Emotion Detection on Spoken Word Corpus

Alexandra Drossos
Anne Marshall
Chandni Shah

Problem Statement

Emotion Detection vs. Sentiment Analysis

Emotion detection has been conducted on many written word corpa:

- Our baseline is the GoEmotions model from Google Research
- Achieved an F1 score of .46

Can we achieve these same results on a spoken word corpus?

The Data - Spotify Podcast Transcripts

Randomly pulled sentences from 100,000 Spotify podcast transcripts spoken by a variety of speakers on a variety of topics
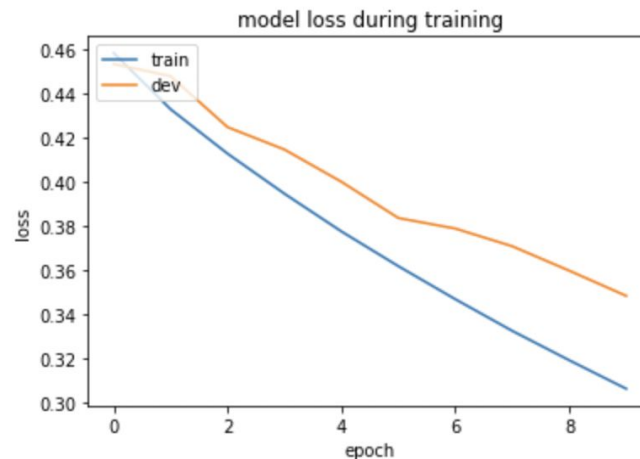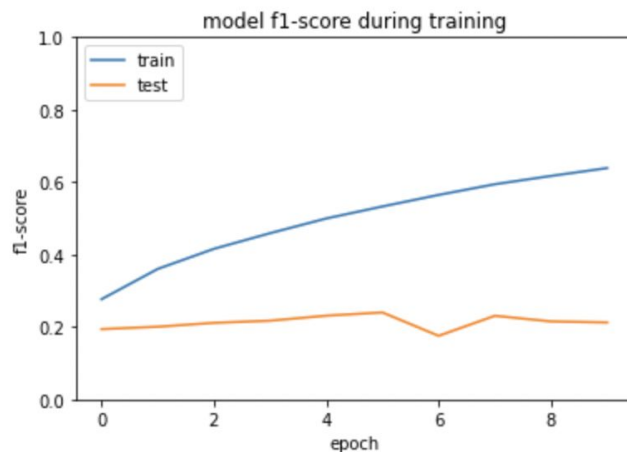
GoEmotions Taxonomy

GoEmotions Taxonomy

Positive Emotions		Negative Emotions		Ambiguous Emotions	
admiration	amusement	anger	annoyance	confusion	curiousity
desire	excitement	disgust	embarrassment	surprise	realization
love	optimism	nervousness	remorse		
approval	caring	disappointment	disapproval	**Neutral Emotion**	
gratitude	joy	fear	grief	neutral	
pride	relief	sadness			

Conducted manual emotion tagging of the selected sentences

Results

Model	Loss	Precision	Recall	F1	Binary Accuracy
Spotify Data Run on GoEmotion Model Out of Box (Zero shot Transfer)	**0.4225**	**.5833**	**.2153**	**.3146**	**.9564**
GoEmotion Model fine tuned on Tagged Spotify Data (Transfer Learning)	**0.3326**	**.4746**	**.2059**	**.2872**	**.9526**

model f1-score during training

model loss during training

Results

Model	Loss	Precision	Recall	F1	Binary Accuracy
Spotify Data Run on GoEmotion Model Out of Box (Zero shot Transfer)	.4225	.5833	.2153	.3146	.9564
GoEmotion Model fine tuned on Tagged Spotify Data (Transfer Learning)	.3326	.4746	.2059	.2872	.9526
BERT fine tuned on Tagged Spotify Data	**.4816**	**.4369**	**.3629**	**.3420**	**.9486**
RoBERTa fine tuned on Tagged Spotify Data	**.1354**	**.5772**	**.3468**	**.3735**	**.9580**

Results

Model	Loss	Precision	Recall	F1	Binary Accuracy
Spotify Data Run on GoEmotion Model Out of Box (Zero shot Transfer)	.4225	.5833	.2153	.3146	.9564
GoEmotion Model fine tuned on Tagged Spotify Data (Transfer Learning)	.3326	.4746	.2059	.2872	.9526
BERT fine tuned on Tagged Spotify Data	.4816	.4369	.3629	.3420	.9486
RoBERTa fine tuned on Tagged Spotify Data	.1354	.5772	.3468	.3735	.9580
BERT on MLM pretraining with Tagged Spotify Data	**.3186**	**.4012**	**.2702**	**.2772**	**.9473**
RoBERTa on MLM pretraining with Tagged Spotify Data	**.2104**	**.5000**	**0.3468**	**.3524**	**.9536**
GoEmotion on MLM pretrained with Tagged Spotify Data (Few Shot)	**.3338**	**.4702**	**0.2863**	**.4431**	**.9517**

Results

Model	Loss	Precision	Recall	F1	Binary Accuracy
Spotify Data Run on GoEmotion Model Out of Box (Zero shot Transfer)	.4225	.5833	.2153	.3146	.9564
GoEmotion Model fine tuned on Tagged Spotify Data (Transfer Learning)	.3326	.4746	.2059	.2872	.9526
BERT fine tuned on Tagged Spotify Data	.4816	.4369	.3629	.3420	.9486
RoBERTa fine tuned on Tagged Spotify Data	.1354	.5772	.3468	.3735	.9580
BERT on MLM pretraining with Tagged Spotify Data	.3186	.4012	.2702	.2772	.9473
RoBERTa on MLM pretraining with Tagged Spotify Data	.2104	.5000	0.3468	.3524	.9536
GoEmotion on MLM pretrained with Tagged Spotify Data (Few Shot)	**.3338**	**.4702**	**0.2863**	**.4431**	**.9517**
RoBERTa Synthetic Tag pretraining	**.6624**	**.3958**	**.3065**	**.2985**	**.9463**
RoBERTa Synthetic Tag pretraining with Tagged Spotify Data	**.3983**	**.2682**	**.1935**	**.1907**	**.9387**
RoBERTa with MLM and Synthetic Tag pretraining	**.6761**	**.3978**	**.2984**	**.2948**	**.9467**
RoBERTa with MLM, Synthetic Tag pretraining and Tagged Spotify Data	**.4029**	**.2927**	**.1935**	**.1991**	**.9410**

Zooming in on best performing model...

GoEmotion on MLM pretrained with Tagged Spotify Data (Few Shot)
- F1 Score: 0.4431
- Learning Rate: 0.00005 *(consistent with non-MLM BERT)*
- Batch Size: 25 *(consistent with non-MLM BERT)*
- Training on all layers *(consistent with non-MLM BERT)*
- Dropout layer with rate of 10% *(unique to MLM)*

While the F1-score is near the published GoEmotions model (0.46), there are limitations
- Loss: 0.3338
- Potential for overfitting

Key Takeaways

- Emotion detection is an inherently difficult task, even for human annotators
- Experimentation with fine-tuning and MLM techniques can go a long way
- We were able to achieve spoken text corpa results (F1 of 0.44) in line with the published written text corpa GoEmotions model (F1 of 0.46)
- There are important limitations to keep in mind with this model, and any model, in the emotion detection space