Proposal: Evaluating & Predicting the Reliability of Redditors

Kashev Dalmia, Ryan Freedman, Terence Nip
{dalmia3, rtfreed2, nip2}@illinois.edu

October 4, 2015

1 Introduction

Reddit, the self-proclaimed "Frontpage of the Internet", is a website containing entirely user-contributed, uncurated content organized by topic. Over the past few years, Reddit has gained prominence within popular media channels for various reasons, all surrounding the content submitted by Reddit users, or Redditors.

Redditors can be extremely engaged and reliable, and therefore useful as sensors. In this paper, we propose a method of determining if Redditors are reliable based on their post and comment history, as well as the reaction of the community to their activity. Furthermore, the resulting analysis can be used to create a predictor of future Redditor activity. For example, this would allow for the creation of nudges to encourage users who demonstrate potential dependability to turn them into actual, reliable users - and conversely, discourage unreliable users from posting on Reddit, increasing the net dependability of the site.

1.1 Why Reddit?

Reddit is quickly becoming a larger and larger social network, and much of its potential remains untapped. It is unique among social networks for a few reasons, the combination of which make it attractive for study and use as a sensor network.

Though more things can be posted, Twitter is primarily used for text. Instagram is for photos, and Vine is for videos. Reddit, however, is extremely content agnostic; Users post photos, links, videos, and text, relating to a variety of topics. Twitter caps users at 140 characters, Reddit does not. Facebook is mostly private content, Reddit is largely public. Reddit also comes with the built in benefit of content being organized by subject matter, into various 'subreddits'. In this way, many positive user behaviors are already encouraged, like posting content in the proper subreddit.

1.2 Finding Reliable Users

There are some issues with Reddit that are less present in social networks like Twitter. For instance, Twitter has verified accounts, and if the verified account `@cnn` tweets something, one can be relatively sure it is researched and vetted information which is likely to be correct. Reddit has no such user verification mechanism built-in natively, nor do they make an effort to provide specialized functionality for accounts belonging to brands or news organizations . Moreover, many Reddit users are completely anonymous and could potentially have multiple accounts, which can cause rise to all sorts of poor behavior. Thus, determining the reliability of users is a challenge not just for most users as it is for Twitter, but for all users.

2 Proposed Work

In light of the issues with Reddit, it is valuable to determine the reliability of Redditors automatically. We plan to do this by creating a Reliability Score, ranging beetween -1.0 and 1.0, in which a more positive number denotes a user who is more reliable, and a negative score denotes a user who is less reliable. Furthermore, we hope to show that based on indicators used to calculate this score over time, that we can anticipate the trend of a Redditor's reliability. This has implications in creating reliable users by predicting their trajectory of behavior. One could take this trajectory, compare it to a more 'desirable' trajectory, and perhaps even nudge Redditors towards a more Reliable trajectory.

The following sections discuss some technical challenges we anticipate, and our plans to overcome them.

2.1 Creating A Reliability Score

In coming up with a metric for determining reliability of a user, there are certain traits of a reddit user that should be addresssed.

'Karma', reddit's crowd-sourced measure of whether or not a user's submitted content is "good", allows other users to rapidly identify whether or not a user tends to post interesting/good content relative to how the "hivemind" feels (based on a user's net post/comment karma). Karma, in itself, however, is not a proper measurement for how reliable a reddit user is, as one could gain karma for posting content that the hivemind simply approves of - puns, cat pictures, references to past popular reddit threads - and not necessarily for content that proves to be important relative to the "real world".

Posting in trusted subreddits, however, is a step forward in turning karma from a measurement of approval by the community into a measurement of reliability as given by the hivemind. By removing karma given by subreddits which have a dearth of "serious" content (such as `r/funny`) in favor of those with a wealth of serious content (such as `r/news`), we skew the existing measurement of karma into one that is based upon trust from redditor to redditor - that is, instead of being a measurement of how much redditors approve of a specific piece of submitted content, karma now becomes a measurement of how trustworthy that specific piece of content is - and in aggregate, becomes a measurement of how trustworthy/reliable that user is.

It is also important to take into account the sharing of content from reputable, reliable sources. Given the nature of reddit, one could easily share serious content from reputable sources like the New York Times on subreddits that may not necessarily be serious, like on `r/nottheonion`, where users share links to articles which sound like they could be written by and published in the Onion but are, in fact, serious articles. In cases like this, it is readily evident that the actual subreddit is less important, in favor of the actual content itself.

There are also numerous other metrics that can serve as a proxy for reliability, such as account duration and the user's choice of subreddits to post in. Account duration could be used as an identifier to determine if an account is being used as a "throwaway", or an account that is merely being used to create controversy/share opinions that might not be popular without linking it to one's primary username. Similarly, a user's choice of subreddits could be used as a proxy to determine whether or not a user is reliable simply by seeing what their interests are. For example, given two users, one frequenting `r/funny`, a subreddit where users post funny content versus one who frequents `r/news`, where users post U.S. news, odds are that the second user will be posting more reliable, serious content than the first user.

In an effort to reap all the benefits that inherently exist within data being generated by redditors and gathered by reddit, we propose this notion of a reliability score, a aggregate score generated by the fusion of all these different factors into a number which is quickly able to identify whether or not a user is reliable.

2.2 Gathering Usernames

Reddit does not have an API for gathering or searching for usernames. It does however, allow one to see popular posts on particular Subreddits, or in general, the front page. We plan to mine usernames by looking at popular posts and comments on those posts, and logging the usernames associated with those activities. This list of usernames can then be plugged into our reliability score calculations, and used to train our classifiers for behavior prediciton.

2.3 Verification

Verification of reliability is tricky. Though it is labour intensive, we will have to select a control group of users, view their history manually, decide if they are 'reliable' or not, and then compare that to the score our classifier gives.

Verification of behavior prediction is less tricky. We can simply use our models to predict the future behavior of old users, and then compare that to their actual trajectory.

3 Proposed Milestones

Writing Milestone	**Date**
Abstract	Sep 15
Introduction	Oct 15
Current Work	Oct 15
Challenges Faced	Nov 15
Open Challenges	Nov 15
Conclusion	Nov 15
References	Nov 15
Technical Milestone	**Date**
User List Data Grab	Oct 15
User Ground Truth	Oct 20
User Characteristic Grab	Nov 5
Classifier Written with Characteristics	Oct 20
Data Run	Oct 25
Data Analyzed	Oct 31
In-class Presentation 1	Nov 1
In-class Presentation 2	Nov 20