

Proposal: Evaluating & Predicting the Reliability of Redditors

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

September 29, 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 Creating Good 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 mechanism, nor accounts for 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

Say what we’re going to do, here, and a little bit about how it’s going to work.

TODO

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

2.1 Gathering Usernames

TODO

2.2 Creating A Reliability Score

Talk about:

1. Relibility Karma vs overall karma
2. Posting in trusted subreddits
3. Posting reliable sources
4. Others
5. Fusion method

2.3 Verification

TODO

3 Proposed Milestones

TODO