Characteristics of a Successful Reddit Post

Stephen Smart, Emily Southern, Shruthi Sukumar, and Adam Winchell\*

University of Colorado Boulder – INFO 4602/5602 – Spring 2018

Abstract

The goal of this project was to determine the characteristics of a successful Reddit post. We analyzed data from eight subreddits consisting of text-only posts. We created three visualizations to assist Reddit users interested in generating the most karma.

**Keywords**: subreddit, topic model, D3.js

# Introduction

Reddit is a popular social news and media aggregation website. Registered users submit content, such as text posts, links, and images, to user-created communities called ‘subreddits’. These communities cover a wide range of topics including, but not limited to, news, science, movies, video games, music, books, and food. Some subreddits allow for text-only submissions, which we focus on here. Specifically, we chose to analyze a collection of subreddits dedicated to posts asking focused questions. These are communities like AskHistorians, AskScience, and AskCulinary, where members of the subreddit pose a question specific to the given community and experts within the community collectively provide their perspectives on the answer.

Each user who views a post can choose to either up-vote (i.e. like the post) or down-vote the post. The sum of the up-votes and down-votes, treated as +1 and -1, respectively, is defined as the karma of the post, which serves as a proxy for the success of a post. We aim to determine the characteristics of a successful text-only post in various Ask-based subreddits. These findings can be used to assist Reddit users interested in generating the most karma.

LEAVE 0.5 INCH SPACE AT BOTTOM OF LEFT COLUMN ON FIRST PAGE FOR COPYRIGHT BLOCK

# Related Works

\*add summary of research related to our project

## Subsection One

\*subsection for related works

## Subsection Two

\*subsection for related works

\* adam.winchell@colorado.edu

# Description of Project

We created three visualizations to help reddit users understand the characteristics of a successful subreddit post. Visualization 1 shows the topics of successful posts in a given subreddit. It also provides information about which subreddit it is easiest to be successful in. Visualization 2 helps users decide which time of day to post in order to generate the most karma. Visualization 3 allows the user to define the karma threshold that distinguishes between a “good” and a “bad” post, and then displays the number of posts in each subjective category for each subreddit.

We collected data from eight Ask-based subreddits using Reddit's provided Python API. We collected data over course of one week from the following subreddits: AskHistorians, AskAcademia, AskAnthropology, AskCulinary, AskMen, AskScience, AskSocialScience, AskWomen. The process consisted of checking each subreddit every hour for new posts and recording the time the post was created, the number of users on the subreddit at the time of the post, and the title of the post. We recorded the karma of the post 24 hours after the post was made.

## Visualization 1

Our first visualization is a bubble chart showing the topics of successful posts in a given subreddit. \*explain topic modeling here

The user can choose which subreddit to visualize by selecting it from the dropdown menu above the bubble chart. Each different colored bubble represents a different topic, which is displayed within the bubble along with the number \*what does this number represent?. The size of the bubble corresponds to… \*does color have any meaning/how did we choose colors?

## Visualization 2

Our second visualization is a line graph showing the popularity of posts in a given subreddit over time. The horizontal axis represents time. The user can choose to visualize the entire week of data (Monday-Sunday) or all 24 hours of one specific day of the week using the dropdown menu. The vertical axis is the mean karma normalized by the number of users on the selected subreddit. \*discuss animation and color choice

## Visualization 3

Our third visualization is a stacked bar chart displaying the number of posts above and below a given karma threshold for each of the subreddits. The user can change the threshold that differentiates between a “good” and “bad” post using the slider tool below the visualization. \*discuss animation and color choice

# Discussion

Our system of visualizations was designed to determine the characteristics of a successful Reddit post.

\*add findings from first visualization

\*add findings from second visualization

\*add findings from third visualization

References

1. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. "Latent dirichlet allocation." Journal of Machine Learning Research 3:993-1022, 2003.
2. Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. "D³ data-driven documents." IEEE Transactions on Visualization and Computer Graphics 17(12):2301-2309, 2001.
3. Liangjie Hong, Ovidiu Dan, and Brian D. Davison. "Predicting popular messages in twitter." Proceedings of the 20th International Conference Companion on World Wide Web, pages 57-58, 2011.
4. Robert Kosara. "Presentation-oriented visualization techniques." IEEE computer graphics and applications 36(1):80-85, 2016.
5. Jey Han Lau, Nigel Collier, and Timothy Baldwin. "On-line trend analysis with topic models: #twitter trends detection topic model online." Proceedings of COLING 2012: Technical Papers, pages 1519-1534, 2012.
6. George Robertson, Roland Fernandez, Danyel Fisher, Bongshin Lee, and John Stasko. "Effectiveness of animation in trend visualization." IEEE Transactions on Visualization and Computer Graphics 14(6):1325-1332, 2008.
7. Amir Salihefendic. "How Reddit ranking algorithms work." Hacking and Gonzo, 2015.
8. Tauhid Zaman, Emily B. Fox, and Eric T. Bradlow. "A Bayesian approach for predicting the popularity of tweets." The Annals of Applied Statistics 8(3):1583-1611, 2014.