

# CURTIS - Cure by Therapy Intelligent System

Kaushik Shakkari

## Objective:

The objective of this challenge is to produce short empathic contextual responses/reflections for user problems/question/queries for a context.

## Example:

**Question Title (Question):** I don't know why my mother cannot trust me

**Question Text (Question Context):** My mother started hating me. She doesn't believe in me and keeps scolding me. I also started hating her

**Expected Response (from bot):** It seems like you are having conflicts with someone important in your life.

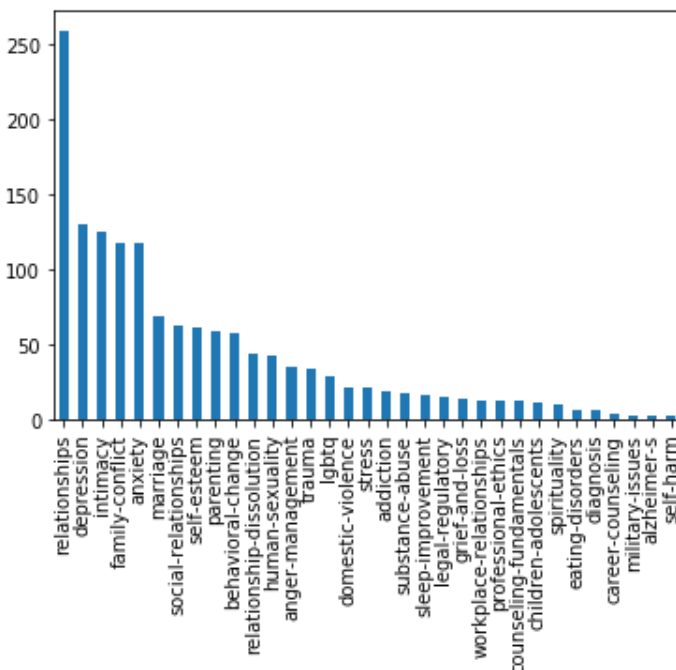
## Data Extraction:

I have written a python script to retrieve all questions' urls from **councilchat.com**. . Then I have retrieved all the answers from therapists using questions' urls. There are totally **12139** therapists' answers from **33** different topics. Topics include depression, anxiety, stress and parenting etc.

**Second Dataset is from Kaggle** ([Therapist Q&A](#)):

## Challenges in using this dataset for modelling:

- In spite of **14268** therapist answers, there are only **831** unique questions. Selecting one therapist's answer for a question to extract reflection is difficult.
- Topic Distribution is following **Zipfian distribution**. Imbalance dataset is not good for modeling if we are using a topic as a feature for generating reflections.



X-axis: Topic Name

Y-axis: Number of questions in respective topic

- I have observed that the same question appears on different topics.

Questions	Topics
Can I change my feeling of being worthless to everyone?	Self-esteem and Depression
How can I be less anxious when dealing with my toddler?	Anxiety and Parenting
How can I deal with the stress of my mother?	Family-conflict and Stress

### Preprocessing to solve above challenges to prepare data for modelling:

- I used upvotes and views as features to select one answer from multiple therapist's answers for each question. I grouped all answers by questions, sorted answers in descending order by upvotes and views, and selected the therapist answer with highest upvotes and views.
- Reflections can be extracted from the answerText feature in the CouncilChat dataset. After observing data I found that in most of the cases a good reflection can be the first sentence in answerText (therapist answer for a question). However, I considered the second sentence if the first sentence is less than 3 words or if the first sentence is a greeting ("hello" or "hey kaushik"). Moreover, if the therapist's message has "Seems like" or "Sounds like", I used that sentence as the reflection.
- To avoid **Zipfian distribution** while modeling, I have created a new feature called root\_topics. The root topic only consists of 3 topics.

Topics	Root Topics
relationships,intimacy,family-conflict,parenting,relationship-dis solution,marriage,domestic-violence	Relationship Conflicts
depression,anxiety,stress,anger-management,trauma	Emotion Conflicts
All other minor categories	Others

- Created a multi label feature for root topics. For example, "How can I deal with the stress of my girlfriend?" gets [1,1,0] as the label because the question can be classified as both Relationship Conflicts and Emotion Conflicts category. This is useful for the baseline model.

### Modeling:

After understanding the problem, I came up with 3 main ideas to generate good reflections for different or new journalizing entries. For all 3 models, I am using deep contextual bidirectional

sequence to sequence models like BERT, XLNet, and Roberta. I have completed implementing Model 1 and Model 2. Model 3 is currently in progress.

Any seq-seq model requires a good word and sentence representation for any supervised or unsupervised task. In our case, It is very important to capture the context of user journalizing entry. For example, let us consider the following journalizing entries.

Journalizing Entry / Question Title	Context 1	Context 2
I become anxious when I go to the <b>apple store</b> .	Company Store	Fruit Store
I am scared of <b>bats</b> . I get bad dreams of <b>bats</b> .	Animal	Sports
I want to <b>break</b> up / Anxiety caused me to <b>break</b> my phone.	Relationship	Destroy

The context or questionText from the dataset will help us to give a better representation of words and sentences while building the model.

### Model 1: (Fine Tuning Multi-label Multi-class Classification Model) - My Baseline Model

#### Description:

This model is similar to finding the intent of the query and generating an intent specific response. Given a **Question Title** and **Context/Question Text**, I built a model to do multiclass classification using fine-tuning. I am adding an output layer (768 nodes \* the number of labels) at the end of the “**BERT BASE UNCASED**” pretrained model which uses 110 million parameters to classify journalizing entries. The topic feature from the dataset acts as labels for classification. After classifying into a topic, topic-specific reflection is outputted.

Metrics: label ranking average precision score = 91.46 %

#### Advantages:

- It is easy to evaluate classification tasks using metrics like accuracy, precision, recall, and F1 score, etc. In our case, I am building the multilabel classification and I used “label ranking average precision score” as metrics to evaluate the model.

#### Disadvantages:

- This model can only generate a specific message for a topic. For example, both journalizing entries “*How do I get my girl to fall back in love with me?*” and “*How do I get over my breakup and concentrate on my career*” gets classified as a “Relationship Conflicts” root topic or “Relationships” topic and generic reflection like “*The decision regarding relationship issues is always a difficult one*” can be generated and it might not be apt for all entries in “relationships” topic.

## Model 2: (Contextual Similarity Model) - My Current Model

### Description:

**questionText** can act as the context of the **questionTitle**. As I explained earlier, it is very important to get the contextual representation for **questionTitle**.

- For each journalizing entry (question title) in the dataset, I generate 768-dimensional contextual vectors using the BERT pre-trained model.
- When the user types the context of the journalizing entry (question text) and the journalizing entry (question title), I generate the BERT contextual vector for the journalizing entry.
- I calculate the similarity between the user's contextual vector with all other contextual vectors in the dataset.
- The reflection of the current user journalizing entry will be the reflection of the journalizing entry with the highest similarity score in the dataset.

### Sample Reflections Generated from Journalism Entries

**\*below examples are not from the training dataset\***

Question / Question Title	Context / Question Text	Reflection from model	Confidence Score (0-1)
Why am I feeling uncomfortable at social situations, especially parties?	I am an introvert and feel anxious in parties, for example I was at an event at USC last Friday, and after 20 min I left because I felt uncomfortable with all the new people there and I just wanted to be by myself. Why is that?	Oh, that is so frustrating! Have you ever heard of social anxiety? Sounds like you may be suffering from it	0.9182
My partner is forcing me to do sex. I cannot do that often.	I recently got married. My wife is not happy with our sex life.	Hi, First and foremost, I want to acknowledge your efforts to gain (your) ideal erectile function	0.8491
Why am I having a stressful life?	I am very stressed in my life. I am spending 10 hours at the office. At same time I have a daughter and I need to look after her. Moreover, my dad is an Alzheimer's patient and needs to look after him.	It sounds like you have several different things happening at the same time.	0.8459

"Are you selling machine gun?"	"Machine Gun is soo good"	You seem to have been a very positive influence in your kids lives	0.6594
--------------------------------	---------------------------	--	--------

From the above table, we can see that the reflection from the model for the last response is irrelevant. We can also observe that confidence for such reflections is low compared to other relevant reflections.

#### **Advantages:**

- Unlike the baseline model, this model will generate journalism entries specific reflections.

#### **Disadvantages:**

- It is very challenging for any unsupervised learning task to evaluate the model performance without some feedback from users.
- Currently, I am using the highest similarity score to predict the reflection. However, I need users' feedback to compute the threshold statistically and analyze if the highest similarity score is greater than or less than the threshold. Later, I can perform error analysis on False Positives and False Negatives to tune the model.

#### **Model 3: (Reflection Generative Model) - In progress**

The first model is a retrieval-based model and the second model is the context similarity-based model. We can also build a sequence to sequence generative models. The architecture of this model is similar to language translation models. However, we have fewer data points to build the generative model. Currently, I am exploring more websites like CouncilChat to scrape more data.

#### **Appendix:**

##### **"BERT BASE UNCASSED" pretrained model architecture:**

This architecture has 3 different embeddings and 12 encoder layers. The three types of embeddings include token, position, and segment embeddings. Each encoder layer has 12 heads which capture different kinds of attention. Totally this architecture uses 110 million parameters.

Vocabulary size (T) = 30522 words (Number of tokens derived from tokenizer)

Tokenizer: Word piece : Handles tokens which are seen for first time also

Eg: Playing -> Play, ##ing

Number of Transformers layers (L) = 12

Multi Headed attention (A) = 12 heads

Embedding Dimension (H) = 768

Feed Forward Dimension (FFD) =  $4 * H = 3072$

Sequence Length = Max number of tokens in a sentence (S) = 512

Positions to be encoded (P) = 512

Query(Q), Key(K), Value(V) = 64

Type	Parameter	Dimensions	# of parameters	Aggregate
<b>EMBEDDINGS</b>	Pos Emb Weight	$P * H$	393, 216	23,837,184
	Segment Embedding	$2 * H$	1536	
	Embedding Weight	$T * H$	23,440,896	
	Layer Norm - Weight	H	768	
	Layer Norm - Bias	H	768	
<b>Encoder Layer</b> <b>(* 12)</b>	Query.Weight * 12	$H * Q * 12$	589,824	2,363,904 * 12 = 28,366,848
	Query.Bias * 12	$Q * 12$	768	
	Key.Weight	$H * K * 12$	589,824	
	Key.Bias	$H * 12$	768	
	Value.Weight	$H * V * 12$	589,824	
	Value.Bias	H	768	
	Combine 12 into 1 (W)	$12 * Z(64) * H$	589,824	
	Combine 12 into 1 (B)	H	768	
	Layer Norm - Weight	H	768	
	Layer Norm - Bias	H	768	
<b>Feed Forward</b> <b>(* 12)</b>	Layer 1 - Weight	$H * FFD$	2,359,296	4,723,968 * 12 = 56,687,616
	Layer 1 - Bias	FFD	3072	
	Layer 2 - Weight	$FFD * H$	2,359,296	
	Layer 2 - Bias	H	768	

	Layer Norm - Weight	H	768
	Layer Norm - Bias	H	768