Emotional Impacts of Viral Content

Sentiment Analysis of Social Media Comments

Veronica Giannotta | 02.13.2019

As user comments have grown in popularity to become a staple of internet behavior, the visibility of consumer content has only increased. Given the reputation of the comments section as an emotionally fraught place, **what are the emotional impacts of the most popular internet content?**

Project Overview

Project Overview

- Collect comment data
- Train a model on Emotion vocabulary
- Feed comment data to the model
- Interpret model results
- Tag posts with associated emotions

Project Overview

Data Sources:

Collect comment data

Train a model on Emotion vocabulary

Feed comment data to the model

Interpret model results

Tag posts with associated emotions

YouTube.com
+
Reddit.com

Project Overview

Data Source:

Collect comment data

Train a model on Emotion vocabulary

Feed comment data to the model

Interpret model results

Tag posts with associated emotions

National Research Council Canada

Project Overview

Method:

Collect comment data

Train a model on Emotion vocabulary

Feed comment data to the model

Interpret model results

Tag posts with associated emotions

Logistic Regression

Project Overview

- Collect comment data
- Train a model on Emotion vocabulary
- Feed comment data to the model
- **Interpret model results**
- Tag posts with associated emotions

Project Overview

- Collect comment data
- Train a model on Emotion vocabulary
- Feed comment data to the model
- Interpret model results
- Tag posts with associated emotions

A Note on Virality...

Data Sources

Data Sources

National Research
Council Canada

YouTube

reddit

Data Collection

Data Collection: Social Media Posts

- Calls to the YouTube and Reddit APIs
- Search for posts that fit “viral” criteria
- Account for the nested nature of comment threads
- Structure the JSON output appropriately

What was in the data sets?

NRC Emotion Lexicon:

- 19236 rows of terms
- Between 2 and 5 synonyms for each term
- Associations of each term with one or more of eight emotions:
    - anger, fear, anticipation, trust, surprise, sadness, joy, or disgust
- Associations of each term with either a negative or positive sentiment

Social Media Data:

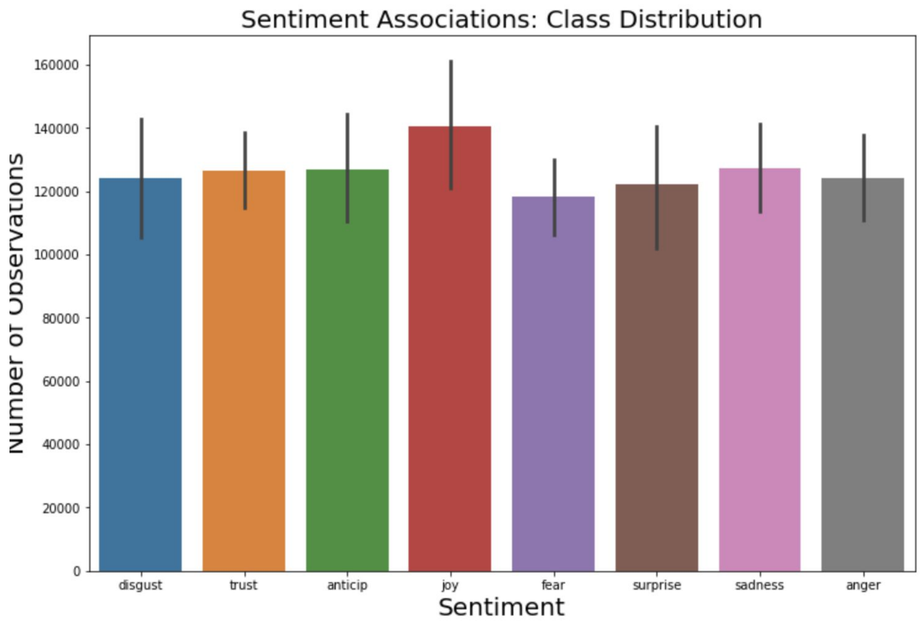
- 651 rows of unique social media posts
- Title, description, likes/dislikes, view count, total comment count, date published
- 50+ comments per post

Exploring the Data

Exploring the Data

Visualizing the NRC Class Distributions

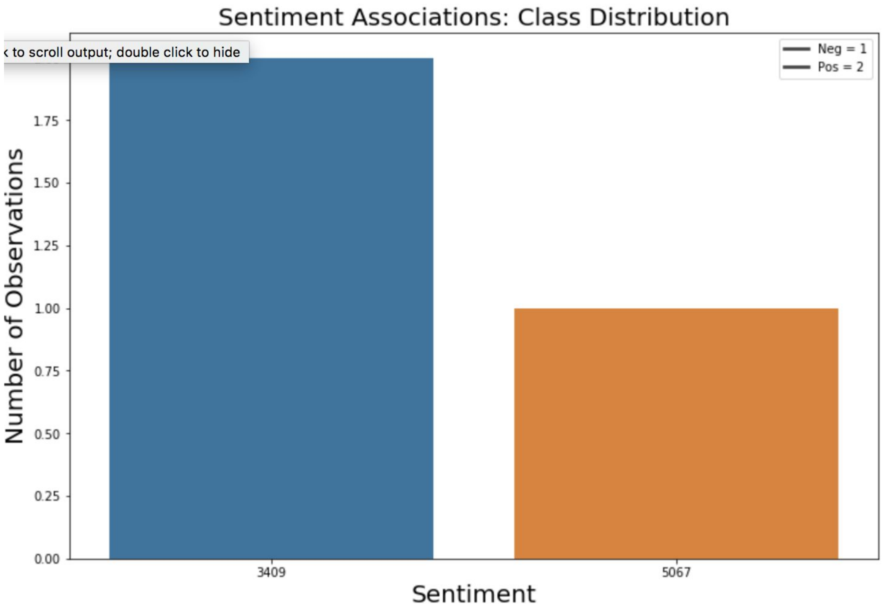
fear	1939
anger	1623
trust	1589
sadness	1513
disgust	1343
joy	1104
anticip	1016
surprise	633



Exploring the Data

Visualizing the NRC Class Distributions

negative	5067
positive	3409



Exploring the Data

Data Structure

	Term	AffectCategory	Flag	Synonym_1	Synonym_2	Synonym_3	Synonym_4
30	guide	fear	0	teaching	direct	breed	None
31	guide	anger	0	teaching	direct	breed	None
32	guide	anticip	0	teaching	direct	breed	None
33	guide	trust	1	teaching	direct	breed	None
34	guide	surprise	0	teaching	direct	breed	None
35	guide	positive	1	teaching	direct	breed	None
36	guide	negative	0	teaching	direct	breed	None
37	guide	sadness	0	teaching	direct	breed	None
38	guide	disgust	0	teaching	direct	breed	None
39	guide	joy	0	teaching	direct	breed	None

Exploring the Data

Data Structure

	Term	AffectCategory	Flag	Synonym_1	Synonym_2	Synonym_3	Synonym_4
30	guide	fear	0	teaching	direct	breed	None
31	guide	anger	0	teaching	direct	breed	None
32	guide	anticip	0	teaching	direct	breed	None
33	guide	trust	1	teaching	direct	breed	None
34	guide	surprise	0	teaching	direct	breed	None
35	guide	positive	1	teaching	direct	breed	None
36	guide	negative	0	teaching	direct	breed	None
37	guide	sadness	0	teaching	direct	breed	None
38	guide	disgust	0	teaching	direct	breed	None
39	guide	joy	0	teaching	direct	breed	None

Modeling

Modeling

Process

Train a model on Emotion vocabulary

Feed comment data to the model

Modeling

Process

Phase 1

Train a model on Emotion vocabulary

- Emotion + Sentiment data from the NRC Lexicon
- *Supervised*

Feed comment data to the model

Modeling

Process

Phase 1

Train a model on Emotion vocabulary

- Emotion + Sentiment data from the NRC Lexicon
- *Supervised*

Phase 2

Feed comment data to the model

- Comment text data from YouTube + Reddit
- *Unsupervised*

Modeling

Part I: Sentiment

Modeling

Part I: Sentiment

Baseline Accuracy:

Modeling

Part I: Sentiment

Baseline Accuracy:

negative	0.597806
positive	0.402194

Modeling

Part I: Sentiment

Baseline Accuracy:

```
negative    0.597806
positive    0.402194
```

Overall Accuracy:

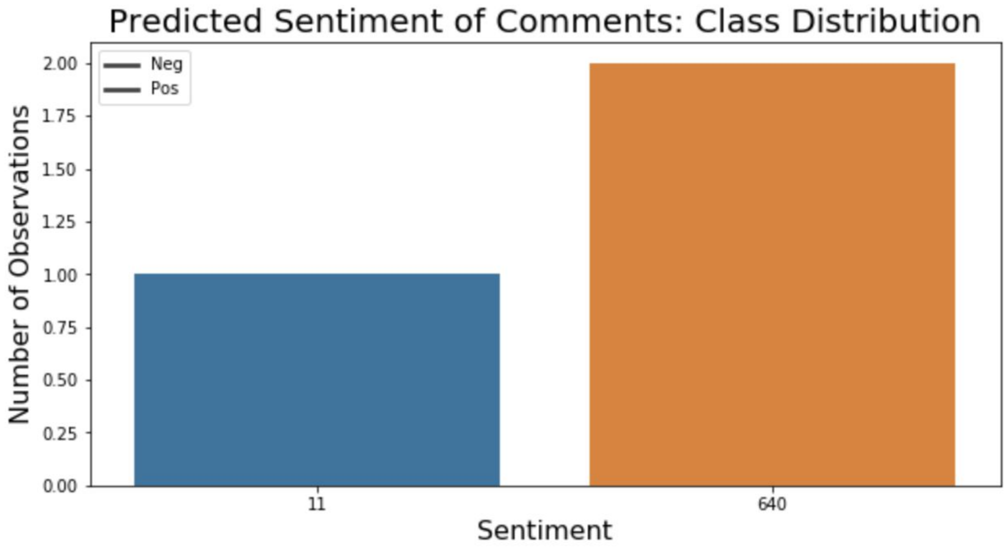
```
Sentiment Train Score:  0.9919773478055687
Sentiment Test Score:  0.887682869279613
```

Modeling

Part I: Sentiment

Test Data: Class Distribution

Negative: 11 comments
Positive: 640 comments



Modeling

Part II: Emotion

Modeling

Part II: Emotion

Approach # 1

Modeling

Part II: Emotion

Approach # 1

Multiclass Classification

Predict all emotion categories at the same time

Modeling

Part II: Emotion

Approach # 1

Multiclass Classification

Predict all emotion categories at the same time

Baseline Accuracy:

fear	0.180204
anger	0.150836
trust	0.147677
sadness	0.140613
disgust	0.124814
joy	0.102602
anticip	0.094424
surprise	0.058829

Modeling

Part II: Emotion

Approach # 1

Multiclass Classification

Predict all emotion categories at the same time

Baseline Accuracy:

fear	0.180204
anger	0.150836
trust	0.147677
sadness	0.140613
disgust	0.124814
joy	0.102602
anticip	0.094424
surprise	0.058829

Train-Test Split Accuracy:

```
Emotion Train Score:  0.6509293680297398
Emotion Test Score:  0.2382899628252788
```

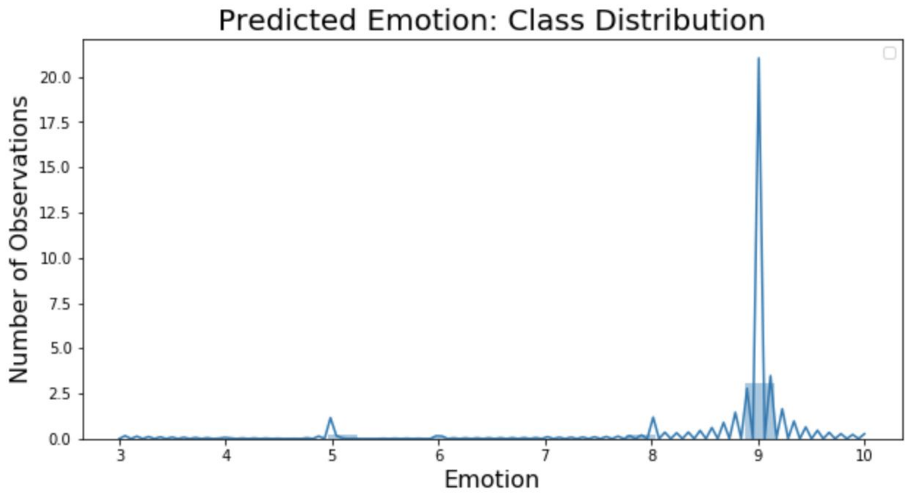
Modeling

Part II: Emotion

Approach # 1

Multiclass Classification

Predict all emotion categories at the same time



Modeling

Part II: Emotion

Modeling

Part II: Emotion

Approach # 2

Modeling

Part II: Emotion

Approach # 2

Multiple Binary Classifications

One binary regression per emotion category

Modeling

Part II: Emotion

```
Emotion Train Score, Target_4:     0.8807930607187113
Emotion Test Score, Target_4:      0.8182156133828996

Emotion Train Score, Target_5:     0.9313506815365551
Emotion Test Score, Target_5:      0.846096654275093

Emotion Train Score, Target_6:     0.8947955390334572
Emotion Test Score, Target_6:      0.8412639405204461

Emotion Train Score, Target_7:     0.9017348203221809
Emotion Test Score, Target_7:      0.866542750929368

Emotion Train Score, Target_8:     0.929368029739777
Emotion Test Score, Target_8:      0.8680297397769516

Emotion Train Score, Target_9:     0.9240396530359356
Emotion Test Score, Target_9:      0.8947955390334572
```

Approach # 2

Multiple Binary Classifications

One binary regression per emotion category

Modeling

Part II: Emotion

```
Emotion Train Score, Target_4:     0.8807930607187113
Emotion Test Score, Target_4:      0.8182156133828996

Emotion Train Score, Target_5:     0.9313506815365551
Emotion Test Score, Target_5:      0.8460966654275093

Emotion Train Score, Target_6:     0.8947955390334572
Emotion Test Score, Target_6:      0.8412639405204461

Emotion Train Score, Target_7:     0.9017348203221809
Emotion Test Score, Target_7:      0.8665427509293680

Emotion Train Score, Target_8:     0.929368029739777
Emotion Test Score, Target_8:      0.8680297397769516

Emotion Train Score, Target_9:     0.9240396530359356
Emotion Test Score, Target_9:      0.8947955390334572
```

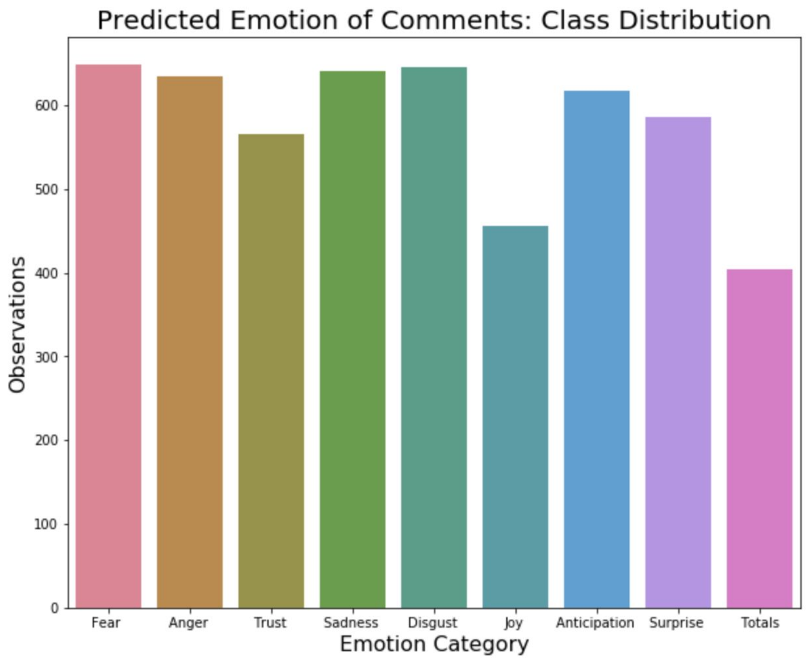
Approach # 2

Multiple Binary Classifications

One binary regression per emotion category

Modeling

Part II: Emotion



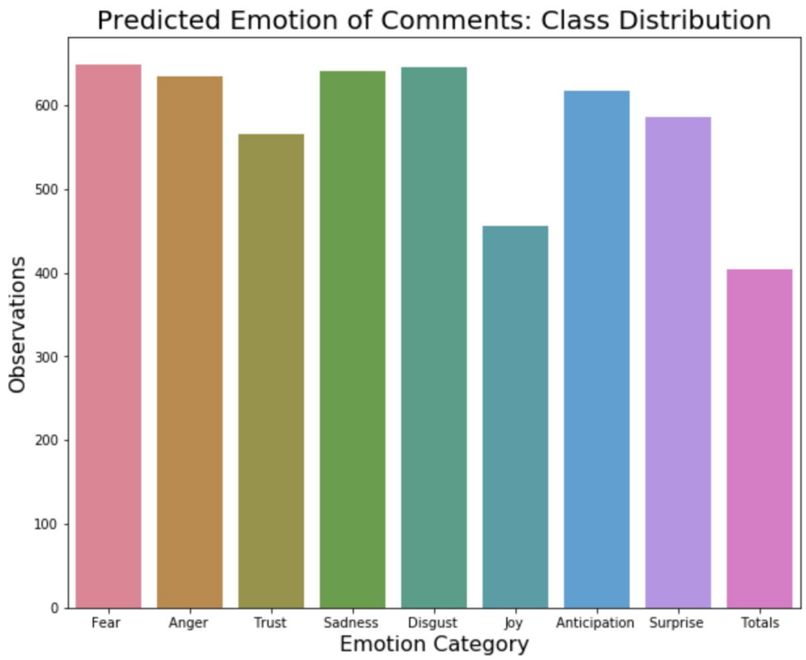
Approach # 2

Multiple Binary Classifications

One binary regression per emotion category

Modeling

Part II: Emotion



Approach # 2

Multiple Binary Classifications

One binary regression per emotion category

```
649 comments associated with Fear
634 comments associated with Anger
565 comments associated with Trust
640 comments associated with Sadness
646 comments associated with Disgust
456 comments associated with Joy
617 comments associated with Anticipation
586 comments associated with Surprise
```

Modeling

Conclusion:

Modeling

Conclusion:

- NLP and logistic regression **can** be used to model emotional sentiment in text data with reasonably reliable results.

Modeling

Conclusion:

- NLP and logistic regression **can** be used to model emotional sentiment in text data with reasonably reliable results.

Suggestions for Future Iterations:

Modeling

Conclusion:

- NLP and logistic regression **can** be used to model emotional sentiment in text data with reasonably reliable results.

Suggestions for Future Iterations:

- Utilize advanced modeling techniques, such as clustering and neural networks
- Incorporate token coefficients into the results, in order to create a valence of emotional intensity
- Experiment with different preprocessing and embedding methods, such as sentence2vec, in order to improve accuracy

Thank You.