Therapize - A Therapy Chatbot

Divyansh Oze

University of Nottingham
ppxdo1@nottingham.ac.uk

May 10, 2022

Abstract

Implementation of a transformer-based fine-tuning and an Artificial Neural Network model using a bag_of_words approach to develop an Intelligent Agent capable of holding conversations with people in distress. Drawn are the workflows for agents in each of the approaches. Evaluation results show how the GPT-2 fine-tuned model can survive some esoteric-deep conversations than the ANN model, which performs decently on the limited corpus. It also concludes with some future applications for the therapy chatbot.

Keywords: Therapy Chatbot, Intelligent Agent, Transformers, Neural Network

I. Introduction

With the rising need for mental health support in the entire world, there has only been a massive rise in chatbot technology. Apart from having a personal humanitarian side for mental health issues, the project aims at building a therapeutic-intelligent agent that can be used in terms of distress and angst for a user to interact within an environment. Although there is a good supply of qualified professionals seeking help related to mental health, there is still an agonising wait for the patients to get an appointment. The idea of conversational chatbots helps ease the pain and presents a low-barrier way to reach out for what can be an intimidating and even life-treating problem in life.

A review of the psychiatric landscape [Aditya Nrusimha Vaidyam, 2019] considers the heterogeneity involved while studying this extensive topic for the use of chatbots in psychiatric consultancy. It states some evidence on how well researched and programmed chatbots can work as Therapeutic-intelligent agent to help in the psychiatric treatment.

This project uses two approaches to developing an intelligent chatbot as a mental-health application. It uses the bag of word [Deepu S,Pethuru Raj and S.Rajaraajeswari (2016)] approach to train an artificial neural network and fine-tune the GPT2 [Ashish Vaswani,2017] for another chatbot prototype.

II. Data Description

There has been a usage of two datasets in developing two different chatbots:

i. Intents

This .JSON file is created manually by storing tags as to what the question is about on a limited scale, different patterns under the specific tag, and the responses for such pattern-based inputs.

ii. Therapy Dataset

This consists of questions people have put up on a platform with qualified therapists responding to those questions. They do not have perfect responses, but they help us with domain-specific expert knowledge.

III. Methodology

The section will explain the networks used in approaching the problems, The two approaches used in developing two chatbot models:

- GPT-2 using Transformers
- Artificial Neural Network using bag_of_words

i. GPT-2 Fine-Tuning

The Transformer-based model network uses multiple attention layers to predict the next token for a sequence. The multiple attention layers help get the global dependencies for input processing. The Transformer network can hold the contextual knowledge for long sequences without falling prey to the vanishing gradient descent problem.

It contains an encoding component and a decoding component. These components are just a collection of encoders and decoders, each of them is one connected to a multi-headed attention layer and a feed-forward neural network, as shown in Figure[1]. The multi-head attention layer employs a self-attention method, which feeds the input to three vectors, i.e. query (Q), key (K), and value (V)]

i.1 Data Acquisition

The model starts with reading the therapy_data .JSON file, converting it into a data frame after loading to efficiently perform actions on the file.

i.2 Data Preparation

We iterate through the data frame to get the column for the questions people in distress have asked the qualified professionals on the platform and the responses to the respective questions.

After preparing the data in a question/answer format, data is stored as a text file to pass into the fine-tuned GPT-2.

i.3 Fine-Tuning GPT-2

The GPT-2 model takes in the parameters mentioned in the model, the new text file, and how many steps to take for the fine-tuning process.

i.4 Parameter Setting

The parameters set on the GPT-2 chatbot model controls the kind of responses it generates.

The length parameter decides the output length of the chatbot response. The 'temperature' parameter lets us control how greedy we want the generative model to be for the responses. A smaller value would give grammatically correct answers but with small variation. Higher values would mean grammatical mistakes but responses with more variation and diversity.

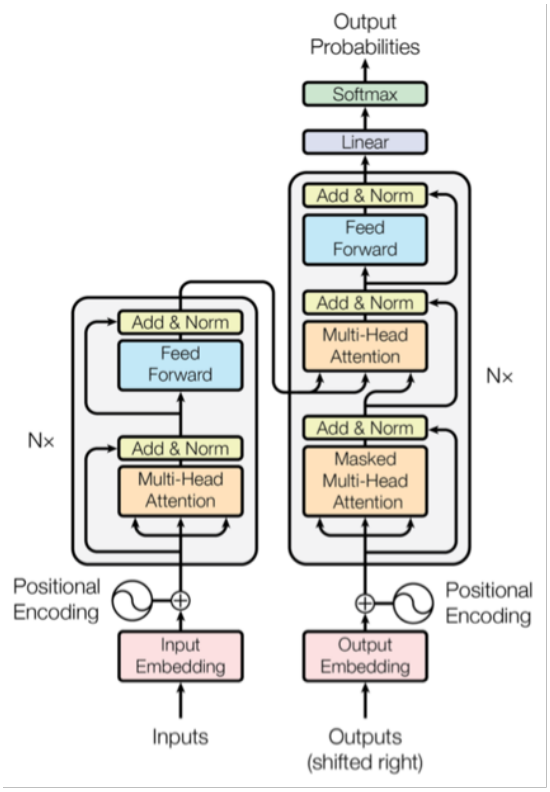

Figure 1: *Transformer Model Architecture*

i.5 Generation

We have two ways of using the agent in the fine-tuned model once it's trained: We let the trained model generate its conservations between a person in distress and some professional advice. We send our input as users to the model and let the model communicate with us in the same way as a general chatbot.

ii. Bag of Word Approach

We are using an Artificial Neural Network [Namrata Bhartiya 2020] to get our inputs processed in such a way that enables the model to generate meaningful responses back to the user. The core functioning for processing the inputs gets handled by Natural Language Processing [Tarun Lalwani, 2018]. NLP starts with breaking the corpus down into smaller components like tokens and further uses tools from the NLTK library like stemming and the usage of stop words to save on computational time and not invest memory on words that do not help the model with predictions.
The architecture for the ANN using the bag_of_words approach is shown below in 2.

We have combined the preprocessing with feature extraction in the NLP preprocessing block as follows:

- **bag_of_words as *features*** : are generated from the preprocessing block to be later stored as features for the training phase, to predict indexes for tags denoting emotions or topics.
- **indexes for tags as *labels*** : we proceed to put indexes of the tags present in the intents file to store as labels for the model.

ii.1 Neural Network

The artificial neural network used for this approach consists of a general input and output layer and one hidden layer for the computation. The length of any bag of word vector in the input set decides the input layer size, and the count of tags in the intents decides the size for the output layer. ReLU is present between each consequent layer as an activation function to only allow the flow of positive values into the network.

After giving predictions for indexed tags based on the bag_of_words passed as inputs into the network, we save the network as a pre-trained classifier.

ii.2 Chat

During a live interaction with the user, as shown in Figure[2]. The input received from the user undergoes preprocessing for its text sequence, generating the bag_of_wordsthat gets passed into the trained classifier for indexed tags.

The indexed tags generated by the trained classifier get checked with the tags present in the intents file. If the particular indexed tag matches with the one in the original corpus, the chatbot generates a random response from the given set of responses designed under that specific 'tag'.

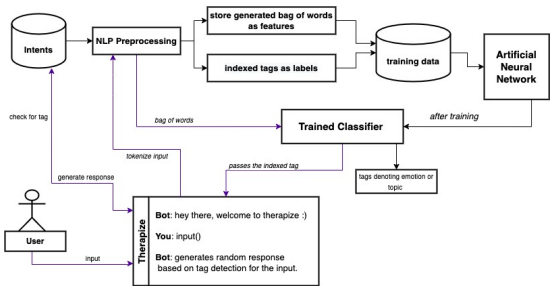

Figure 2: *Architecture for the Bag of Word Approach*

IV. Workflows for the Agents

i. Workflow for the GPT-2 Fine Tuned Model

i.1 Reading the Data

It receives the user input as a sequence of text.

i.2 Generation

The GPT-2 on the therapy_data generates live responses for the user input based on the fine-tuning done on the therapy_data.

ii. Workflow for the bag of word approach

ii.1 Data Collection

The user receives the inputs in the form of texts and sequences and transfers it to preprocessing for preparation of data.

ii.2 NLP Preprocessing

After cleaning the live data received, It returns the bag_of_words generated on the input as shown in Figure[3] and passes it to the trained classifier for prediction.

ii.3 Trained Classifier

The model would predict indexed tags based on the bag_of_words generated by the NLP preprocessing block.

ii.4 Chat

It generates random responses given based on the tag predicted from the neural network on the particular text received by the user.

```
Hey, Welcome to Therapize. Let's chat! (type 'quit' to exit)
--------------------------------------------------------------------------------
You: Hey how are you

Tokenized: ['Hey', 'how', 'are', 'you']

Bag of words: [0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]

Therapize: Hi there, how can I help? :)
--------------------------------------------------------------------------------
```

Figure 3: *Tokenized input and Bag of Words generation*

V. Experiments and Results

i. Bag of Words Approach

The trained ANN with hyperparameters, i.e. num of epochs, batch_size, learning rate, Adam Optimizer and the Cross-Entropy, to calculate the loss between predicted and true labels. Model can reach the minima sooner with high learning rates as shown in Figure[1].

It can be inferred that the model with 150 epochs gives out the same loss like the one running it for 250 and 500 epochs at a learning rate of 0.01, but it is important to note that it is the training loss, and it is plausible for the model to be more overfitting.

So it is not advisable to have much higher values for LR at the expense of not generalising well on the data it has not seen yet.

lr	epochs	loss
0.001	150	0.119
0.001	250	0.022
0.001	500	0.001
0.01	150	0.0003
0.01	250	0.0001
0.01	500	0.0001

Table 1: *Hyper-parameter tuning results*

Softmax at the output layer gives out a probability distribution for scores. Tuning the value for that score with how it matches the tag in the intents is a hyper-parameter for us to use here.

Setting the value for something as low as 45, the model responds to inputs that the machine will not understand; otherwise, it grows softer on the prediction, as shown in Figure [4]. The default for the model is picked to be at a score of 75, which makes it balanced and does not give out responses for tags it does not belong to in the corpus.

```
Hey, Welcome to Therapize. Let's chat! (type 'quit' to exit)
--------------------------------------------------------------------------------
You: Hell
Therapize: Hope we were of any help, see you soon!
--------------------------------------------------------------------------------
You: hel o
Therapize: Bye! Come back again soon :)
--------------------------------------------------------------------------------
You: hell o
Therapize: See you later, thanks for visiting Therapize
--------------------------------------------------------------------------------
```

Figure 4: *Responses generated by the ANN*

ii. GPT-2 Fine-Tuned Model

Accuracy was measured by fine-tuning the GPT2 model on the (therapy_data). Three different step sizes were experimented with to measure the accuracy of predicted tokens and sequences. The resultant losses with different step sizes are shown in Figure [2].

steps	loss
500	1.96
1000	0.46
1500	0.13

Table 2: *Resultant Losses for GPT2 Fine-Tuned*

A conversation made between the User and the 124M GPT2 model fine-tuned on 1500 steps, showed some remarkable results as shown in Figure[5], seeing the data for the first time.

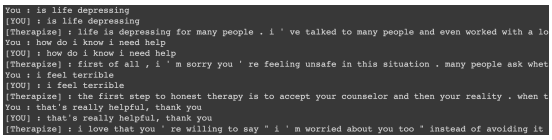

Figure 5: *Responses generated by GPT2 Fine-Tuned*

VI. Conclusion and Future Prospects

- Overall, the bag_of_words approach for chatbots develops an agent capable of generating decent responses.However, the model cannot give unexpected esoteric answers with a broader view if such patterns are not fixed in the intents mentioned.

- Personalities and other information about the user if securely accessed, can be used to generate unique answers for individuals that would have greater impact on them.

- Link the therapy bot to some qualified mental health practitioner available nearby to the user, at the user's permission, to help in any case of emergencies.

References

[Aditya Nrusimha Vaidyam, 2019] Aditya Nrusimha Vaidyam, Hannah Wisniewski, John David Halamka,Matcheri S. Kashavan,John Blake Torous, (2019). Chatbots and Conversational Agents in Mental Health: A Review of the Psychiatric Landscape, *Can J Psychiatry*.

[Deepu S,Pethuru Raj and S.Rajaraajeswari (2016)] A Framework for Text Analytics using the bag_of_words(BoW) Model for Prediction, *IJANA*.

[Ashish Vaswani,2017] Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin, (2017). Attention Is All You Need *IJERT Journal*. arXiv:1706.03762.

[Purushothaman Srikanth, 2019] E.Ushitaasree, G. Paavai Anand, (2019). Conversational Chatbot with Attention Model, *IJITEE*. ISSN:2278-3075.

[Namrata Bhartiya 2020] Namrata Jangid,Sheetal Jannu, Purvika Shukla, Radhika Chapaneri (2020). Artificial Neural Network Based University Chatbot System *IEEE*.

[Tarun Lalwani, 2018] Shashank Bhalotia, Ashish Pal, Shreya Bisen, Vasundhara Rathod (2018). Implementation of a Chatbot System using AI and NLP *IJIRCST*.