

My Automoderator is the future of the internet but still not perfect.

The AutoMod Proposal

Reddit is an American social news aggregation, web content rating, and discussion website. As of October 2020, Reddit ranks as the 17th-most-visited website in the world and 7th most-visited website in the US, according to Alexa Internet, with 40.9% of its user base coming from the United States, followed by the United Kingdom at 10.0% and Canada at 5.2%. The challenge for Reddit is to keep its online forums clean of spam and abusive users.

A moderator is just a regular Redditor except with a few humble duties within a particular community, which include the following: Configure parameters for the community, like what its description should be or whether it should be considered "Over 18", set custom logo and styling, if any, mark links or comments as the community moderator's submission, which just adds an "[M]" and turns their name in green, remove links and comments from their community if they find them objectionable or off topic. They can ban a spammer or other abusive user from submitting to their community. (This has no effect elsewhere on the site), add other users as moderators and while Moderators have no special powers outside of the community they moderate, and are independently appointed. Mods are reporting for duty on the front lines, within the social site's volunteer army.

The solution would be to find a way to programmatically moderate the forums using cutting edge machine learning technology. Natural Language Processing has its roots in the 1950s, with Alan Turing's Turing test as a criterion of intelligence, a task that involves the automated interpretation and generation of natural language, but at the time not articulated as a problem separate from artificial intelligence. Up to the 1980s, most natural language processing systems were based on complex sets of hand-written rules. Starting in the late 1980s, however, there was a revolution in natural language processing with the introduction of machine learning algorithms for language processing. With this solution, reddit could expand their overall market share on the web and grow community overall size.

The power of NLP could be seen just with my initial results, with just 6 weeks of exposure to programming and just 1 week of exposure to NLP, I was able to deploy an ML model that classified two different forums posting with an 97% accuracy. My support vector machine classifier was able to distinguish between Reddits with near perfect ROC/AUC score of 99%, 1 being perfect and 0 being horrible as differing between the two classes.

I think this could be perfected to distinguish between spam and not spam, abusive or not and even compare the posted text in posts to see if there relevant to the subreddit. This will greatly improve the user experience on the website and allow for

Humans and Technology

further commercial applications, in addition to reducing the workload of those on the front lines of the internet. There are serious human implications to spam and internet trolling. This is my call to action Reddit, are you ready to fight the good fight?

Said no-one,

Anonymous User431