Expressively Vulgar

The socio-dynamics of vulgarity and its effects on sentiment analysis in social media

Isabel Cachola, Eric Holgate, Daniel Preotiuc-Pietro, Junyi Jessy Li

Eric Holgate
@eholgate

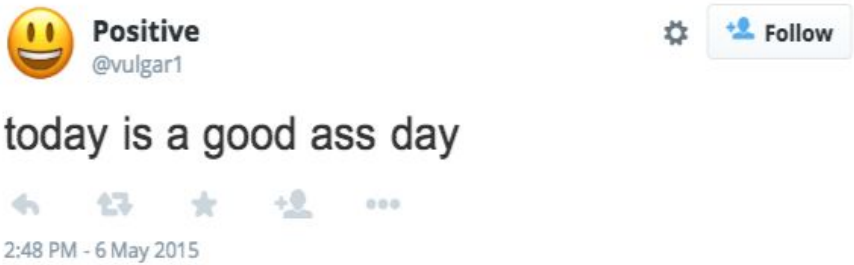
Can't wait to hear more kickass talks today at #coling2018 in beautiful #santafe! #expressivelyvulgar

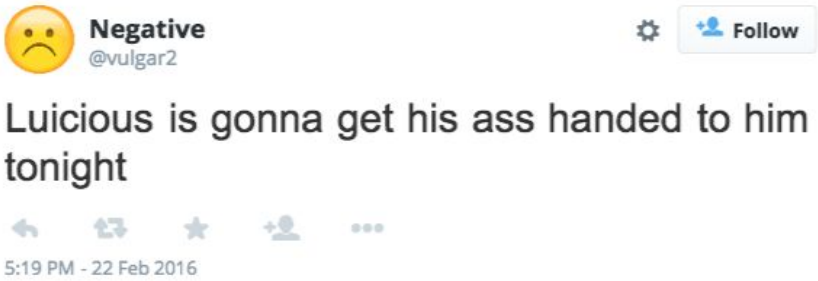
7:07 AM - 23 Aug 2018

Motivation

1. Vulgarity is prevalent in daily communication
2. Vulgarity is employed purposefully
3. Vulgarity is used for different goals

Motivation


Positive
@vulgar1
Follow
today is a good ass day
2:48 PM - 6 May 2015


Negative
@vulgar2
Follow
Luicious is gonna get his ass handed to him tonight
5:19 PM - 22 Feb 2016


Neutral
@vulgar3
Follow
[URL] volunteering at the big ass indie art craft fair
2:48 PM - 6 May 2015

Research Questions

Do demographic factors impact the expression of vulgarity?

1

Does vulgarity impact perception of sentiment?

2

Does modeling vulgarity explicitly help sentiment prediction?

3

Agenda

- Data
- Demographic Analysis
- Perception of Vulgarity
- Modeling

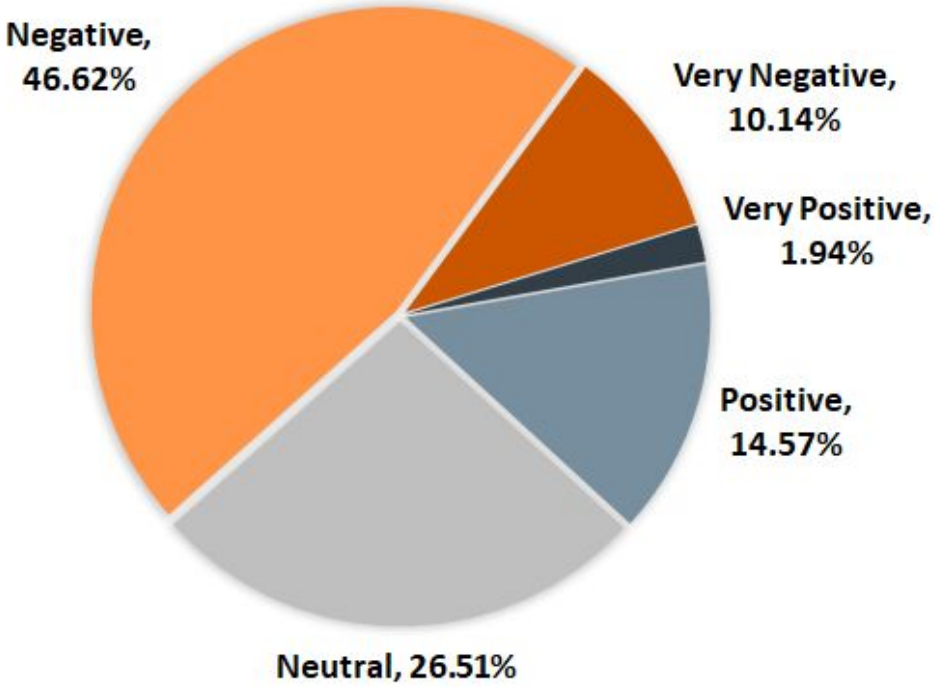
Data

- We introduce a corpus of 6,800 vulgar tweets annotated for sentiment
  - Sourced from 4,132 users with demographic info (Preotiuc-Pietro et al., 2017)
    - Gender, age, education, income level, faith, political ideology
  - Vulgarity defined with a list from www.noswearing.com
    - Regular expressions include spelling variation and self-censorship e.g., *damnnnnn* or *a$$*

Data

- Annotated for sentiment
  - MTurk: each tweet rated by 9 annotators
  - 5 point scale + Not Applicable
  - Annotators with agreement < 0.3 excluded
  - Tweets with < 5 annotations excluded
- Available at: https://github.com/ericholgate/vulgartwitter

Sentiment Distribution



Research Questions

Do demographic factors impact the expression of vulgarity?

1

Does vulgarity impact perception of sentiment?

2

Does modeling vulgarity explicitly help sentiment prediction?

3

Demographic Analysis

Partial Pearson correlation

- Dependent variable: vulgar frequency (per user)
- Control for age & gender (e.g., Schwartz et al., 2013)

Demographic Trait	Pearson r	p-value
Gender	-0.077	1.61^{-4}
Age	-0.233	6.64^{-31}
Education	-0.100	7.62^{-07}
Income	-0.087	1.73^{-05}
Faith	-0.187	2.74^{-20}
Political Ideology	0.124	8.69^{-10}

Demographic Analysis

Results

- Females are less likely to post vulgar tweets than males
- Younger users and political liberals are more likely to post vulgar tweets
- Higher education, income, and faith are anti-correlated with usage of vulgarity

Demographic Trait	Pearson r	p-value
Gender	-0.077	1.61^{-4}
Age	-0.233	6.64^{-31}
Education	-0.100	7.62^{-07}
Income	-0.087	1.73^{-05}
Faith	-0.187	2.74^{-20}
Political Ideology	0.124	8.69^{-10}

Research Questions

Do demographic factors impact the expression of vulgarity?

1

Yes!

Research Questions

Do demographic factors impact the expression of vulgarity?

1

Does vulgarity impact perception of sentiment?

2

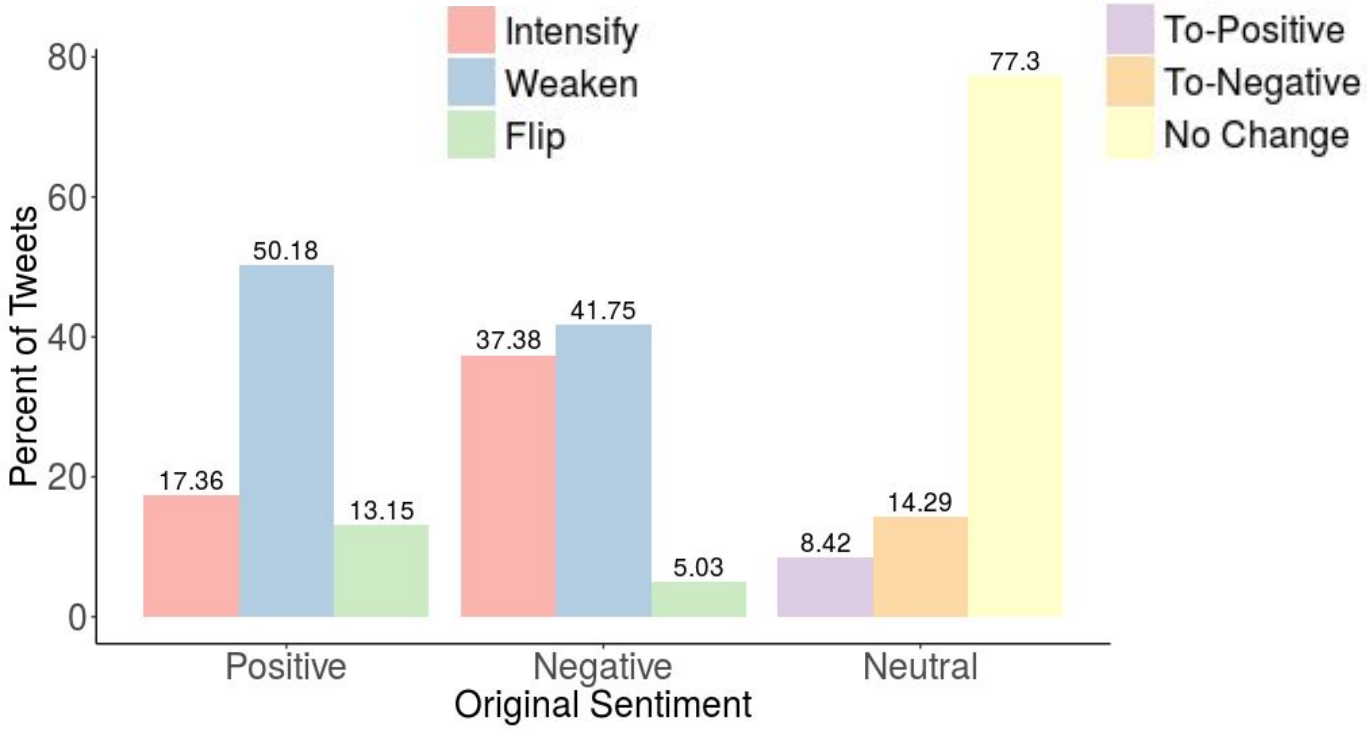
Does modeling vulgarity explicitly help sentiment prediction?

3

Perception of Vulgarity

- Subset of 5,681 tweets were censored and re-annotated for sentiment along the same guidelines
- Example
  - Original: mixing all the flavors together at the drink fountain was a good idea when i was 8 now it just makes a shitty drink
  - Censored: mixing all the flavors together at the drink fountain was a good idea when i was 8 now it just makes a _____ drink

Perception of Vulgarity



Research Questions

Does vulgarity impact perception of sentiment?

2

Yes!

Research Questions

Do demographic factors impact the expression of vulgarity?

1

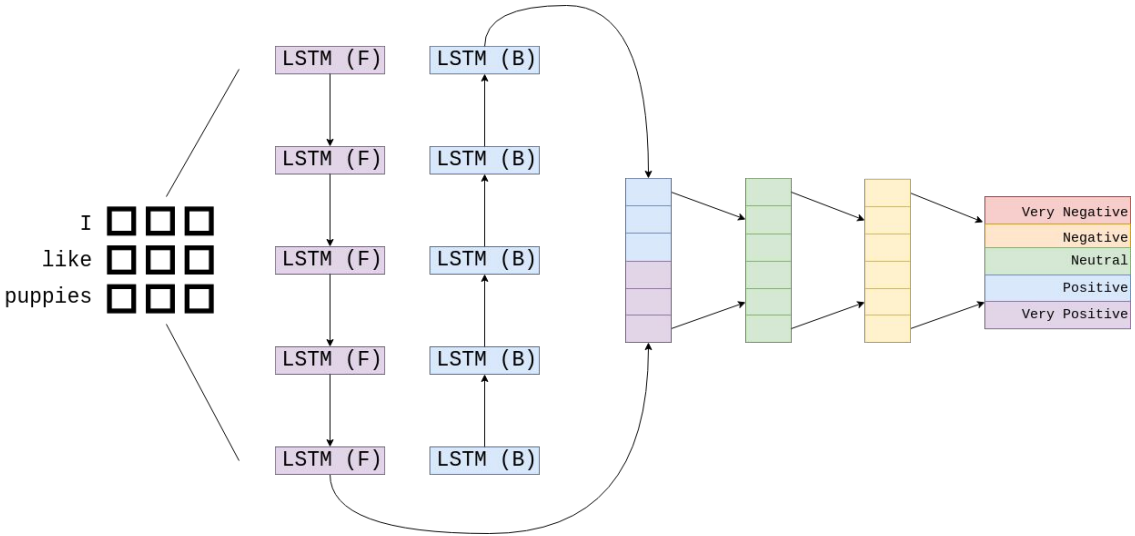
Does vulgarity impact perception of sentiment?

2

Does modeling vulgarity explicitly help sentiment prediction?

3

Modeling



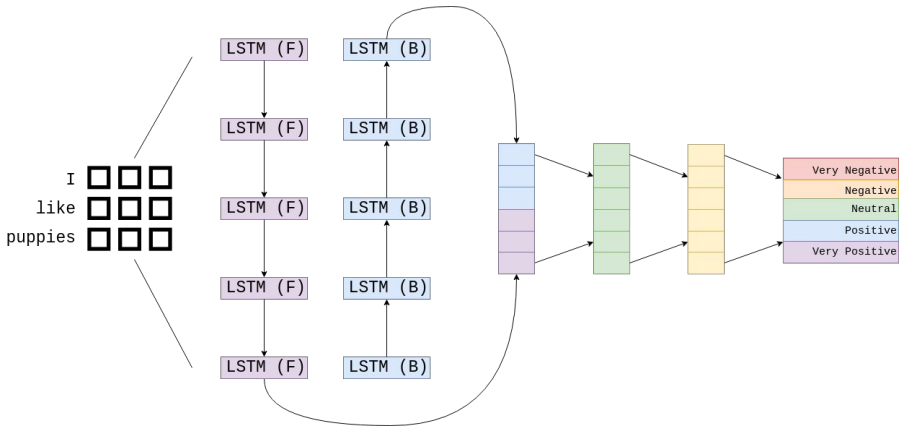
Baseline Architecture: Bidirectional LSTM

Modeling

1 **Masking:** This is the <VG>.

2 **Insertion:** This is the shit <VG>.

3 **Concatenation:** This is the shit → 1



Modeling: Vulgar Twitter Results

Model	Bi-LSTM	Insertion	Concatenation	Masking
Error	0.791	**0.759**	**0.761**	0.898

Lower is better.

Modeling Examples

Text	True Label	Baseline Prediction	Insertion Prediction	Concat Prediction
welcome to my personal hell	**Negative**	Neutral	**Negative**	**Negative**
so fucking excited	**Very Positive**	Neutral	Negative	Negative

Modeling: SemEval 2017 Task 4 Results

Model	Bi-LSTM	Insertion	Concatenation	Masking
Error	1.320	**1.068**	**1.148**	1.666

Lower is better.

Research Questions

Does modeling vulgarity explicitly help sentiment prediction?

3

Yes!

Take Aways

- Vulgarity is intentional
  - It serves many pragmatic functions
  - The use of vulgarity impacts sentiment perception
- Vulgar frequency correlates with sociodemographic factors
- Modeling vulgarity is useful for NLP

Future Work

- Can we explicitly categorize and model pragmatic functions of vulgar words?

Spoiler Alert: Yes!

- EMNLP 2018:

  Why Swear? Analyzing and Inferring the Intentions of Vulgar Expressions

  Eric Holgate, Isabel Cachola, Daniel Preotiuc-Pietro, Junyi Jessy Li

- Categorize, annotate, and predict vulgar functions
  - Aggression
  - Emotion
  - Emphasis
  - Auxiliary
  - Signal Group Identity
  - Non-Vulgar Use
- Vulgar functions are useful downstream

Thank You!

Contact:

Isabel Cachola: isabelcachola@utexas.edu
Eric Holgate: holgate@utexas.edu