

Generate Reflections and Paraphrases out of Distress Stories in Mental Health Forums

Zheng Wang

Supervisor: Kalpani Anuradha Welivita, Dr. Pearl Pu Faltings
Human Computer Interaction Group, IC, EPFL

Abstract

Due to the increasing living pressure, pandemic and other social problems, more and more people are suffering from emotional distress and mental health issues. However, the shortage of counselling services and prejudice against mental illness allow only a limited number of people to receive psychotherapy. To address these problems, therapeutic conversational agents is an excellent assistance to provide counselling services to those people who has difficulties to reach mental health workers. A counselling chatbot can be available 24/7 to assist anyone suffering from distress. Therefore, development of counselling chatbots is of vital importance and should be emphasized. There are already many works in area of conversational agents focusing on counselling. But simulating the counselor is not an easy task since many specific therapeutic techniques are used in the domain of counselling. In this paper, we show the importance of one such technique termed “reflection and paraphrasing” and how we can train chatbots to generate reflections and paraphrases. In this regard, we extracted reflections and paraphrases from existing dialogue datasets discussing distress related issues and formed a large-scale dataset containing reflections and paraphrases. To show the significance of this dataset, we fine-tuned a GPT-2 model to automatically generate reflections and paraphrases on 103K data. We discuss the limitations encountered during this process and how the performance of the reflection and paraphrase generator can be further improved in the future.

Introduction

Mental health is one of the most neglected areas of health globally. The COVID-19 pandemic has further worsened the status of mental health. According to the World Health Organization(WHO), the numbers of people suffering from mental illness are staggering. Today, nearly 1 billion people live with a mental disorder and in low income countries more than 75% of people with the disorder do not receive treatment[26]. Every 40 seconds, a person dies by suicide[26]. In addition, mental health is not just a health issue but also related to a country’s economy. However, there are several reasons which result in the difficulties for addressing mental illness. The first is that in clinical treatment doctors mainly rely on medication to treat patients instead of psychological counselling which is not what patients want.

Copyright © 2022, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

In addition, compared to the number of people who need counseling that health workers is far from adequate and people they can serve is limited. The last but not the least, the prejudice and stigma of mental illness is also an important factor that makes people reluctant to receive psychotherapy. In this case, technology is an excellent aid to mitigate the above issues by providing means of professional psychological counselling through means of artificial intelligence.

Nowadays, dialog systems play an important role in natural language processing. They perform chit-chat with humans or serve as assistants via conversations. A multitude of woks on this area are carried out. Researchers have tried to train various deep neural networks to generate semantically and syntactically correct sentences. These experiments has showed outstanding performance in dealing with specific problems in certain domains and open domain chat. Furthermore, It also can be trained to generate counselling-like responses to help address mental health issues.

Reflection is one of the important conversational strategy used in psychotherapy. It is like holding up a mirror: repeating the patient’s words back to them exactly as they said. Sometimes, counselors might reflect back the whole sentence, or select a few words, or even one single word from what the patients said [12]. There are several reasons why reflection is an effective skill. First of all, it makes the patients feeling heard and in turn deepens trust within the therapeutic relationship. Secondly, it makes the patients listen to what they said again. When they hear their own words, they might think about those words and clarify the inappropriate content. Finally, reflection can be used to express the understanding to patients. What’s more, paraphrasing has been regarded as an influential reflection that greatly contributes to the process’s progress of counselling [11]. It is rephrasing the main content of the patient’s message (usually in a shortened form) to clarify the essence of what he or she has just said [11]. This kind of reflection can encourage additional thoughts and new expressions, which then aid the patient in examining conflicts. Using paraphrasing during counselling also assists the mental health workers to clarify and brighten the patient’s expressions. Hence, when designing chatbots to aid people suffering from emotional distress and mental health issues, reflecting and paraphrasing are very important skills to train them with.

There are many conversation datasets built to train the di-

alog systems to generate paraphrases. For instance, Quora dataset [5] which is duplicated questions collected from Quora ¹ website as the paraphrasing sentences pairs. It has 50K training instances and 20K testing instances. Another dataset often used is MSCOCO [14] with 94K training instances and 23K testing instances. It has human annotated captions of images and each image contains five captions. There also are some datasets such as ACL Anthology Sentence Corpus (AASC) [1], ArXiv Bulk Data [2] and Semantic Scholar Open Research Corpus (S2ORC) [16] consisting of paraphrasing sentences from papers. However, all these datasets are open-domain datasets and has no relevancy to reflections and paraphrases related to the therapeutic domain.

There has been some work done to improve the counselling interactions between chatbots and clients. For example, Tanana et al. [34] introduced a recursive neural network to model the counselor and client dialog turns, and therefore identify the counselor statements. Althoff et al. [4] analyzed a large amount of text-message-based counselling dialogs to understand the language style and symmetry in counselling conversations. Besides, understanding the client’s emotion is also helpful for dialog system to be more like a real counselor. In EmotionLines [6] and OpenSubtitles [15], they labeled the emotion of each utterance collected from TV or movie transcripts by Amazon MTurkers.

Perez-Rosas et al. [22] built a psychological dataset consisting of counselors’ behaviors during Motivational Interviewing encounters. It has 276 MI conversations annotated at utterance level with counselor verbal behaviors including “reflections” using the Motivational Interviewing Treatment Integrity 4.0 (MITI). Shen et al. [2020] develop an automatic dialogue system generating reflections using another counselling dataset: Alexander Street Dataset, which is a collection of Counseling & Therapy videos’ transcripts downloaded from Alexander Street Press ². But the above datasets are not publicly available. To address this issue, we construct a dataset containing reflections and paraphrases by extracting reflecting and paraphrasing sentences from publicly available counselling conversation datasets. After that, we utilize the obtained dataset to fine-tune a pre-trained GPT-2 model to see the quality of the generated reflections and paraphrases. The whole process is shown in Figure 1. During the process of extracting reflections and paraphrases, we use two methods, which are extracting sentences using n-grams and calculating the similarity scores of the listener’s sentence with that of the speaker’s utterance. In the following section, we will describe these two methods in detail and present the statistical analysis of the results obtained by these two methods. The main contribution of this paper is that we collect a dataset containing reflections and paraphrases and the corresponding speaker utterances. This dataset can be used to train dialogue systems to generate reflections and paraphrases. We also test the performance of a generation model trained on the collected data. Furthermore, we discuss the limitations of our work and point to

some future directions that can improve the performance of the reflection and paraphrase generator.

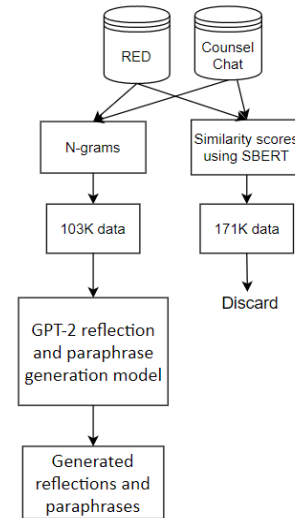


Figure 1: The process of whole project.

Literature Review

There have been significant efforts put in building dialog systems that can generate reflections and paraphrases. The general methods to deal with the tasks of Natural Language Generation (NLG) mainly contain: RNN Seq2Seq [33], Transformer [36], Attention Mechanism [7], GAN [9], Memory Network [32], GNN [29], and Pretrained language models. The pre-trained language models can be divided into two categories: non-contextual and contextual. The non-contextual pre-trained models [18],[21] can learn high quality of word and phrase representations, and are widely used to initialize the embeddings and improve the model performance for generation tasks. The contextual models are mainly based on LSTM or Transformer architecture. For example, ELMo [28] uses deep bidirectional LSTM, BERT [8] uses the Transformer encoder. Meanwhile, more and more effective pre-training methods are proposed to improve the model’s general representation ability. For example, BERT [8] presents a masked language model pre-training objective, MASS [31] designs a pre-training objective to guide the decoder predict the masked sentence fragments in the encoder side.

For adapting the NLG system on the counselling domain, there are many researchers designing different models and techniques to make NLG system generating rich and reliable counselling responses. In 2013, Han et al. [27] designed a conditional random field algorithm to extract “who, what, when, where, why, how” (5W1H) information to counsel and user emotions (happy, afraid, sad, and angry). They use this information to recognize the users’ statement and predict the conversation context. The built system can generate three kinds of counselling responses: paraphrasing, questioning and reflecting feelings. In the same year, a listening-oriented dialog system is introduced by Meguro et al. [35].

¹<https://www.quora.com>

²<https://alexanderstreet.com/>

This system is based on a model trained by a partially observable Markov decision process using human dialog corpus. However, it is limited because it generates responses by selecting utterances from the corpus and only can respond to some general domain utterances. Then, Han et al.[10] further improved the counseling dialog system by extracting more details such as emotion-, problem-, and reason-oriented information and used this information to select the appropriate counseling responses from an external knowledge base (KB). For those utterances which are not relevant to counselling, they adopted a “back-off strategy” to reply a general chat to encourage the speaker to continue interacting. At the same time, they also tried to build a system to generate rich and relevant counselling responses by exploiting a knowledge base (KB). The system extracts an important named entity from a user utterance and then scans the KB to extract contents related to this entity. In 2017, Lee et al.[13] introduced a novel chatbot system for psychiatric counselling service. Their system understands content of conversation based on an LSTM model with emotion recognition. They also generated personalized counseling responses from user input. To implement this, they utilized additional constraints to the generation model for proper response generation which can detect conversational context, user emotion and expected reaction.

The above mentioned articles are trying to emulate the counselor to users in the conversation. Tanana et al.[17] built an automatic dialogue generation system which can help counselors to practice counselling skills. The authors implemented an artificial standardized user that interacts with the counselors and gives the counselor real-time feedback and suggestions based the the specific counselling skills they use. Shen et al. [30] also did some work in this direction. The authors utilized the pre-trained GPT-2 [24] architecture by operating entirely in a sequence-to-sequence way to generate reflections and fine-tune the model with conversations in the counselling domain. They also improved the language models by the context augmentation techniques and applied the domain adaptation on an additional counselling corpus to make the system generating a variety of counselling styles. However, the datasets that they have used are not publicly available and hence their work is not reproducible. In this work, we aim to use dialogues extracted from Reddit emotional distress related conversations and the publicly available Counsel Chat dataset to create a large-scale dataset containing reflections and paraphrases and show how this dataset can be used to generate reflections and paraphrases by fine-tuning a GPT-2 generation model.

Methodology

In this section, we will describe the methodology used to extract reflections and paraphrases from two existing datasets and train a transformer based model to automatically generate reflections and paraphrases given a distress narrative.

First to extract reflections and paraphrases, we experimented with two methods, which are text search using n-grams and obtaining similar sentences to the distress

narratives using cosine similarity. Then we fine-tuned a language generation model GPT-2 [24] on the extracted sentences.

Datasets

The datasets we used to extract the reflections and paraphrases are the RED (Reddit Emotional Distress) dataset [37] and the Counsel Chat dataset [20],[19]. The dialogues in RED are collected from eight subreddits on Reddit that are devoted to provide emotional support for people in distress. The RED dataset contains many dialogues spanning different topics related to mental illness such as depression, anxiety, suicide and financial crisis. Counsel Chat dataset is collected from CounselChat³ on which people can ask questions from professional counselors. In the RED dataset, we consider the speakers to be people undergoing emotional distress and listeners as “amateur” counselors who reflect on what the speakers describe and console them. In the Counsel Chat dataset, the content of questions is the mental health and emotional distress-related issues people experience, and the answers are given by experienced counselors.

	No. of Dialogs	No. of Turns
Entire	1,275,4506	3,398,747
anxietyhelp	8,297	18,351
depressed	10,892	23,804
depression	510,035	1,396,044
depression_help	23,678	51,849
mentalhealthsupport	3,551	7,931
offmychest	437,737	1,064,467
sad	18,827	42,293
suicidewatch	262,469	791,737
consul_chat	20	2271
consul_chat2	-	1482

Table 1: Statistics of the datasets

Extracting Reflections and Paraphrases

Text Search using N-grams Psychologists use a wide range of different keywords such as “sounds like”, “I gather”, and “it looks like” as lead-ins to reflect and paraphrase what the speaker has said. To extract such reflections and paraphrases, we used text search using these keywords. The keywords we utilized are illustrated in Table 2. They are the words listed in the guide for health workers [23] and a helping tool for future mental work students [25]. For the RED and the Counsel Chat dataset, we separated the listener utterances into individual sentences and extracted those that include one or more of the above keywords. Those sentences containing the keywords are sentences that most likely contain reflections and paraphrases. They were written as output to a file along with the corresponding speaker utterance containing the distress narrative.

³<https://counselchat.com/>

Lead-ins Keywords for Paraphrasing
hear you saying
sounds like, sound like
i'm understanding is, i am understanding is, iam understanding is
in other words
what you're saying is, what you are saying is, what youre saying is
do you mean
sounds as if
you sound
i gather
how you felt was
what happened was
you seem to be saying
it seems like, it seem like
from where you sit
it appears as though, it appear as though
it looks like, it look like

Table 2: The leads-in keywords for paraphrasing.

Using Text Similarity Scores We experimented with using text similarity scores to extract paraphrases since paraphrases are more or less similar to what is already stated by the speaker. To find a suitable similarity threshold to extract such paraphrasing sentences, we first calculated the similarity scores between each sentence in the listener utterance and the corresponding speaker utterance. To calculate the scores, we used the Sentence-BERT (SBERT) model [3] to generate the embeddings of the speakers' utterances and the individual sentences in the listeners' utterances and calculated the cosine similarity between the speaker utterance and each sentence of the corresponding listener utterance.

We found that the similarity scores mainly range from 0.2 to 0.6. However, when we observed the listener sentences with high similarity scores, we found that these sentences are describing experiences of the listeners, which are similar to what is experienced by the speakers. Thus, these sentences were not necessarily paraphrasing what the speakers said but rather were sentences disclosing their own experience. Hence, we had to abandon this approach in collecting more reflections and paraphrases.

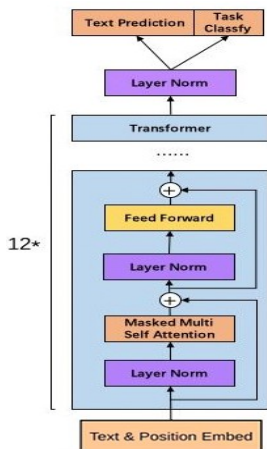


Figure 2: The architecture of GPT-2.

Generating Reflections and Paraphrases

To automatically generate reflections and paraphrases given a distress narrative, we fine tuned the GPT-2 language generation model pre-trained by Huggingface based on the works of "therapy-reflections-generator" [19]. The architecture of GPT-2 is showed in 2. The data used to fine tune the model were the reflections and paraphrases extracted by text search using n-grams because the sentences extracted by using similarity scores did not guarantee to be paraphrasing sentences. There were many causal chatting sentences among them. The total number of data is 102822. The speaker utterances were fed as input to the model and the reflecting and paraphrasing sentences were fed as the output. We divided the dataset into training, validation and test sets according to the ratio 8:1:1. We ran the training for 50 epoches and set the learning rate as 5e-5. The generated reflections and paraphrases are shown in the "Results" section.

Results

In this section, we illustrate the results of each method discussed above.

Text Search using N-grams

The number of reflections and paraphrases we were able to extract using each keyword is illustrated in Figure 3. Figure 4 shows the number of reflections and paraphrases extracted from each dataset by the two methods. Some examples of extracted reflections and paraphrases are shown in Table 3. The detailed statistics are shown in Table 4.

Text Similarity Scores

The number of extracted reflections and paraphrases using text similarity from each dataset is illustrated in Figure 4.

Generating Reflections and Paraphrases

The average perplexity scores of validation and test sets are 16.85 and 16.90, respectively. Some examples of generated reflections and paraphrases and original reflections and paraphrases are shown in Table 5. There are several issues of the output. First of all, the model generated some general responds repeatedly such as "I don't know what to do", "It sounds like you are in a really tough place right now", and "It sounds like you are a good person". Secondly, sometimes it will generate an incomplete sentence. We will discuss some reasons in the nest section.

Discussion

In this paper, we constructed a reflection and paraphrasing dataset with 279,906 reflections and paraphrases extracted from RED and Counsel_Chat datasets. But we discard the data extracted by similarity scores approach because a large number of irrelevant sentences in it. Then we fine-tuned a GPT-2 model based on 82,257 training data and tested its performance on 1,282 testing data. We outlined in detail the

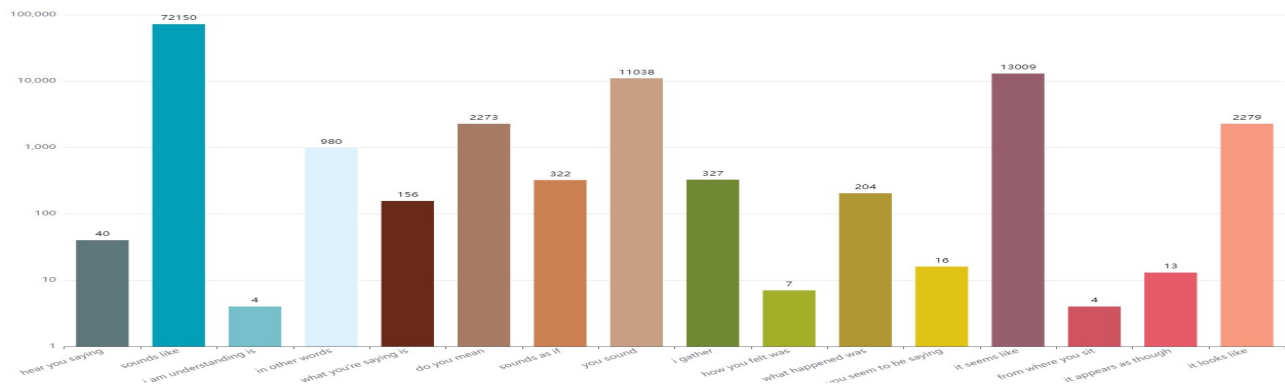


Figure 3: Extracted results of each keywords.

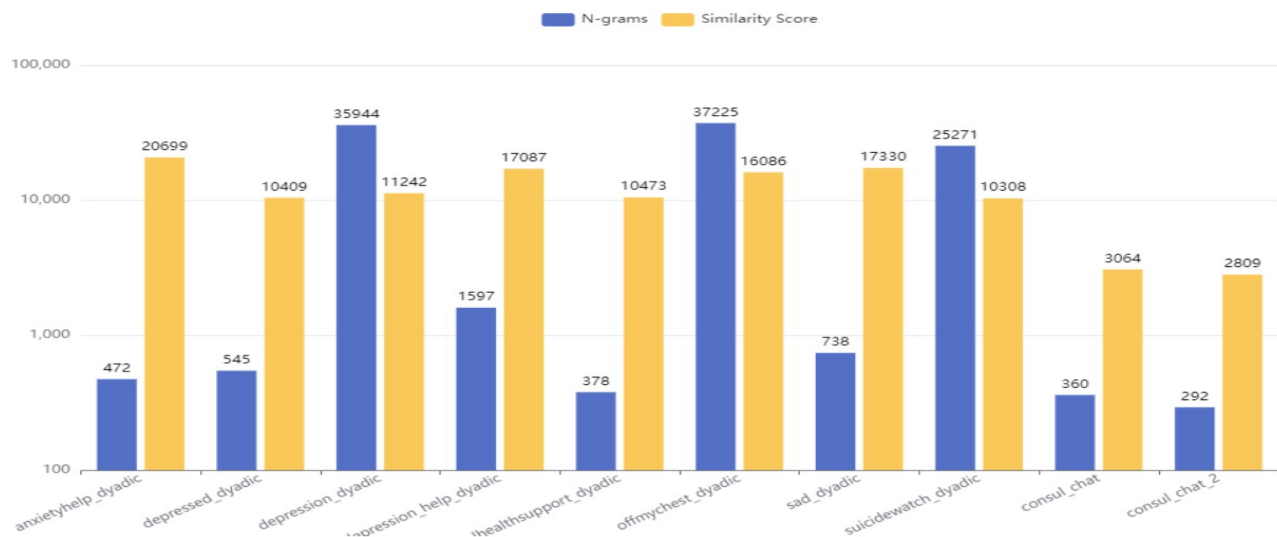


Figure 4: Comparison of extracted results from the two methods.

steps of the process and the final results of each process. As a result, we found that the performance of the fine-tuned reflection and paraphrase generator was not upto the standard we expected. In this section, we will discuss the potential reasons that may have caused this.

There are limited available dialogue datasets focusing on counselling. So, there are limitations in finding good quality reflections and paraphrases. In addition, there are some errors in the original RED dataset, which leads to the incomplete extraction of reflections and paraphrases. For instance, several speaker and listener utterances are stacked in a row. In addition, we did not deal with the typos and special words, which might have influenced the performance of the generation system.

Regarding the reflections and paraphrases extracted using the n-grams methods, we can observe from the results that the distribution of numbers is imbalanced. A large number of sentences are extracted by a few specific n-grams such as “sounds like”, “it seems like”, and “you sound”. There are some keywords that are seldom used in the reflecting

and paraphrasing sentences. However, we cannot recognize which keywords are most used as lead-ins for reflecting and paraphrasing. Hence the way using the keywords one by one to extract sentences is inefficient when there are a large amount of keywords. For improving this way, we can firstly to detect the frequency of them on a medium dataset and then choose the most frequent keywords to apply to the whole dataset. In addition, this method is limited because we only use 16 keywords and their variants to extract paraphrasing sentences. There may be some other keywords which are used in paraphrasing sentences but we are not aware of. Lastly, the final dataset can not guarantee that all the extracted sentences are reflections and paraphrases. There are many general chatting sentences which also use these keywords.

To analyse the similarity score method, there are also several disadvantages of using this method. Firstly, it is a little difficult to choose a suitable score as the threshold. Because, the scores are a bit random and there are many paraphrasing sentences which have the scores outside the

It sounds like you are still in mourning after losing someone.
 You sound like you have some anxiety in your life and that's something you should talk about and seek some help with where you can.
 I hear you saying you've done a lot of things to make him happy.
 what you're saying is, what you are saying is, what you're saying is
 When you mean locked up, do you mean jail or in a mental hospital?
 You sound like you really want to get where you set out to go, but setbacks seemed to have stifled your enthusiasm.
 And from what I gather, your opinion is that she's awesome, unique, and absolutely deserving of everything she wants to get from life.
 What happened was you did not act to take her and someone else got in first.
 You seem to be saying that you can't escape the cycle day-in and day-out.
 Acknowledging this is a big step but it seems like you do a really good job at recognizing those symptoms already.
 It appears as though you have a "grand story" to tell.
 It makes it look like you already know what your issues are and you are ignoring doing anything about it.

Table 3: Paraphrasing sentences extracted by ngrams method.

	anxietyhelp	depressed	depression	depressionhelp	mentalhealthsupport	offmychest	sad	suicidewatch	consulchat	consulchat2
hear you saying	0	0	12	0	0	9	0	8	6	5
sounds like	310	366	24920	1165	288	27276	485	16769	291	233
i am understanding is	1	0	1	0	0	2	0	0	0	0
in other words	4	6	412	19	19	305	6	177	17	15
what you are saying is	1	1	68	1	1	43	1	40	0	0
do you mean	9	14	816	26	6	413	1	40	0	0
sounds as if	1	1	89	2	2	131	2	73	11	10
you sound	31	52	3422	169	19	4463	88	2781	7	6
i gather	4	6	237	7	1	95	3	83	1	0
how you felt was	0	0	0	0	0	4	0	3	0	0
what happened was	0	1	45	4	1	4	0	3	0	0
you seem to be saying	0	0	6	0	0	3	0	7	0	0
it seems like	46	78	5177	175	33	3658	117	3684	22	19
from where you sit	0	1	1	0	0	1	0	1	0	0
it appears as though	0	0	5	0	0	4	0	4	0	0
it looks like	0	19	861	29	8	701	17	641	2	1

Table 4: The extracted number of reflections and paraphrases from each dataset using different keywords.

range. Secondly, using the threshold to extract sentences can not avoid extracting many sentences which also get the scores in this range. The interval we chose is also the interval where many causal chatting sentences would be concentrated.

For the generation model, we find that the final results are not very convincing. There are several reasons leading to it. Firstly, the number of epochs which we used for training is not enough. Secondly, the speaker utterances used as input are too lengthy. Applying a summarization technique or cutting the useless content to reduce the length of the input data may improve the performance. Finally, the overall data used for training may not be enough.

Future work

In the future, we can improve the performance of the reflection and paraphrase generator from several aspects. For example, we can collect more counselling chat data from CounselChat website or other counselling forum, and find more lead-in keywords for reflections and paraphrases. In addition, we can correct the errors in the RED dataset and clean the extracted text which is used for fine-tuning the

generation model. Regarding the generation model, we can use more data for training and train the model for more epochs.

Acknowledgment

I would like to warmly thank everyone who helped me to complete the project. The biggest nod of appreciation goes to my supervisor, Kalpani Anuradha, who faithfully monitored my progress every week and gave me a lot of suggestions and help. I learned a lot of things from her. In addition, I express my gratitude to Dr. Pearl Pu for arranging the project, enabling me to deepen my knowledge in the NLP area.

References

- [1] 2020. AASC: ACL Anthology Sentence Corpus. <https://www.kaggle.com/aymenmouelhi/quora-duplicate-questions>.
- [2] 2020. ArXiv Bulk Data. <https://arxiv.org/help/bulkdata>.
- [3] 2021. SEBRT.net. https://www.sbert.net/docs/usage/semantic_textual_similarity.html.

<p>Speaker: The thought of living with out him, feels like half of me is gone... Empty. I feel worthless. And crazy.</p> <p>Output: It sounds like he is really struggling with his feelings for you.</p> <p>Ground truth: People stigmatize it so it sounds like someone is broken.</p>
<p>Speaker: I dedicated my time to entering contests and calling into radio shows to get free tickets but nothing worked. I am particularly upset about missing this festival. I can't help but feel angry and upset towards my boyfriend who promised me we would go.</p> <p>Output: It sound like a big deal but I just don't know what to do.</p> <p>Ground truth: You sound like a brat expecting your boyfriend to buy you expensive tickets when it's not even His favorite band.</p>
<p>Speaker: Sometimes I feel like my boyfriend was put on this Earth by Satan to breed hate, and I'm agnostic. He wasn't always like this, or at least didn't show me this side of him, hence why I love him and stay.</p> <p>Output: You sound like you're a good person.</p> <p>Ground truth: This sounds like a really unhealthy relationship and should leave him.</p>
<p>Speaker: I made a mistake that changed my life. It resulted in losing my girlfriend of five years. Every day I wake up and go to sleep to guilt, regret, self hatred, anger over my decisions, and a general lack of motivation.</p> <p>Output: You sound like you have some pretty strong feelings about your girlfriend, and I think you should talk to her about it.</p> <p>Ground truth: Sounds like you still have some unresolved issues with your break up.</p>
<p>Speaker: Depression has turned me into a vile thing- a sniveling, resentful, bitter, and utterly self-destructive creature. It's eaten away at my hopes like a rabid wolf and has turned me into a fearful individual. The dark circles under my eyes have grown just like the darkness in my heart has grown. I feel as though Depression has taken away everything that I was and turned me into a hollow shell, empty of meaning and life.</p> <p>Output: I don't know what to do, but it sounds like you're in a really dark place right now.</p> <p>Ground truth: Sounds like we're going through the same thing.</p>

Table 5: Some examples of generated sentences with the original sentences.

- [4] Althoff, T.; Clark, K.; and Leskovec, J. 2016. Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, 4: 463–476.
- [5] AYMENMOUELHI. 2018. Quora Question Pairs. <https://www.kaggle.com/aymenmouelhi/quora-duplicate-questions>.
- [6] Chen, S.-Y.; Hsu, C.-C.; Kuo, C.-C.; Ku, L.-W.; et al. 2018. Emotionlines: An emotion corpus of multi-party conversations. *arXiv preprint arXiv:1802.08379*.
- [7] Chorowski, J.; Bahdanau, D.; Serdyuk, D.; Cho, K.; and Bengio, Y. 2015. Attention-based models for speech recognition. *arXiv preprint arXiv:1506.07503*.
- [8] Devlin, J.; Chang, M.-W.; Lee, K.; and Toutanova, K. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
- [9] Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.
- [10] Han, S.; Kim, Y.; and Lee, G. G. 2015. Micro-Counseling Dialog System Based on Semantic Content. In *IWSDS*.
- [11] Kelly, K., ed. 2017. *PARAPHRASING IN COUNSELLING*. Counsellor Tutor Ltd.
- [12] Kelly, K., ed. 2017. *REFLECTION IN COUNSELLING*. Counsellor Tutor Ltd.
- [13] Lee, D.; Oh, K.-J.; and Choi, H.-J. 2017. The chatbot feels you - a counseling service using emotional response generation. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 437–440.
- [14] Lin, T.-Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*, 740–755. Springer.
- [15] Lison, P.; Tiedemann, J.; and Kouylekov, M. 2018. Opensubtitles2018: Statistical rescoring of sentence alignments in large, noisy parallel corpora. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*.
- [16] Lo, K.; Wang, L. L.; Neumann, M.; Kinney, R.; and Weld, D. S. 2019. S2ORC: The semantic scholar open research corpus. *arXiv preprint arXiv:1911.02782*.
- [17] Michael J Tanana, V. S. D. C. A., Christina S Soma; and Imel., Z. E. 2019. Development and evaluation of clientbot: Patient-like conversational agent to train basic counseling skills. *Med Internet Res*, 21(7):e12529.

- [18] Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G. S.; and Dean, J. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, 3111–3119.
- [19] Nicolas Bertagnolli. 2020-08-28. Counsel Chat: Bootstrapping High-Quality Therapy Data. <https://towardsdatascience.com/counsel-chat-bootstrapping-high-quality-therapy-data-971b419f33da>.
- [20] NSanjay. 2021-02-11. therapy-reflections-generator. <https://github.com/NSanjay/therapy-reflections-generator>.
- [21] Pennington, J.; Socher, R.; and Manning, C. D. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 1532–1543.
- [22] Pérez-Rosas, V.; Mihalcea, R.; Resnicow, K.; Singh, S.; and An, L. 2016. Building a Motivational Interviewing Dataset. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, 42–51. San Diego, CA, USA: Association for Computational Linguistics.
- [23] Project, P. M. H., ed. 2015. *Basic Counselling Skills*. University of Cape Town.
- [24] Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; and Sutskever, I. 2019. Language Models are Unsupervised Multitask Learners.
- [25] Revathi D/O Thangavel. 2009-05-21. for the helpers by the helpers. <https://2105rocks.wordpress.com/the-beginning-stage/listening-response-2-paraphrasing/>.
- [26] RIALDA KOVACEVIC. 2021-02-11. Mental health: lessons learned in 2020 for 2021 and forward. <https://blogs.worldbank.org/health/mental-health-lessons-learned-2020-2021-and-forward>.
- [27] Sangdo Han, D. L., Kyusong Lee; and Lee., G. G. 2013. Counseling Dialog System with 5W1H Extraction. In *Proceedings of the SIGDIAL2013 Conference*, (349-353).
- [28] Sarzynska-Wawer, J.; Wawer, A.; Pawlak, A.; Szymanowska, J.; Stefaniak, I.; Jarkiewicz, M.; and Okruszek, L. 2021. Detecting formal thought disorder by deep contextualized word representations. *Psychiatry Research*, 304: 114135.
- [29] Scarselli, F.; Gori, M.; Tsoi, A. C.; Hagenbuchner, M.; and Monfardini, G. 2008. The graph neural network model. *IEEE transactions on neural networks*, 20(1): 61–80.
- [30] Shen, S.; Welch, C.; Mihalcea, R.; and Pérez-Rosas, V. 2020. Counseling-Style Reflection Generation Using Generative Pretrained Transformers with Augmented Context. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, 10–20. 1st virtual meeting: Association for Computational Linguistics.
- [31] Song, K.; Tan, X.; Qin, T.; Lu, J.; and Liu, T.-Y. 2019. Mass: Masked sequence to sequence pre-training for language generation. *arXiv preprint arXiv:1905.02450*.
- [32] Sukhbaatar, S.; Szlam, A.; Weston, J.; and Fergus, R. 2015. End-to-end memory networks. *arXiv preprint arXiv:1503.08895*.
- [33] Sutskever, I.; Vinyals, O.; and Le, Q. V. 2014. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, 3104–3112.
- [34] Tanana, M.; Hallgren, K.; Imel, Z.; Atkins, D.; Smyth, P.; and Srikumar, V. 2015. Recursive neural networks for coding therapist and patient behavior in motivational interviewing. In *Proceedings of the 2nd workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, 71–79.
- [35] Toyomi Meguro, R. H., Yasuhiro Minami; and Dohsaka., K. 2013. Learning to Control Listening-Oriented Dialogue Using Partially Observable Markov Decision Processes. *ACM Transactions on Speech and Language Processing*, (Vol. 10, No. 4, Article 15).
- [36] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. In *Advances in neural information processing systems*, 5998–6008.
- [37] Yeh, C.-H. 2015. A Dialogue Dataset Containing Emotional Support for People in Distress. *arXiv preprint arXiv:1503.08895*.