Emotion Detection

Angela Ramirez, Christopher Garcia-Cordova, Mamon Alsalihy

Sentiment Analysis

Supervised Multi-class classification task

Can be modeled as learning probabilities of the following form:

$P(y_i \mid x)$, where $\boldsymbol{x}$ is the **text** and $y_i \in \{y_1, y_2, \ldots, y_n\} = \mathbf{Y}$ the **emotion set**.

Example

Text	Emotion
I was in line paying for groceries when i found out i was $4 short someone behind me volunteered to pay the rest for me.	Grateful
I am very happy to have been first over 300 students during this year's at my engineering school.	Joyful
Last night I felt extremely guilty I ate 12 Big Mac's from McDonalds.	Guilty

Motivation

We are investigating the efficacy of the empathetic dialogue system’s ability to understand emotions.

How?

We do this by using the same data to classify emotions.

Why does that matter?

The point is that if the classifier does well or poorly to detect emotions then what does that say about the dialogue system’s ability to identify these emotions and its ability to generate appropriate emotional responses to the speaker.

Related Work

The Circumplex Model of Affect: An Integrative Approach to Affective Neuroscience, Cognitive Development, and Psychopathology

- *Clinicians and researchers found that individuals do not experience, or recognize, emotions as isolated, discrete entities, but that they rather recognize emotions as ambiguous and overlapping experiences*

MIME: MIMicking Emotions for Empathetic Response Generation

- *We argue that empathetic responses often mimic the emotion of the user to a varying degree, depending on its positivity or negativity and content.*

CAiRE model fined-tuned for dialogue emotion detection and empathetic response generation.

- *However, data-driven end-to-end empathetic chatbot currently suffers from two limitations: 1)model capacity and 2) the sparsity of data for both emotion recognition and empathetic response generation (Rashkin et al., 2019).*
- Created a model that out performed the Empathetic Data approach

Dataset

- Created using the ParlAI platform to interact with Amazon Mechanical Turk Workers, **810 US workers**, a pair of workers are asked to
    1. Select an emotion word each and describe a situation where they felt X emotion (Prompts)
    2. Talked to someone about the situation for Y amount of rounds (Utterances)

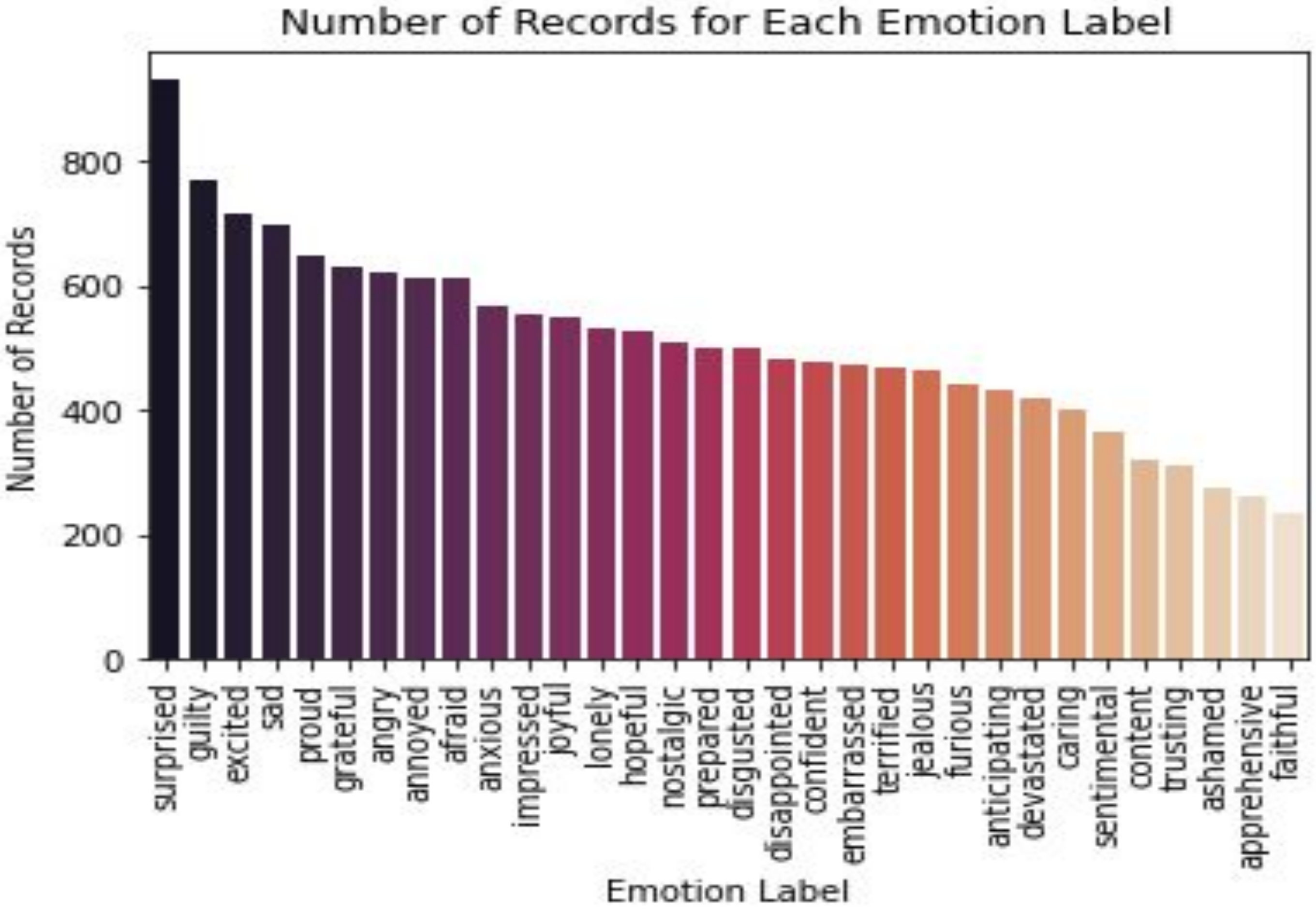
- Median # of conversations per worker was 8, while some averaged 61, for quality check they **manually checked random subsets of conversations from most-frequent workers**

Dataset

Emotion Space: 32 emotional labels

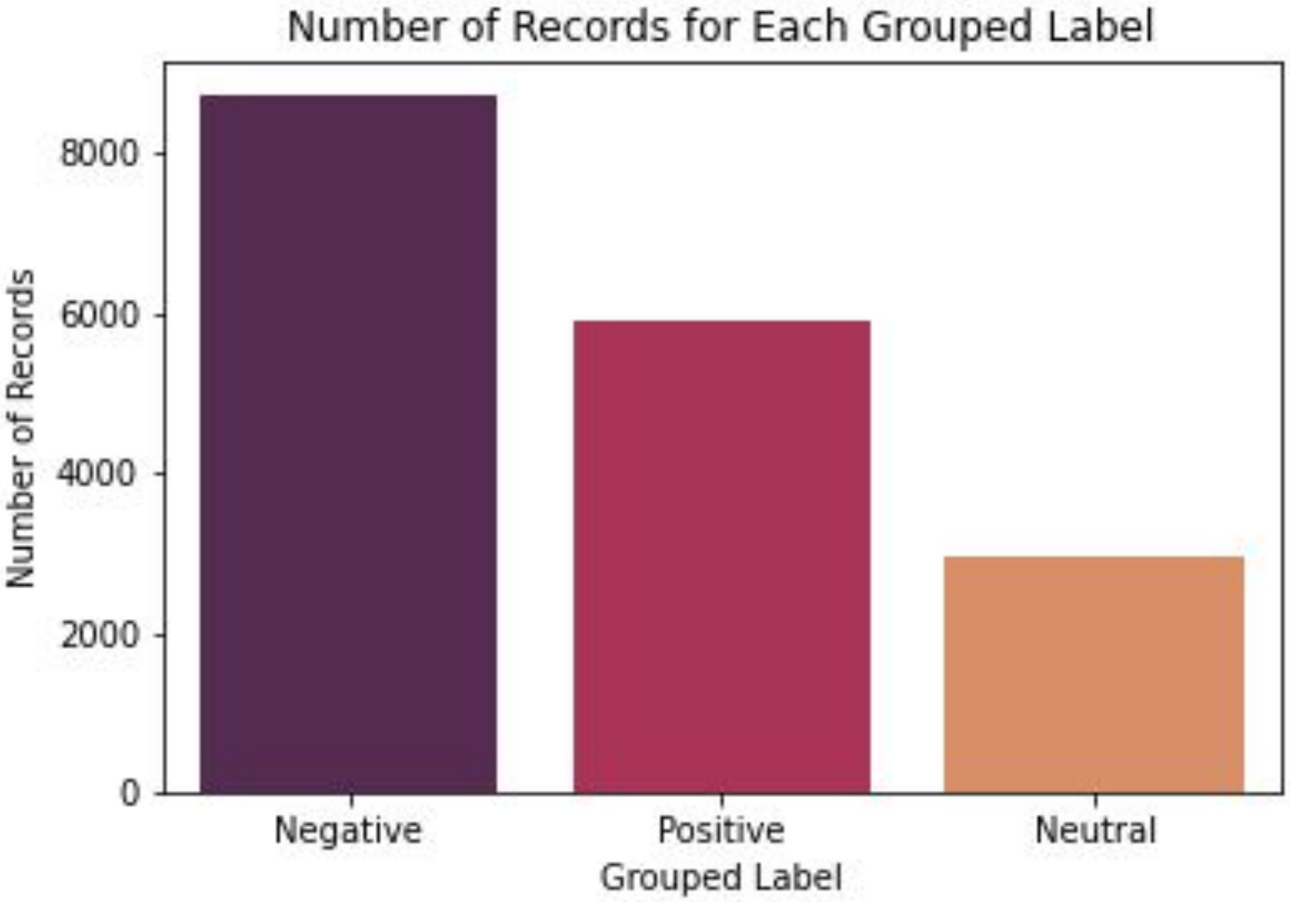
'caring', 'anticipating', 'proud', 'disgusted', 'surprised', 'apprehensive', 'ashamed', 'terrified', 'nostalgic', 'anxious', 'confident', 'annoyed', 'sentimental', 'angry', 'lonely', 'trusting', 'prepared', 'joyful', 'jealous', 'faithful', 'content', 'disappointed', 'sad', 'devastated', 'hopeful', 'afraid', 'embarrassed', 'excited', 'grateful', 'impressed', 'guilty', 'furious'

Dataset



Number of Records for Each Emotion Label
Number of Records
0
200
400
600
800
surprised
guilty
excited
sad
proud
grateful
angry
annoyed
afraid
anxious
impressed
joyful
lonely
hopeful
nostalgic
prepared
disgusted
disappointed
confident
embarrassed
terrified
jealous
furious
anticipating
devastated
caring
sentimental
content
trusting
ashamed
apprehensive
faithful
Emotion Label

Dataset


Number of Records for Each Grouped Label
Number of Records
8000
6000
4000
2000
0
Negative
Positive
Neutral
Grouped Label

Proposed Approach

- Evaluate models’ ability to classify all 32 emotions
- Collapse emotion space to 3 emotions (positive, negative, neutral)
- Tfidf vectorization, word embeddings.
- Baseline: Support Vector Machine, Multilayer Perceptron, and Convolutional Neural Network
- Metrics: macro F1, precision, recall, and accuracy.

Emotion2Sentiment Mapping

Methodology

Intuition/common-sense based; ambiguous/borderline-cases decided by majority vote.

Intuition of emotion groupings was similar to word similarity groupings from spacy. e.g. angry - furious 0.74, outlier -> e.g. surprised - disappointed 0.844

Emotion2Sentiment Mapping

Condensed Emotion	Mapped To
Negative	afraid, angry, annoyed, anxious, apprehensive, ashamed, devastated, disappointed, disgusted, embarrassed, furious, guilty, jealous, lonely, sad, terrified
Neutral	anticipating, faithful, hopeful, nostalgic, prepared, trusting
Positive	caring, confident, content, excited, grateful, impressed, joyful, proud, sentimental, surprised

Experiments and Analysis

- **Experiments**
  - 3 different models (SVM, MLP, CNN)
  - Classifying the Prompts into 32 Emotion Groups that were given
  - Classifying the Prompts into Positive, Negative, Neutral (reduced emotion groups into 3 labels)
  - Classifying the Utterances into 32 Emotion Groups
  - Classifying the Utterances into Positive Negative, Neutral
- **Analysis**
  - Feature Engineering/Analysis - Linguistic Inquiry Word Count (LIWC) e.g. word count, adverbs, positive emotion words, etc.
  - Emotion similarity using word2vec

Model	Architecture	Hyperparameters	Optimizer	Loss Function
SVM	TFIDF, one v. rest	Regularization, L2	N/A	Squared Hinge
MLP	TFIDF word vector Two linear layers with drop out probability	Learning_rate: 1e^-2 Hidden Size: 200	AdaGrad	Cross Entropy
CNN	Embedding layer with dropout, three one dimensional convolution layers, max pooling.	Learning_rate: 1e^-3 Channel: 100 Kernel sizes: [3,4,5]	Adam	Cross Entropy

Results - Sentiment Labels: Prompts

Model	Precision	Recall	F1	Accuracy
SVM	.73	.64	.66	.73
MLP	.75	.75	.75	.75
CNN	.72	.72	.71	.72

Results - Sentiment Labels: Utterances

Model	Precision	Recall	F1	Accuracy
SVM	.57	.53	.53	.65
MLP	.69	.7	.69	.7
CNN	.66	.67	.66	.67

Results - 32 Emotion : Prompts

Model	Precision	Recall	F1	Accuracy
SVM	.46	.46	.44	.46
MLP	.46	.46	.46	.46
CNN	.38	.38	.38	.38

Results - 32 Emotion : Utterance

Model	Precision	Recall	F1	Accuracy
SVM	.28	.27	.26	.27
MLP	.38	.39	.38	.3
CNN	.32	.31	.31	.31

Data Analysis with SVM baseline

Confusion matrix shows SVM's difficulty in distinguishing emotions with the given features.

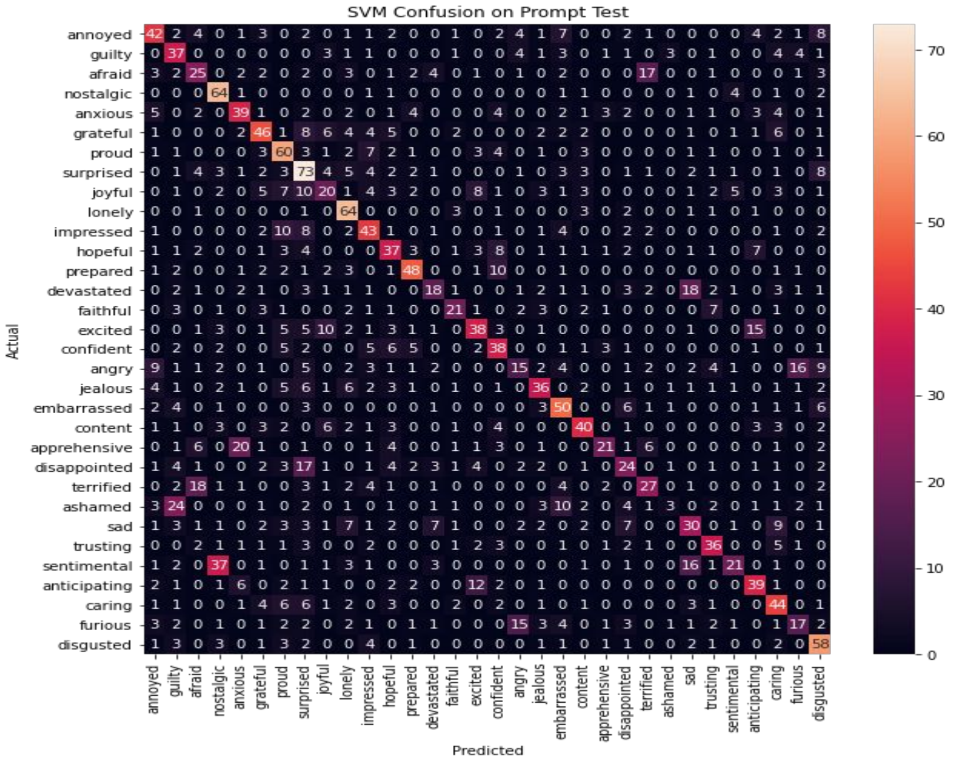
SVM trained on prompts performs better than when trained on the utterances; indicative of data quality.

Since SVM features based on BOW, confusion between emotions implies similar word distribution.

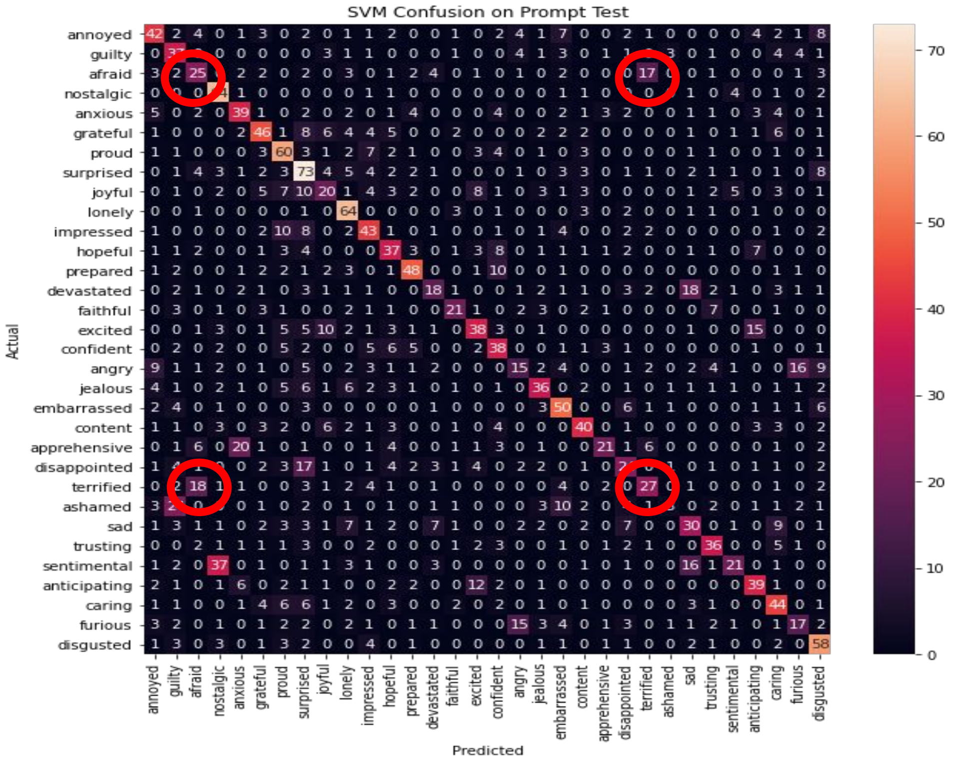
Better features, accounting for key words, word order, usage of adverbs, adjectives might help model distinguish emotions – though performance using LIWC suggests not.

Compressing emotion space improves model performance, suggesting word distribution better separates data.

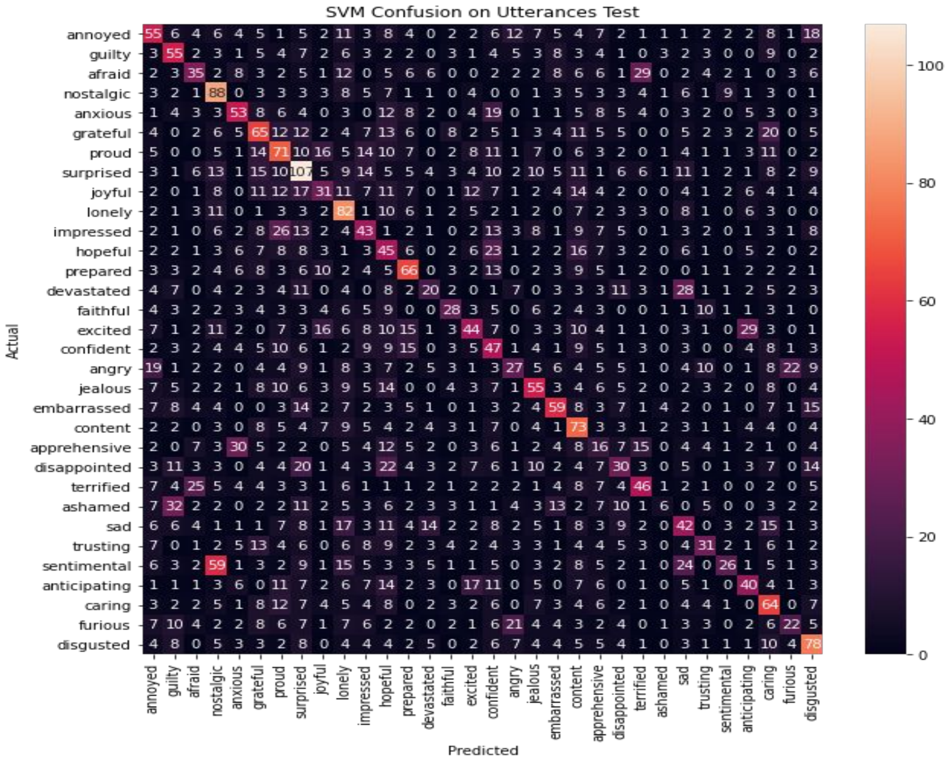
Analysis


SVM Confusion on Prompt Test
Actual
Predicted
annoyed
guilty
afraid
nostalgic
anxious
grateful
proud
surprised
joyful
lonely
impressed
hopeful
prepared
devastated
faithful
excited
confident
angry
jealous
embarrassed
content
apprehensive
disappointed
terrified
ashamed
sad
trusting
sentimental
anticipating
caring
furious
disgusted
70
60
50
40
30
20
10
0

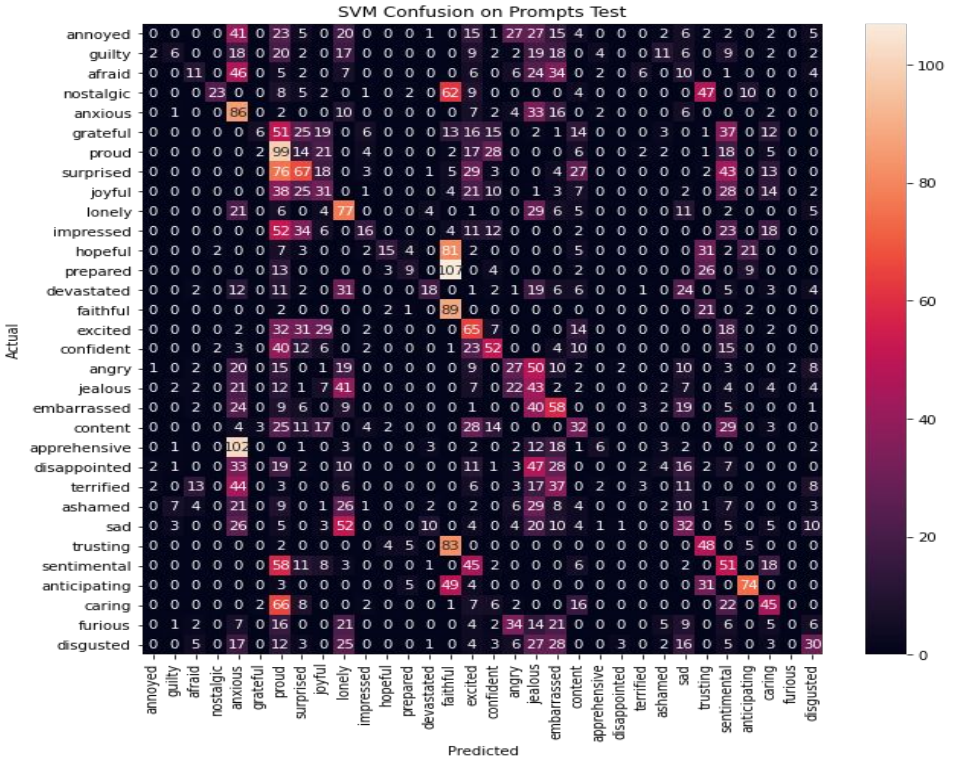
Analysis


SVM Confusion on Prompt Test
Actual
Predicted
annoyed
guilty
afraid
nostalgic
anxious
grateful
proud
surprised
joyful
lonely
impressed
hopeful
prepared
devastated
faithful
excited
confident
angry
jealous
embarrassed
content
apprehensive
disappointed
terrified
ashamed
sad
trusting
sentimental
anticipating
caring
furious
disgusted
70
60
50
40
30
20
10
0

Analysis


SVM Confusion on Utterances Test
Actual
Predicted
annoyed
guilty
afraid
nostalgic
anxious
grateful
proud
surprised
joyful
lonely
impressed
hopeful
prepared
devastated
faithful
excited
confident
angry
jealous
embarrassed
content
apprehensive
disappointed
terrified
ashamed
sad
trusting
sentimental
anticipating
caring
furious
disgusted
100
80
60
40
20
0

Analysis

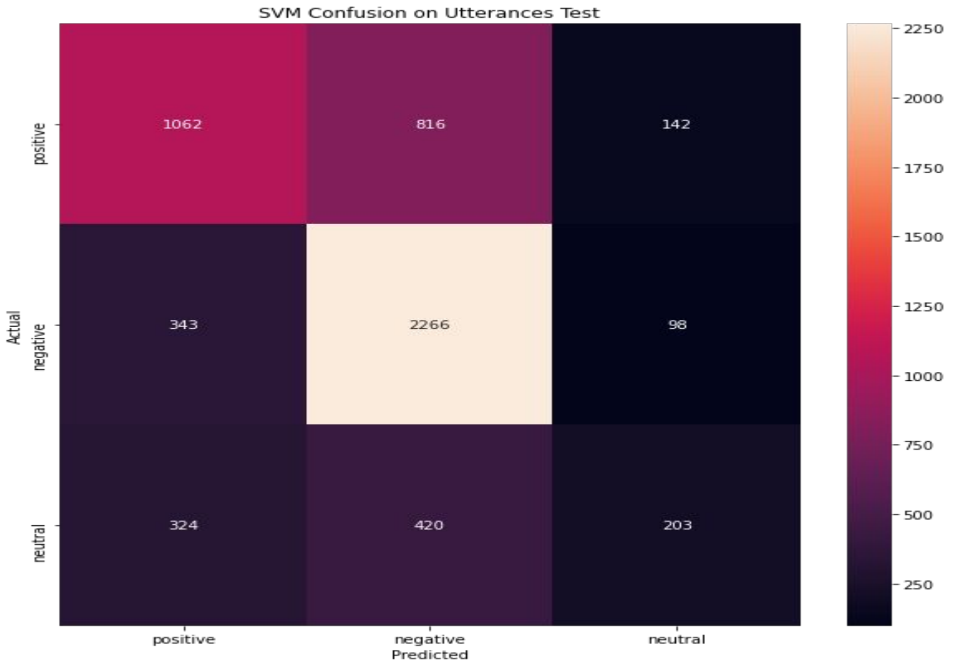


SVM Confusion on Prompts Test
Actual
Predicted
annoyed
guilty
afraid
nostalgic
anxious
grateful
proud
surprised
joyful
lonely
impressed
hopeful
prepared
devastated
faithful
excited
confident
angry
jealous
embarrassed
content
apprehensive
disappointed
terrified
ashamed
sad
trusting
sentimental
anticipating
caring
furious
disgusted
100
80
60
40
20
0

Analysis

SVM Confusion on Prompts Test
592
261
42
80
1129
13
133
148
139
positive
negative
neutral
Actual
Predicted
1000
800
600
400
200

Analysis



SVM Confusion on Utterances Test
positive
negative
neutral
Actual
1062
816
142
343
2266
98
324
420
203
positive
negative
neutral
Predicted
2250
2000
1750
1500
1250
1000
750
500
250

Conclusion

- Reducing Labels down to 3 groupings improved all model performance by ~0.3 on all F1 macro scores
- Adding additional features from LIWC and combining utterances and prompts lowered scores
- Issues observed
  - Overlap of similar emotions caused discrepancies
  - Not an equal number of positive, negative, and neutral emotion types

Citations

- Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2018. Towards empathetic open-domain conversation models: a new benchmark and dataset.
- Zhaojiang Lin, Peng Xu, Genta Indra Winata, Farhad Bin Siddique, Zihan Liu, Jamin Shin, and Pascale Fung. 2019. Caire: An empathetic neural chatbot.
- Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, Soujanya Poria. MIME: MIMicking Emotions for Empathetic Response Generation
- Posner, J., Russell, J. A., & Peterson, B. S. (2005). The circumplex model of affect: an integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and psychopathology*, *17*(3), 715–734. https://doi.org/10.1017/S0954579405050340

Question and Answers