r/BlackPeopleTwitter and r/WhitePeopleTwitter

Comments in two subreddits based on race

Reddit and subreddits

- User submissions
  - Links
  - Images
  - Text
- Comments
- Upvotes
- Subreddits, r/all, and hot

r/BlackPeopleTwitter

From the page description:

"Screenshots of Black people being hilarious and insightful on social media, it doesn't need to just be twitter but obviously that is best.

Black culture has a unique way of examining the everyday and we are here to showcase that."

r/WhitePeopleTwitter

- No subreddit description
- Both r/BPT and r/WPT formed 3 years ago
- r/BPT has 1,743,296 subscribers as of June 2, 2018
- r/WPT has 522,242 subscribers as of June 2, 2018

Preliminary Assumptions

Topics involving race should be approached with caution

The racial identity of participants is unknown

Are posters in either subreddit predominantly one race? Are commenters predominantly one race? These questions are beyond the scope of this analysis. Additionally,attempting to answer such questions would introduce the possibility of racial discrimination.

"Framing" of online spaces

Even though we can't know how participants identify themselves, we do know that the two subreddits are framed explicitly in terms of race.

Classifying post + top comment

Can we build a model that, given the text and top comment of a particular post, will be able to classify these combined texts into r/BlackPeopleTwitter or r/WhitePeopleTwitter?

Text alone

According to this interpretation of the question, we will look only at the text (post title + top comment).

A secondary question here is whether the language in these two subreddits is enough to differentiate them--do users employ different enough ways of talking for us to build an effective model?

Adding features

For this interpretation of the question, we will use more than just the text to classify our posts. Additional features will be included:

The amount of time the post has been on Reddit

The total number of comments for the post

The number of upvotes for the top comment

The number of characters in the title + top comment

Why?

"I studied Barack Obama’s speech patterns and I’ve found that his speech changed depending on his audience. Because of prevailing racist, xenophobic and Islamophobic ideas about who he was, he had to speak in such a way that made as many people comfortable with him as possible."

-Samy Alim, co-editor of

Raciolinguistics: How Language Shapes our Ideas about Race

https://news.stanford.edu/2016/12/27/link-language-race-new-book/

Why?

The data science question is whether we can classify posts.

The broader question is about the social construction of race in online communication. If we want to avoid discrimination on the basis of race, we need to be aware of how race is constructed. This classifier might be able to address this problem.

Scraping Reddit

By making requests to Reddit's API, we have access to lots of features of each post/comment combination. But not all of those features make sense.

What could we scrape?

['subreddit_id', 'approved_at_utc', 'ups', 'mod_reason_by', 'banned_by', 'removal_reason', 'link_id', 'author_flair_template_id', 'likes', 'no_follow', 'replies', 'user_reports', 'saved', 'id', 'banned_at_utc', 'mod_reason_title', 'gilded', 'archived', 'report_reasons', 'author', 'can_mod_post', 'send_replies', 'parent_id', 'score', 'approved_by', 'downs', 'body', 'edited', 'author_flair_css_class', 'collapsed', 'is_submitter', 'collapsed_reason', 'body_html', 'subreddit_type', 'can_gild', 'subreddit', 'name', 'score_hidden', 'permalink', 'num_reports', 'stickied', 'created', 'author_flair_text', 'created_utc', 'subreddit_name_prefixed', 'controversiality', 'depth', 'mod_reports', 'mod_note', 'distinguished']

What could we scrape?

There are potentially many different features to extract from each post and title. Most of these are empty, and the ones that aren't are probably noise.

Rather than starting with all and paring them down, it's better to start with a few and add ones that make sense.

Which features?

- Total comments for the post
- Total upvotes for the top comment
- Time on reddit
- Text of title and top comment
- Length of text

For most of the posts after about 100, there aren't very many comments or upvotes... even starting from the most popular, you'll get into run-of-the-mill posts eventually.

Models for different questions

Since we are interested in two questions, we have two sets of models. Some models seem to work better to address one question over the other.

Text-only

The basic question of classifying posts+comments based on text alone proved to be relatively difficult to build a model for. I tried fitting the data using the following:

- Logistic Regression
- k-Nearest Classifier
- Decision Tree
- Random Forest
- AdaBoost
- Gradient Boosting

Text-only

The model that produced the best score against the test set was…

Logistic Regression!

Sometimes it's what you try first that does the best.

Modeling on all features

When the numeric features were considered, the models that had failed to out-perform logistic regression showed their value. AdaBoost and Random Forest models proved to be particularly strong, giving accuracies of over 80%.

What do these models suggest?

Can we answer our inital questions? How far could we extend this project?

Differentiating on text alone

The low accuracy of the models suggests limitations for classifying text as belonging to r/BlackPeopleTwitter and r/WhitePeopleTwitter without other information. However, the models perform better than a coin flip.

Problems to resolve

Issue:

Model reflects current topics that are making the news

Solutions:

- Update the model regularly (add 250 new posts from each sub, throw out 500 old posts)
- Use this information to make decisions about what's important to these two online communities

Problems to resolve

Issue:

These models are difficult to interpret.

Solution:

More feature engineering based on these results.

Extensions

Based on these models we can extend the original questions. Are the same users active in both communities? If so, do they participate in different ways in each one?

Extensions

Could we extend this framework to classify different writing genres?

A final note

Is this project vulnerable to misinterpretation?