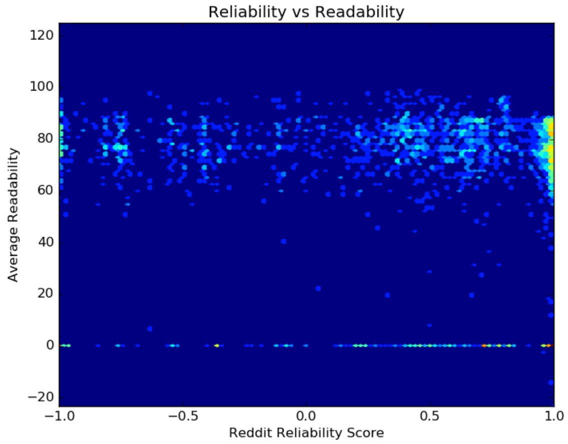

Figure 6: The reliability score s_r plotted against the Flesch–Kincaid readability of their comments.

[9] /u/mikesizz. redditlist.com – tracking the top 5000 subreddits, 2015.

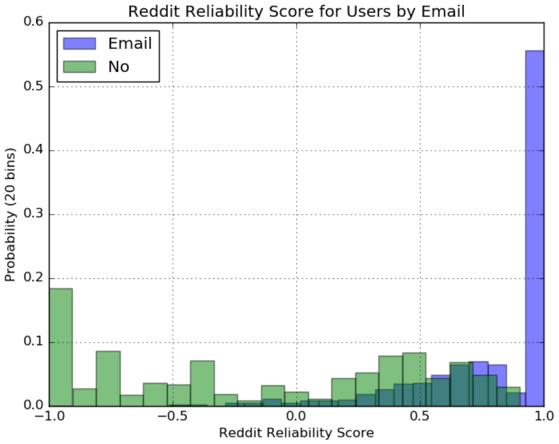

Figure 7: The distribution of the reliability score s_r of the sampled redditors, binned into twenty bins, separated by if they have a verified email address or not.