

User Manual

Project Title: Detecting brigading on social media with machine learning

Student Name: Adam Pegman

Student ID: 14351741

Stream: CASE4

Project Supervisor Name: Ray Walshe

Reddit Moderator Bot

Sample Sets

r_9ww ✖

r_cats ✖

r_dogs ✖

r_animals ✖

Name:

Link:

Size:

10000

Add sample set

Monitored Subreddits

Subreddit name:

Add subreddit

Config

Client id

Client secret

Username

Password

submit

Triggers

Name:

Regex rule:

Add Rule

Name:

Type

Bayesian

Good Comment Sample

r_9ww

Bad Comment Sample

r_9ww

Add Rule

Actions

Name

Message

Add Message action

Name

Type

Approve

Add action

Rules

Name:

Subreddit

Trigger

Action

Add Rule

This is the user guide for a Reddit moderation bot. The goal of the bot is to replace Automoderator, the current bot used by Subreddit moderators to help moderate their subreddits. The service can watch a subreddit and apply rules, which can be triggered by regex or by using machine learning to compare them against collected samples. The service is designed to run locally on the moderator's machine.

Installation

The service is provided as a jar file. When this jar file is run, the service creates a database and starts running on <http://localhost:8080/> Visit this page to continue configuring.

Config

Client id

Client secret

Username

Password

submit

The first step of configuring the service is to create a Reddit account with permission to moderate your Subreddit. Once this is done, you can obtain a `client_id` and `client_secret` from <https://www.reddit.com/prefs/apps> and enter them into the config form to bind the bot to your subreddit, then pressing the submit button.

Monitored Subreddits

Subreddit name:

Add subreddit

The next step is to add your subreddit to the list of watched subreddits for the service. This can be done by pasting a link to your subreddit or “/r/[name]” into the monitored subreddits form and clicking submit. The name of your Subreddit should appear in the list above, along with delete button that can be used to remove it.

Sample Sets

r_aww ✖

r_cats ✖

r_dogs ✖

r_animals ✖

Name:

Link:

Size:

10000

Add sample set

The service uses samples of comments from subreddits or threads to trigger it's rules. You can gather samples of desirable and undesirable comments to use in your rules by entering a name, a link to the comments on Reddit and the number of comments you want to collect, then pushing the "Add sample set" button. The sample set will initially appear orange until the service finishes collecting the comments and calculating statistics about them.

Triggers

Name:

Regex rule:

Add Rule

Name:

Type

Bayesian

Good Comment Sample

r_aww

Bad Comment Sample

r_aww

Add Rule

Triggers are used by the service to decide when a rule should be run.

Regex triggers use the same regular expressions Automod uses to detect comments.

The other available triggers are:

1. The Bayesian trigger service, which requires a good set and a bad set, and classifies a comment as good or bad based on the words in the comment.
2. The Similarity trigger, which takes a good set and a spam set, triggers if 2 of the 3 nearest comments by word counts are spam.
3. The
4. The All trigger, which triggers on every comment.

Actions

Name

Message

Add Message action

Name

Type

Approve ▼

Add action

Actions are the actions taken by the service when a rule is triggered.

The available actions are:

1. Posting a message replying to the comment.
2. Removing the comment.
3. Banning the user who posted the comment.
4. Approving the comment.

Rules

Name:

Subreddit

Trigger

Action

Add Rule

A rule is a Trigger, and an Action that should be triggered by it, as well as a Subreddit the rule should be applied to. This form should be filled out last. Rules should appear in a list above the form once created and can be deleted via the delete button.