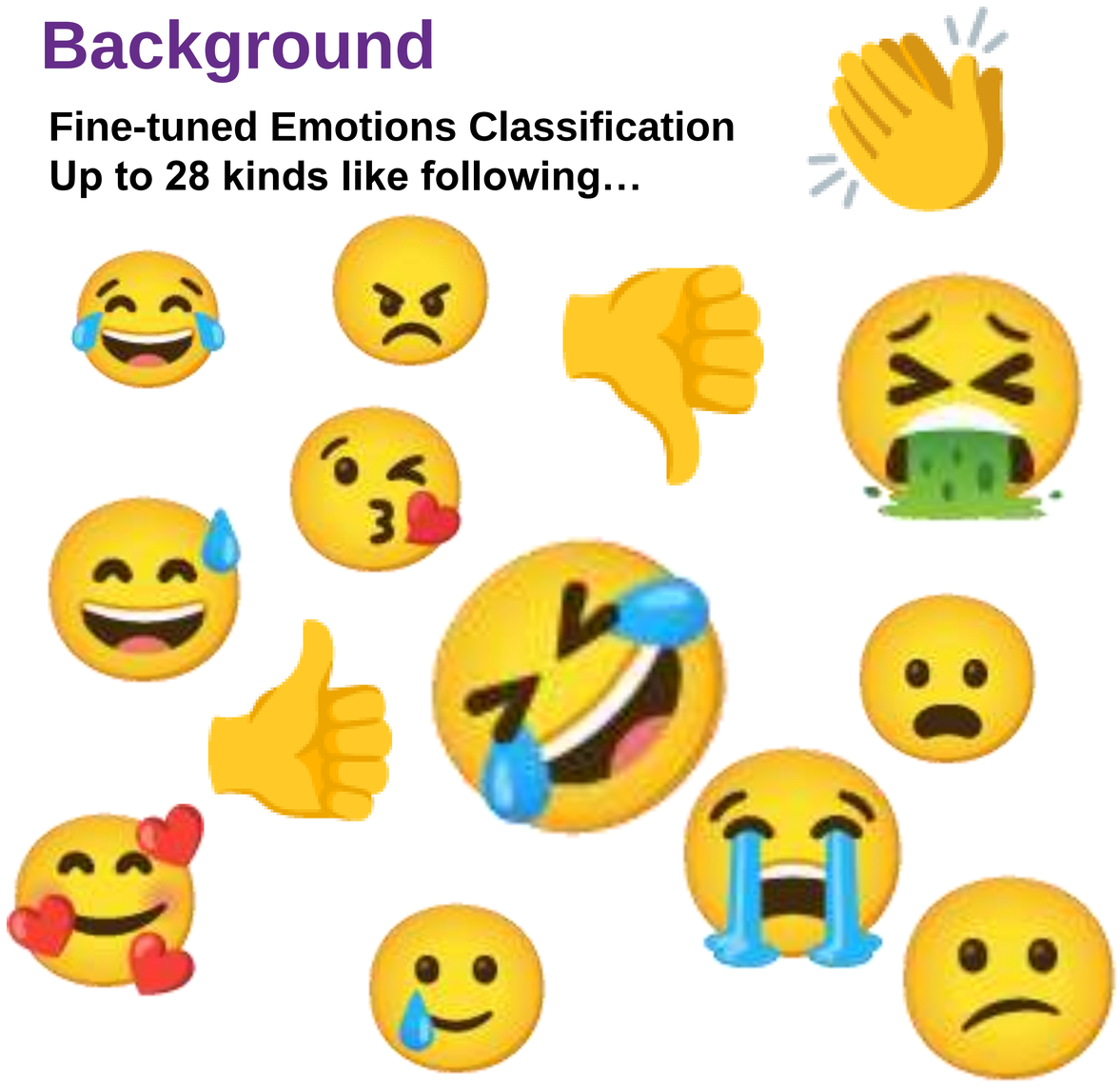
Exploring the Efficacy of BERT on a Fine-Grained Emotion Dataset : GoEmotions

Liangsheng Yin (hnyls2002@sjtu.edu.cn)

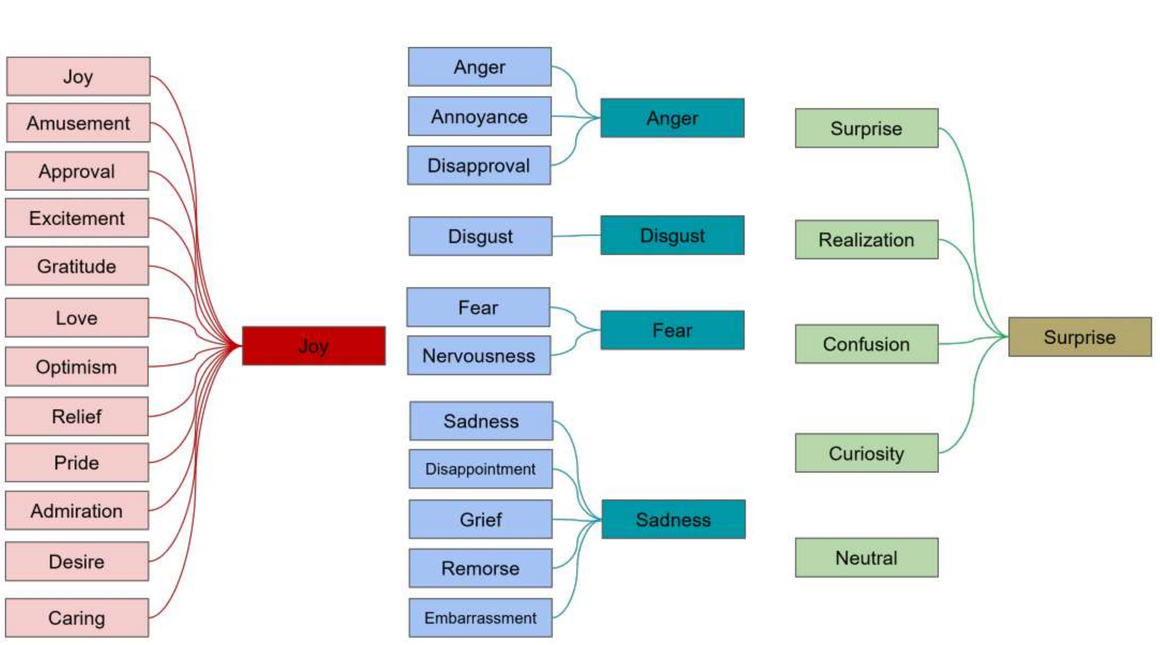
上海交通大學 SHANGHAI JIAO TONG UNIVERSITY

Background

Fine-tuned Emotions Classification
Up to 28 kinds like following…



GoEmotion Dataset



Task : figure out the emotion!

Sample Text	Label(s)
Just thinking about all the stupid stuff I did as a kid makes me very afraid for my 5 year old.	`fear`
Wow! You must have pretty low standards for teachers in your state/school! I guess that's why Massachusetts is always first for education.	`admiration, curiosity, nervouness`
I think someone else should open a drive through so I can enjoy the flavours without the dishes aha	`joy`
Lol you're an idiot.	`amusement`
Agree. Not a perfect coach but pretty, pretty good	`admiration`

Expanding Dataset : Concatenation

This book reads good. + The joke is so funny.

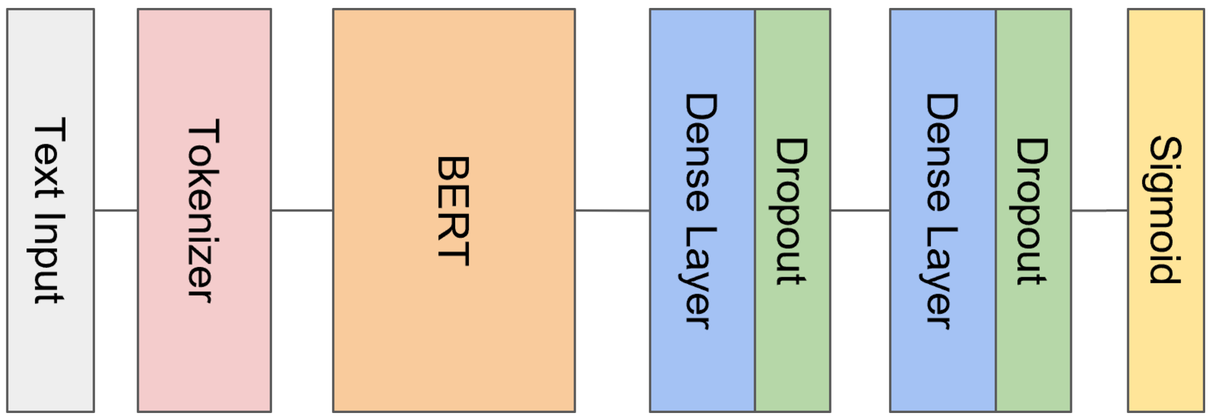
👍 plus 🤣 = 👍🤣

You are so disgusting. + You are adorable.

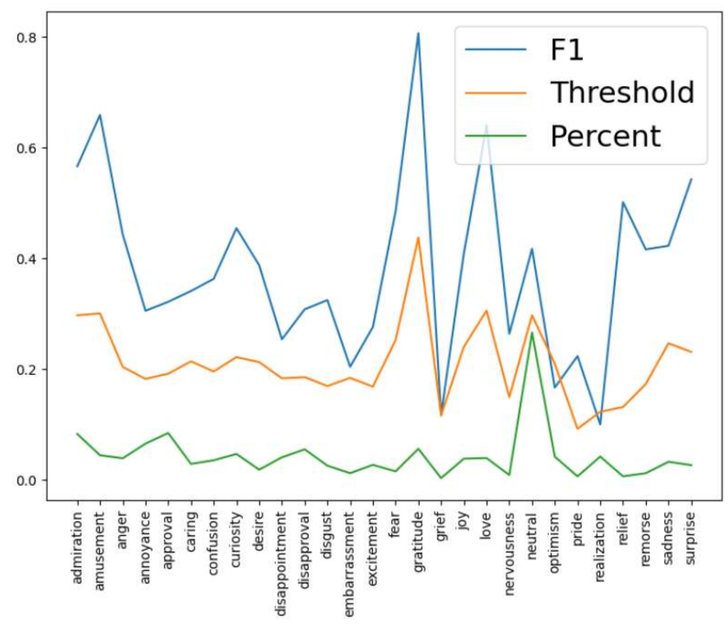
🤢 plus 🥰 = ???

Our Approach

A simple BERT-based model :
work on multi-label classification

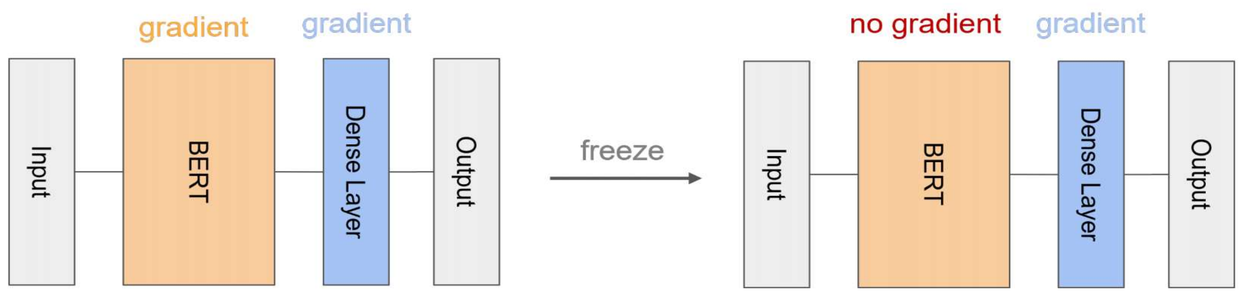


Threshold optimization



For each emotion among the 28 labels, specify its threshold, maximizing the F1 score on training set, hoping it would achieve good performance on the validation set.
The threshold result shows that some emotions such as grief , pride, relief is so challenging. And most positive emotion acquires good results!

Freeze or not ?



BERT model's parameters are so huge and well trained. If we continuously training our small dataset on BERT model, it would definitely cause overfitting. To avoid this, we apply a two-stage training for our model : 1. first not freeze the BERT model 2. freeze the BERT model, discard the parameters in our inference head, and train it again.

Experiments

Experiment	Epochs	M-F1	W-F1	Acc
DenseLayer(1) + Dynamic lr	2	0.27	N/A	0.22
DenseLayer(1) + Fixed lr(5e-5)	4	0.26	N/A	0.23
DenseLayer(1) + Fixed lr(5e-5) + TwoStage	2 + 2(frz)	0.29	N/A	0.24
DenseLayer(1) + Dynamic lr + TwoStage	2 + 4(frz)	0.29	N/A	0.24
DenseLayer(1) + Dynamic lr + Dropout(0.1) + TwoStage	2 + 6(frz)	0.29	N/A	0.24
DenseLayer(1) + Dynamic lr + Dropout(0.1) + TwoStage + OptOnThreshold	2 + 3(frz)	0.38	0.44	0.23
DenseLayer(1) + Dynamic lr + Dropout(0.1) + TwoStage + OptOnThreshold	4 + 4(frz)	0.39	0.44	0.25
DenseLayer(1) + Fixed lr(5e-5) + OptOnThreshold + ExpanedDataset	1	0.35	0.41	0.18
DenseLayer(2) + Dynamic lr + Dropout(0.1) + TwoStage + OptOnThreshold	4 + 3(frz)	0.39	0.44	0.23
DenseLayer(2) + Dynamic lr + Dropout(0.2) + TwoStage + OptOnThreshold	4 + 3(frz)	0.39	0.45	0.24
DenseLayer(1) + Dynamic lr + Dropout(0.1) + OptOnThreshold + TwoStageOnExtended	1 + 3(frz)	0.37	0.42	0.24