

# TWO IMPORTANT MODELS IN CLOSED DOMAIN QUESTION ANSWER CHATBOT

**David Tong([ytong82@berkeley.edu](mailto:ytong82@berkeley.edu))**

# TWO MODELS FOR CLOSED DOMAIN CHATBOTS



## Closed Domain Chatbots

- Designed to chat about a specific topic
- Expect for accurate answers
- Target for specific customers and economic values

## Question Answering Model

- Responsible for providing accurate answers

## Emotion Detection Model

- Detect customers emotion change
- Decide when people customer support will involve

# QUESTION ANSWERING MODEL

## Baseline Model

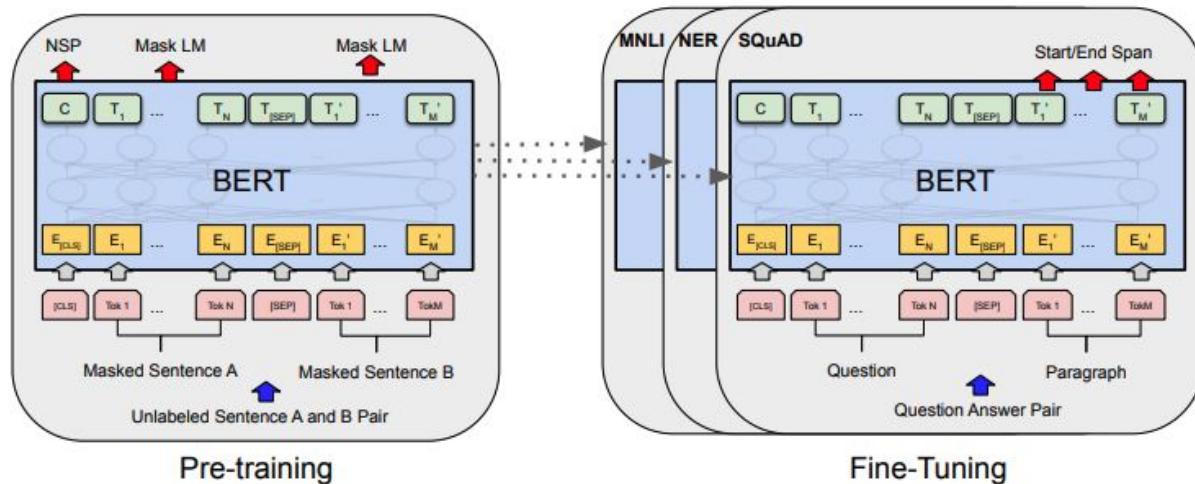
- DISTILBERT - distilbert-base-uncased

## Dataset

- SQuAD 2.0

## Metrics

- EM
- F1



# IMPROVEMENTS

## Type I Improvements

- Use model with more trainable parameters
  - ALBERT - albert-base-v2
- Use model with better training target, suitable for start and end span answer task
  - SpanBERT - SpanBERT/spanbert-large-cased

## Type II Improvements

- Use ensemble model
  - Skim-Read Model
  - Elbow Method

# HIGHLIGHT

## Skim-Read Model

- Skim Part
  - Predict if a question is answerable
- Read Part
  - Predict start and end span of an answer in the text



# RESULTS

## SQuAD2.0

The Stanford Question Answering Dataset

Model	Model Type	Improvement Type	EM	F1
DISTILBERT distilbert-base-uncased	Discriminative Model	Baseline	64.9457	68.3097
ALBERT albert-base-v2	Discriminative Model	Improved Type 1	78.2111	81.2822
SpanBERT SpanBERT/spanbert-large-cased	Discriminative Model	Improved Type 1	83.1382	86.3032
T5 t5-small	Generative Model	Improved Type 1	N/A	N/A

# RESULTS

## SQuAD2.0

The Stanford Question Answering Dataset

Skim-Read Model David-Tong/squad2-s kim-read-predictor Skim Model Logistic Regression	Ensemble Model	Improved Type 2	83.381	86.519
Skim-Read Model David-Tong/squad2-s kim-read-predictor Elbow Method	Ensemble Model	Improved Type 2	84.0984	87.0521

# EMOTION DETECTION MODEL

## Baseline Model

- 1-stage classification model

## Dataset

- GoEmotions

## Metrics

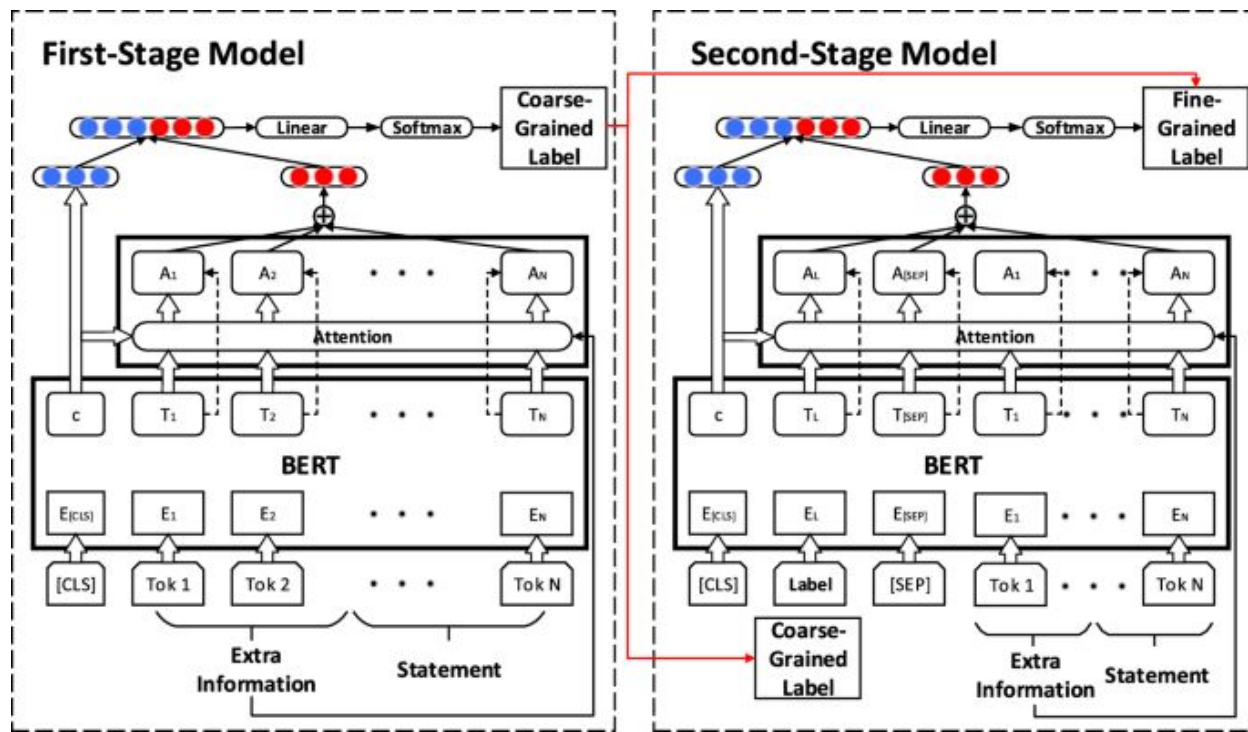
- Accuracy
- Macro Avg
- Weighted Avg





# IMPROVEMENTS

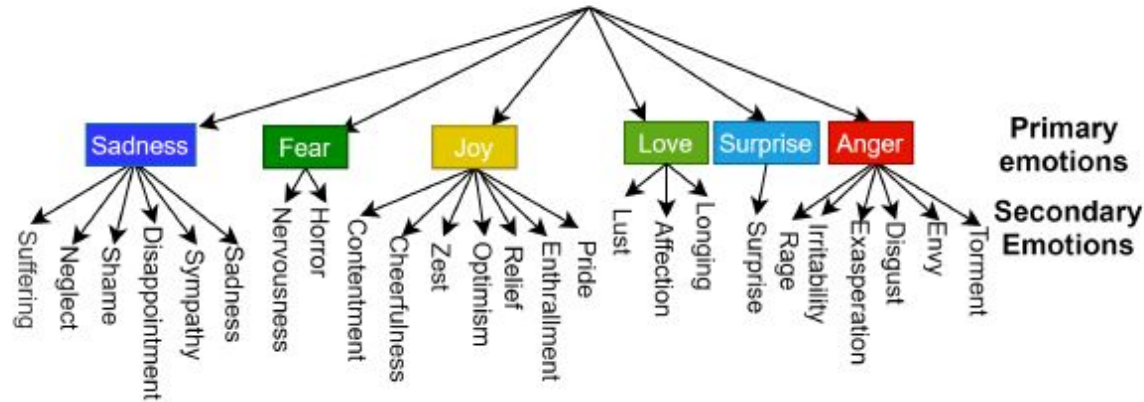
## 2-Stage Classification Model



# IMPROVEMENTS

## Type I Improvement

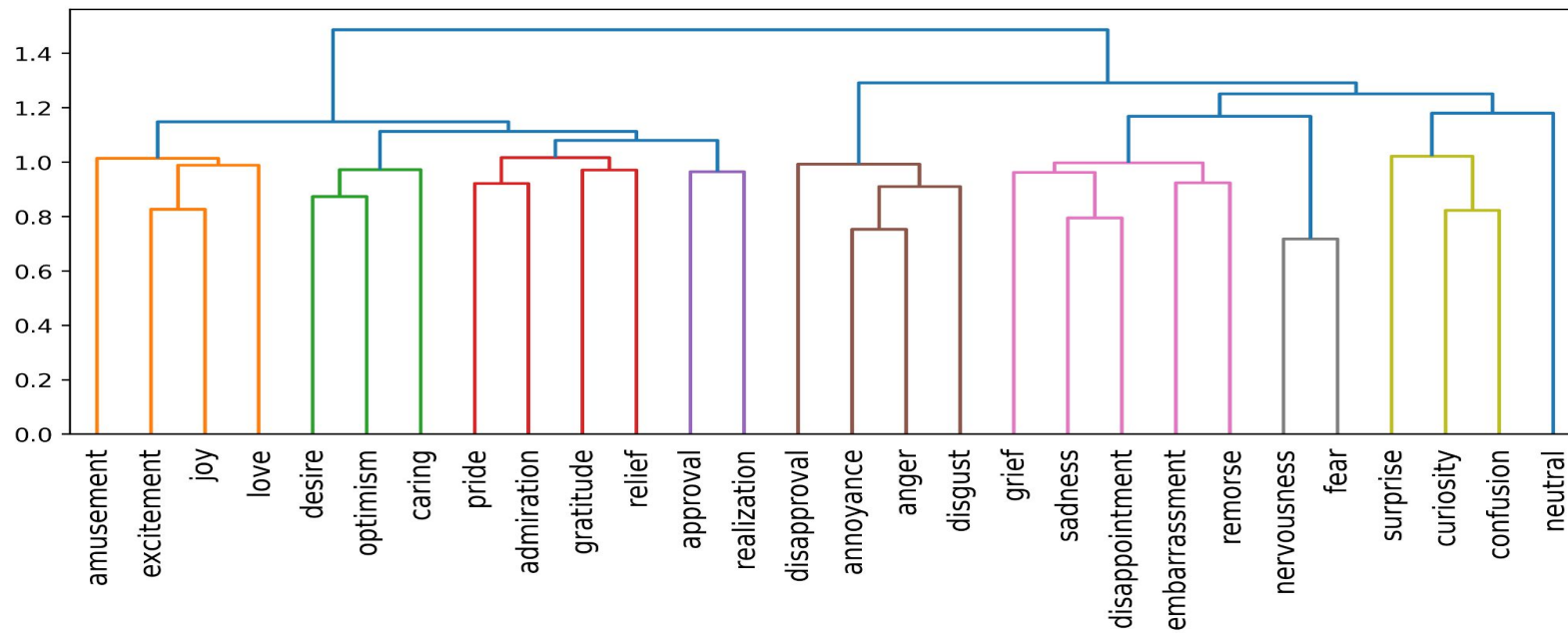
- 2-stage classification using Parrot's emotion model



# IMPROVEMENTS

## Type II Improvement

- 2-stage classification using data clustering model



# RESULTS

Two-Stage classification doesn't bring improvement in performance but even a little downgrade in performance.

Data based clustering has a different model from psychological emotion classification model.



# RESULTS

Model		F1		
		accuracy	macro avg	weighted avg
One-Stage Classification bert-base-cased	Baseline	0.60	0.49	0.59
Two-Stage Classification (Parrott's Emotions model)	Improved Type 1	0.57	0.48	0.57
Two-Stage Classification (Data Clustering Modell)	Improved Type 2	0.58	0.45	0.57

# LOOK FORWARD

## Question Answering Model

- Use generative models like GPT or T5

## Emotion Detection Model

- Remove neutral labeled data, focus on emotion and classification
- Use other clustering method and compare clustering results with existing psychological classification models
- Cast emotions to a 2D plane and watch relationships among them.