Detecting Objectifying Language in Online Professor Reviews

Angie Waller and Kyle Gorman
The Graduate Center, City University of New York

Our experiments show that an ensemble of classifiers can accurately detect objectifying language in online reviews, allowing analysis of professor reviews at scale.

We find that after RateMyProfessors removed the chili pepper rating from their website, students are less likely to comment on the attractiveness of their professors.

We also find that objectifying comments on the website have been in decline over the past ten years.

Acknowledgments

Martin Chodorow: guidance in statistical analysis
Deepali Advani: data preparation
Jonathan Butterick: data collection assistance
Sara Morini: data annotation
William Jordan : data annotation and proofreading
Anonymous reviewers: helpful feedback

Contact: aw@angiewaller.com

THE GRADUATE CENTER
CITY UNIVERSITY OF NEW YORK

Background

SCREEN CAPTURE OF PROFESSOR REVIEW FROM RATEMYPROFESSORS.COM.

Until recently RateMyProfessors.com used a chili pepper design feature to encourage a "hot or not" rating of college professors.

Objectifying comments about professors remain and continue to be posted in this online context.

This study measures correlations between objectifying commentary, changes to the review website interface, and teacher gender across a ten-year period.

Data

28,158 Professors / 358,969 Reviews
15,067,393 Tokens / 71,710 Unique
Dates: 10/13/1999 – 10/01/2019
46 total public and private universities were sampled representing all U.S. regions

4000 reviews are annotated for training and dev. set.
600 reviews are sampled and labeled for test data.

Materials & Methods

We implement classification techniques with unique strengths for capturing the qualities and contexts of objectifying comments.

1. A **chunk tagger**, trained on spans of text that make up attractiveness commentary, represents a bottom-up strategy.

2. A **document classifier** that considers features of the entire review uses top-down processing and a richer feature set.

Results

Both classifiers have relatively high accuracy but significantly lower precision and recall.

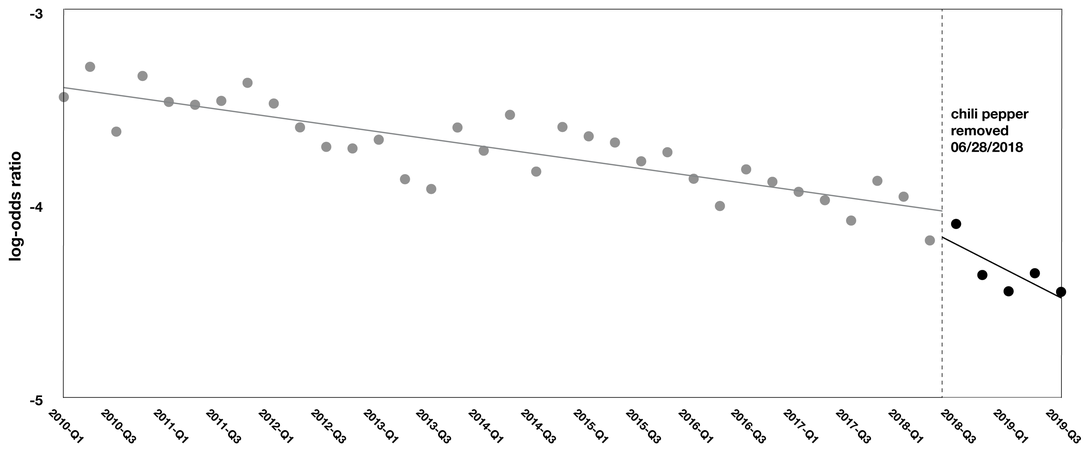
Classifier	Prec.	Recall.	F1	Acc.	Kappa
Chunk tag.	.42	.21	.28	.89	.23
Doc. class.	.44	.23	.30	.93	.26

After the chunk tagger and doc. classifier are trained, a simple **document-level ensemble** of these models is applied. These ensembles consist of two forms of voting:

In ensemble 1 we consider reviews with disagreement as non-objectifying reviews					
Ensemble 1	.72	.44	.55	.89	.50
In ensemble 2, we completely discard reviews when the classifiers disagree to achieve higher accuracy					
Ensemble 2	.74	1.00	.84	.99	.83

Analysis

Log-odds of attractiveness commentary in reviews from 2010 to August 2019



	Estimate (log-odds)	Std. err.	Wald χ^2	$p(\chi^2)$
(Intercept)	-3.111	.143	476.18	< .001
pepper removed	-.428	.136	9.93	.002
time (quarters)	-.020	.002	79.44	< .001
difficulty-high	-.075	.022	11.49	< .001
quality-high	-.051	.026	3.76	.053
female profs.	-.528	.174	9.19	.002
high quality x female profs.	.097	.043	5.09	.024

Other factors that contribute to less attractiveness commentary are teacher difficulty rating and time interval–attractiveness commentary has been in decline over time. Female teachers rated high quality are more likely to have attractiveness commentary in their reviews, while male teachers are more likely to receive attractiveness commentary overall.

Using generalized estimating equation (GEE), we find removal of the chili pepper coincides with a decline in reviews mentioning professor attractiveness.

Download paper