The Project Proposal

Name: James Liao ID: 490851 Program: MSBA-Fintech

Topic:

Rank relative ratings of toxicity between comments

Main issues or problems:

When we ask human judges to look at individual comments, without any context, to decide which ones are toxic and which ones are innocuous, it is rarely an easy task. In addition, each individual may have their own bar for toxicity. It is important to construct the model that can discern the severity of the comments.

Background information about the project:

For open platform that allows people to leave the review or comments, some people tend to send the toxic comments on purpose. Theses comments usually has more influence than positive review to effect new customers or client's decision and unreasonably downgrade the corporate image. It is important for business owner to filter out those violent comments (but not the negative comments) to keep the customers' feedback more reasonable and precise.

Data Description:

- comments_to_score.csv - for each comment `text` in this file, your task is to predict a `score` that represents the relative toxic severity of the comment. Comments with a higher degree of toxicity should receive a higher numerical value compared to comments with a lower degree of toxicity; scores are relative, and not constrained to a certain range of values. NOTE: the rerun version of this file has ~14k comments that will be scored by your submitted model.

- sample_submission.csv - a sample submission file in the correct format

- validation_data.csv - pair rankings that can be used to validate models; this data includes the annotator `worker` id, and how that annotator ranked a given pair of comments; note, this data contains comments that are not found in `comments_to_score`.

114890	" Gjalexei, you asked about whether there is an ""anti-editorializing"" policy here. There is, a...
732895	Looks like be have an abuser , can you please look into this? thanks.
1139051	I confess to having complete (and apparently blissful) ignorance of Jordan, but I've glanced at the ...

What questions and/or concerns do you have about your project?

The main problem I expected to counter is:

1. The model is unable to discern between negative comments and toxic comments
2. The model may overfitting easily due to the pattern of texting is different people to people