

Automatic Classification of Persuasive Arguments

Josh Berlinski Gleb Promokhov

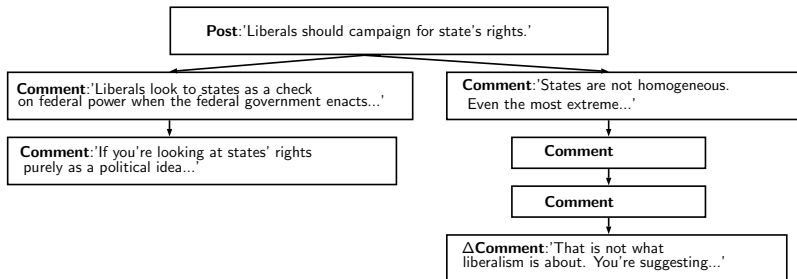
December 4, 2017

Introduction

- ▶ Try to answer the question: *"Are there identifiable stylistic features in effective arguments?"*
- ▶ With a corpus gathered from /r/ChangeMyView¹, we trained a model for classifying novel arguments as 'persuasive' and not
- ▶ Through tuning the model and experimenting with our feature set, we found that certain features had high correlation to changing someone's view

¹<https://reddit.com/r/ChangeMyView>

Figure 1: Example comment tree



- ▶ Corpus collected from /r/ChangeMyView, an active Reddit community which encourages sharing views and discourse in a moderated public forum.
- ▶ Comments on views may be awarded a 'delta point', indicating it has successfully changed the post author's view.

Features

- ▶ Style features include parts-of-speech, hyperlinks, sentiment and markdown formatting distributions.
- ▶ Clustered comments into 5 topics to capture non-stylistic features

Example: Topic 0

The fact that it may be emotionally difficult for some people might be due to living in a society that for hundreds of years has advocated heterosexual monogamous unions...

Topic Examples

Example: Topic 1

... For homosexuality, it's much the same, except that the event that “activates” the gene happens in utero, and the event is actually a hormone deviation from the “normal” amount of hormones experienced...

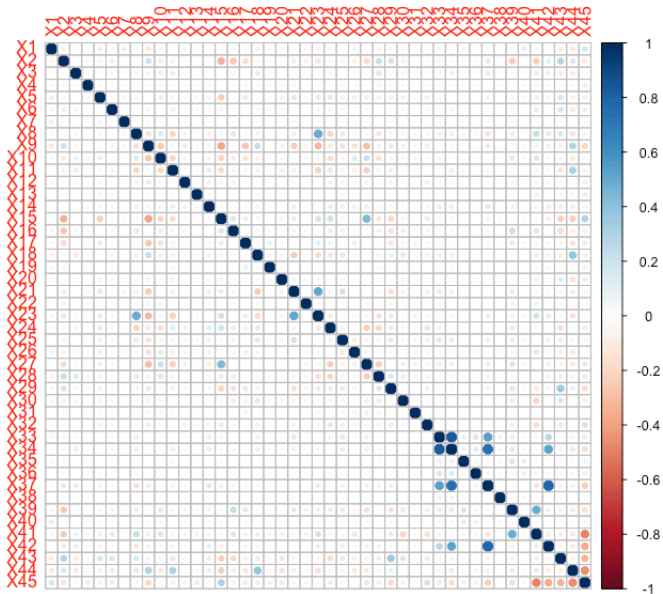
Example: Topic 4

...You need a few more axioms to get to utilitarianism ... One is some sort of impartiality constraint – that my suffering/wellbeing is no more important than your suffering/wellbeing...

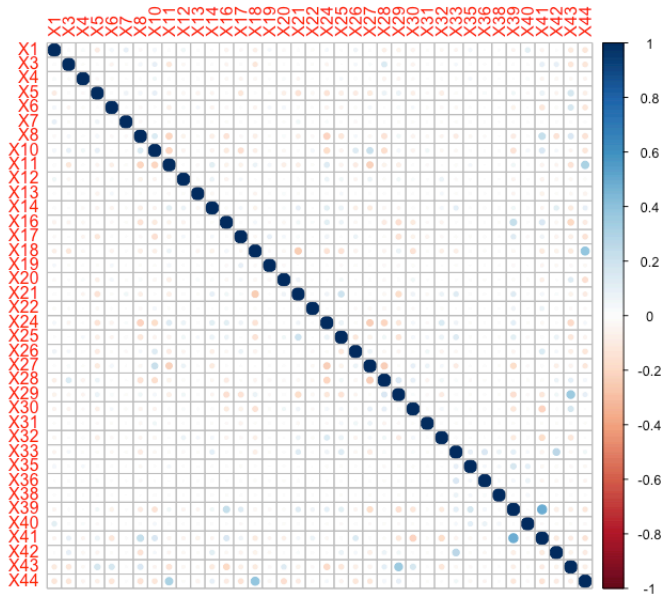
Classifier

- ▶ Extreme class imbalance in original dataset
- ▶ Undersampled to 66% non-delta and 33% delta comments
- ▶ Trained LASSO (L_1 regularized) logistic regression with 80% training and 20% testing split.
- ▶ Manually pruned POS features
- ▶ Automatically removed other features based on correlation

Correlation of Features



Correlation of Features



Results and Discussion

Majority baseline accuracy: **66%**

Classifier accuracy without topic features: **num%**

With topic features: **98.7%**

- ▶ The high value of POS tags features, counted as ratio of count of tag to total words in comment, in classification indicate that well-HOW EXACTLY?
- ▶ The high value of topic features suggest that users of CMV hold variably malleable views on different subjects, by rewarding higher number of delta points in specific topics.

Discussion

Notable Features:

- ▶ Comparative adjectives
- ▶ Possessive Pronouns
- ▶ Philosophical Topic

Conclusion